Springer Series in Wireless Technology

Manish Mandloi
Devendra Gurjar
Prabina Pattanayak
Ha Nguyen *Editors*

# 5G and Beyond Wireless Systems

## PHY Layer Perspective

Springer

# Springer Series in Wireless Technology

**Series Editor**

Ramjee Prasad, Center for TeleInFrastruktur, C1-107, Aalborg University, Aalborg, Denmark

**Springer Series in Wireless Technology** is a series of monographs, contributed titles and advanced textbooks exploring the cutting edge of mobile telecommunication technologies and promulgating them for the benefit of academic researchers, practicing engineers and students. The series encourages contributions in the theoretical, experimental and practical engineering aspects of wireless communications—voice, data and image transmission. Topics of interest to the series include but are not limited to:

- coding and modulation;
- cognitive radio;
- full-duplex wireless communication;
- model-free design;
- multiple access;
- resource allocation;
- uses of digital signal processing in wireless systems;
- wireless energy transfer;
- wireless networks: 4G, 5G and beyond and next-generation WiFi; adhoc wireless networks, device-to-device networks; heterogeneous mobile networks; wireless sensor networks;
- wireless optical communications.
- Proposals for this series (please use the proposal form that can be downloaded from this page), can be submitted by e-mail to either the:

**Series Editor**
Professor **Ramjee Prasad** Department of Business Development and Technology, Aarhus University, Birk Centerpark 15,8001, Innovatorium, CGC, 7400 Herning, Denmark **e-mail:** ramjee@btech.au.dk

or the

**In-house Editor**
Mr. **Oliver Jackson** Springer London, 4 Crinan Street, London, N1 9XW, United Kingdom **e-mail:** oliver.jackson@springer.com

More information about this series at http://www.springer.com/series/14020

Manish Mandloi · Devendra Gurjar ·
Prabina Pattanayak · Ha Nguyen
Editors

# 5G and Beyond Wireless Systems

## PHY Layer Perspective

Springer

*Editors*
Manish Mandloi
Department of Electronics
and Telecommunication Engineering
SVKM's NMIMS (Deemed to be
University) Shirpur Campus
Shirpur, India

Prabina Pattanayak
Department of Electronics
and Communication Engineering
National Institute of Technology Silchar
Silchar, India

Devendra Gurjar
Department of Electronics
and Communication Engineering
National Institute of Technology Silchar
Silchar, India

Ha Nguyen
Department of Electrical
and Computer Engineering
University of Saskatchewan
Saskatoon, SK, Canada

ISSN 2365-4139 ISSN 2365-4147 (electronic)
Springer Series in Wireless Technology
ISBN 978-981-15-6389-8 ISBN 978-981-15-6390-4 (eBook)
https://doi.org/10.1007/978-981-15-6390-4



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

# Preface

The use of wireless communication systems has become an integral part of our daily lives. With the rapid evolution of wireless technologies, it is now possible to transfer a massive amount of data around the world in fractions of seconds, while simultaneously supporting a large number of wireless users. Recent advancements in wireless technologies such as smart home appliances, wireless health monitoring gadgets, wireless traffic control management services, and ultra-high-definition quality multimedia have added unprecedented comfort for human being. Moreover, through Internet of things (IoT), it is expected that information and control of surrounding objects can be gathered using a single control device. However, the practical implementation of these concepts requires cutting-edge research toward developing new techniques for enabling wireless data transfer among large number of wireless devices with very high data rates. In recent years, several advanced technologies have been proposed to enhance the spectral and energy efficiencies for 5G and beyond wireless systems. The concepts of multiple-antenna devices, power-domain multiplexing, very high-frequency communication, and high channel mobility communication have triggered new dimensions of research in multiple-input multiple-output (MIMO) systems, non-orthogonal multiple access (NOMA), millimeter wave (mm-wave) communications, and vehicular communication, respectively.

This book provides brief introduction, basic concepts, recent advancements, and opportunities for future research in various key enabling technologies for 5G and beyond wireless systems. It also serves as a handy information source for the researchers and practitioners working in all areas of wireless communications, with particular emphasis on physical (PHY) layer techniques. These techniques are massive MIMO, mm-wave communications, spectrum sensing and sharing, device-to-device (D2D) and vehicle-to-everything (V2X) communications, and machine learning.

In massive MIMO systems, a base station (BS) employs hundreds to thousands of antennas. Using a large number of BS antennas enhances diversity and multiplexing gains, which in turn improves the reliability and speed of data transmission, respectively. However, there are several challenges that need to be solved to enable

practical realization of massive MIMO systems, such as pilot contamination, channel estimation, precoding, and multi-user detection. Several chapters provide background on different methods for resolving implementation challenges, discuss limitations of existing techniques, and suggest possible future directions for research in massive MIMO. On the other hand, spatial multiplexing of information streams in MIMO systems requires all the antennas to be active all the time. This generally induces interference among data streams transmitted through different antennas at the receiver and also requires a dedicated RF chain for each transmit antenna. The interference among different data streams, also known as inter-channel interference (ICI), limits the bit error rate (BER) performance. Furthermore, using dedicated RF chains for all antennas increases the system's complexity. To mitigate ICI and to reduce the number of RF chains, index modulation (IM) techniques have been proposed. IM techniques include spatial modulation (SM), generalized SM, media-based modulation (MBM), and orthogonal frequency-division multiplexing with IM (OFDM-IM). This book includes chapters on these advanced modulation techniques, which discuss fundamental concepts, performance comparison, and challenges for future research.

In orthogonal multiple access (OMA) techniques, orthogonal resources are used by different users in either time, frequency, or code domain for avoiding inter-user interference. OMA techniques face several challenges due to its limited connectivity and spectral efficiency. Non-orthogonal multiple access (NOMA) techniques can solve the challenges related to massive connectivity and high spectral efficiency by allocating a higher number of users and sharing the bandwidth occupied by a user with other users. Furthermore, given the availability of vast spectrum resources, mm-wave communications can offer much higher data rates, typically in the order of 10 Gbps. Recent developments in mm-wave communications have emphasized on the exploitation of directional beam-forming at both transmitter and receiver via large-dimensional antenna arrays. Dense deployment of BSs yields more efficient transmission of energy by beam-forming along the line-of-sight (LoS) path to the receiving node. The use of a narrow beam width certainly results in better PHY layer security performance. Moreover, it can also avoid the severe impact of co-channel interference. Unfortunately, the mm-wave signals are susceptible to blockage by various common materials, including human bodies, trees, buildings, and mountains. This requires further research and demonstration on the viability of mm-wave technology in wireless networks. For example, in tactical scenarios, it is not clear whether increasing the number of transmitters helps improve the performance where RF transmitters and receivers are wearable devices. To make the reader aware of these important techniques, several chapters have been devoted to NOMA and mm-wave communications.

Wireless spectrum is one of the most vital components in wireless communications, but unfortunately it is becoming increasingly congested as the number of devices grows day by day. To address this problem, the system needs to rely on flexible techniques for spectrum sharing that can fulfill the growing requirements of the next-generation wireless networks. Various spectrum sharing schemes have been developed for cognitive radio networks in which the key idea is to exploit

white spaces of wireless spectrum. In particular, various schemes have been shown to help reducing the cellular traffic through the BS by enabling direct communications among end devices. As a result, D2D communication has been regarded as one of the pivotal techniques in 5G and beyond wireless communications and networks due to its unique advantages that include proximity gain and efficient BS off-loading. For example, when two wireless devices are in close proximity, enabling direct communication between them can provide a crucial solution for several real-time applications, e.g., real-time monitoring, video gaming, file sharing, etc. A few chapters in this book discuss different spectrum sensing techniques and resource allocation for D2D communications.

This book also contains chapters that extensively cover various forms of vehicular communications and the role of physical layer security in such security-prone systems. In general, vehicular communication is a promising technology to realize automatic vehicle systems and smart traffic handling, and improving the safety of drivers, passengers, and pedestrians. It aims to modernize the operation of vehicles, help drivers to take appropriate decisions, and manage vehicle traffic along with offering useful applications for passengers. This technology enables a platform for sending and receiving data related to traffic, road conditions, and other environmental conditions in the form of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). For instance, if any vehicle detects a hazard, the sensors of that vehicle automatically alert the corresponding platform, which can then spread the information to the nearby vehicles connected through the same platform so that these nearby vehicles can apply risk measures. However, due to the mobile nature of vehicles, the topology for vehicular networks turns out to be highly dynamic and therefore poses several challenges for its implementation. Diverse communication environments and mixtures of different communication forms of V2X communications make the modeling of propagation environments complex and challenging. The high density of vehicles with high mobility and random surroundings makes characteristics of vehicular channels far from trivial. This means that channel models that have been extensively studied for the cellular networks might not be directly suited for vehicular networks. Taking these facts into account, a few chapters in this book present fundamental details of different kinds of vehicular communications and examine challenges and roadblocks that need to be resolved.

## Organization of the Book

Different chapters covering the fundamentals and PHY layer design aspects of different techniques can be useful to the readers to get familiar with the fundamentals and to identify research problems.

In Chap. 1, the authors discuss fundamental concepts of conventional single-antenna systems and highlight the motivations for introducing multiple-antenna communication systems. Multiple-antenna (MIMO) system helps to achieve higher

system throughput (by enabling multiple data streams) and reliability (by exploiting multiple diversity paths). This chapter focuses on various aspects of designing MIMO antennas. Chapter 2 provides insights into massive MIMO and covers the pilot contamination aspect in such systems. In massive MIMO, the BS requires complete and accurate channel state information (CSI) for maximizing the benefits of massive MIMO technology. The CSI can be obtained either through feedback from users or channel reciprocity techniques. The use of non-orthogonal pilot sequences sent by users for determining CSI is the main source of pilot contamination in the uplink data communications. Various pilot contamination scenarios and possible solutions are extensively studied in this chapter. Chapter 3 examines the problem of symbol detection in uplink massive MIMO systems. The traditional linear detectors such as zero-forcing and minimum mean-square error detectors are presented, and their limitations are discussed. Several approximate forms of inverse and iterative detection techniques are also presented as low-complexity alternatives to the linear detection schemes. This chapter also includes discussion on BER comparison and computational complexity. In Chap. 4, various index modulation (IM) techniques are introduced, which serve as efficient low-complexity solutions for reliable information transmission in 5G and beyond wireless systems. IM schemes based on indexing across different resources are discussed in detail, including system models and spectral efficiencies. Furthermore, the maximum likelihood detection scheme is discussed for each of the IM schemes presented in this chapter.

Chapter 5 covers the design and detection of a sparse code multiple access (SCMA) scheme. It also provides a brief introduction to the code-domain-based SCMA and non-orthogonal multiple access (NOMA) for achieving high spectral efficiency and low latency in 5G and beyond wireless systems. Furthermore, this chapter also presents the hybrid multiple access schemes based on code-domain and power-domain NOMA with simulation results showing the impact of SCMA. In Chap. 6, performance of practical NOMA systems under hardware impairment is examined. In particular, closed-form expressions for different performance metrics such as outage probability and throughput are derived and useful insights are drawn. Moreover, detection ability with both imperfect successive interference cancelation (SIC) and perfect SIC in the presence of hardware impairments is also discussed. Chapter 7 presents an overview of various challenges associated with mm-wave-based cellular networks and introduces key design concepts. This chapter also introduces a stochastic geometry-based framework to assess the effect of exploiting mm-wave for cognitive-based multi-tier cellular networks. Several metrics such as outage probability, area spectral efficiency, and energy efficiency are theoretically derived, with numerical results to demonstrate the advantages of the proposed approach.

Chapter 8 examines recent works done in the area of mobile multi-user full-duplex (FD) relaying systems. Major challenges for the deployment of FD systems and the modeling of residual self-interference at a FD node are also discussed. Further, works related to the multi-user half-duplex and FD with different relaying schemes are presented in detail. This chapter also highlights the

performances of various multi-user scheduling schemes and suggests various potential applications of FD relaying for mission-critical communications. Chapter 9 presents a comprehensive overview of various filter techniques for reducing the peak-to-average power ratio (PAPR) of waveforms envisioned for 5G and beyond wireless systems. The authors also conduct comparative performance analysis of these techniques to elucidate their merits. They also discuss different non-orthogonal waveforms such as generalized frequency-division multiplexing, filter bank multicarrier (FBMC), biorthogonal frequency-division multiplexing, and universal filtered multicarrier.

In Chap. 10, the authors describe a strategy for minimizing the total power consumption while satisfying the desired coverage of participating nodes to provide the minimum required throughput over a wireless network. In order to achieve the goal, they provide an optimization framework for the deployment of BSs, their number, and transmit power under different scenarios. In Chap. 11, different types of spectrum sensing (SS) algorithms are introduced, which include blind SS techniques and non-blind SS techniques. In the discussion, different noisy scenarios are considered such as white Gaussian noise and colored noise. It also introduces the concept of cooperative spectrum sensing and FD communications for 5G and beyond wireless systems. Chapter 12 investigates performance of an energy harvesting (EH)-enabled cooperative cognitive radio network (CRN) under Weibull and Erlang distributions. In this chapter, the authors analyze the performance in terms of prediction error, collision constraint, number of CR nodes under cooperation, number of frames, and shape parameters of the distribution functions on the performance of EH-CRN. In Chap. 13, the authors analyze the impact of the buffer length on blocking probability and cumulative handoff delay (CHD) for various proactive spectrum handoff schemes. Moreover, this chapter includes detailed comparison of results for blocking probability and CHD in terms of the primary user's arrival rate and mobility parameter of spectrum holes.

Chapter 14 discusses the integration of IoT into satellite networks, which is one of the key focuses in 5G and beyond systems. In this chapter, the authors present comprehensive performance analysis of an overlay multi-user satellite-terrestrial network (OMSTN) where terrestrial secondary devices cooperatively assist multi-user primary satellite communications to access spectrum for secondary communications. They perform analysis on the achievable diversity order of the considered OMSTN under two scenarios, namely when interferers' power is fixed and when it varies with the transmit power of the main satellite and IoT nodes. The authors also introduce a scheme for adaptively choosing the power splitting factor.

In Chap. 15, the authors first give a brief introduction to the potential usage and different modes of operation of D2D communication in the next generation of wireless networks. Then, to seamlessly integrate D2D communications in cellular networks, they present the challenges in designing optimal resource allocation schemes. The authors then classify existing resource allocation schemes based on challenges of different D2D scenarios. Chapter 16 discusses various forms of vehicular communications such as V2V, V2I, V2P, and V2N. This chapter also highlights recent advancements done for different modes of vehicular

communications, collectively referred to as V2X communications. Moreover, it identifies potential research directions to address challenges for bringing V2X communications into operation. As security is one of the major concerns in vehicular communications, Chap. 17 provides a framework to ensure secure communication in cooperative vehicular relay networks with the help of physical layer security techniques. In this chapter, the authors examine a cooperative vehicular network in which a moving source vehicle communicates to a fixed destination infrastructure terminal with the help of a moving relay vehicle in the presence of a passive eavesdropper vehicle. For the considered network, the authors evaluate the system secrecy performance in terms of secrecy outage probability and other performance metrics.

Recent developments in machine learning algorithms in which the models are capable of learning from the data themselves have been proved very promising for various wireless communication problems and scenarios. Motivated by this, Chap. 18 identifies design problems faced in 5G infrastructure that can be modeled as machine learning problems. The authors provide detailed explanation on how to leverage well-explored machine learning models for solving the problem at hand. They also discuss the idea of exploiting spatial and temporal information present in both user-level and network-level data in solving many infrastructure planning problems.

Shirpur, India — Manish Mandloi
Silchar, India — Devendra Gurjar
Silchar, India — Prabina Pattanayak
Saskatoon, Canada — Ha Nguyen

# Contents

# Editors and Contributors

## About the Editors

**Dr. Manish Mandloi** is an Assistant Professor in the Department of Electronics and Telecommunication Engineering, SVKM's NMIMS (Deemed to be University) Shirpur Campus, Shirpur, India. He received his B.Tech. in electronics and communications engineering from SASTRA University, Thanjavur, India, in 2011, Master's in telecommunications engineering from Indian Institute of Science Bangalore, India, in 2013, and Ph.D. in electrical engineering from Indian Institute of Technology Indore, India, in 2017. His major areas of research include low-complexity signal processing techniques for massive MIMO communication systems, and simultaneous wireless information and power transfer. He has published several papers in reputed international journals. He received the IEEE Communications Letters exemplary reviewer recognition for the year 2017.

**Dr. Devendra Gurjar** is working as an Assistant Professor in the Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Assam, India. He received the B. Tech. in electronics and communications engineering from Uttar Pradesh Technical University, Lucknow, India, in 2011, the M. Tech. in wireless communications and computing from the Indian Institute of Information Technology Allahabad, India, in 2013, and his Ph.D. in electrical engineering from the Indian Institute of Technology Indore, India, in 2017. He was with the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, Canada, as a Post-Doctoral Research Fellow. His current research interests include MIMO communication systems, cooperative relaying, device-to-device communications, physical layer security, and simultaneous wireless information and power transfer. He has numerous publications in peer-reviewed journals and international conferences. He is a member of the IEEE Communications Society and the IEEE Vehicular Technology Society.

**Dr. Prabina Pattanayak** is working at the National Institute of Technology Silchar, Assam, India, as Assistant Professor in the Department of Electronics and Communication Engineering. He has received the B. Tech. in Electronics and Telecommunication engineering in 2007 and M. Tech. in Electronics and Communication Engineering (with specialization in Wireless Communication Technology) in 2012 from Biju Patnaik University of Technology, Odisha, India. He obtained his Ph.D in Electrical Engineering from the Indian Institute of Technology Patna, India, in 2017. He was the recipient of Gold medal for his M.Tech degree. He has also served HCL Technologies Ltd. in the capacity of Lead Engineer from 2007 to 2010. His current research interests include multi-user MIMO, MIMO-OFDM, Massive MIMO communication systems, Soft Computing for various communication areas, Smart Grid Communications etc. He has published several articles in peer-reviewed journals and reputed national and international conferences. He is a member of the IEEE Communications Society, the IEEE Vehicular Technology Society, and the IEEE Information Theory Society.

**Prof. Ha Nguyen** joined the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, Canada, in 2001, and currently holds the position of NSERC/Cisco Industrial Research Chair in Low-Power Wireless Access for Sensor Networks. He received the B.Eng. degree from the Hanoi University of Technology, Hanoi, Vietnam, in 1995, the M.Eng. degree from the Asian Institute of Technology, Bangkok, Thailand, in 1997, and the Ph.D. degree from the University of Manitoba, Winnipeg, Canada, in 2001. He co-authored the textbook "A First Course in Digital Communications" (published by Cambridge University Press). His current research interests include communication theory, wireless communications, and statistical signal processing. Dr. Nguyen was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2011 and IEEE WIRELESS COMMUNICATIONS LETTERS from 2011 to 2016. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is a Fellow of the Engineering Institute of Canada.

## Contributors

**Abhishek** Indian Institute of Information Technology, Allahabad, India

**Bhargav Appasani** School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Wasim Arif** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Gaurav Singh Baghel** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Ajay Bhardwaj** School of Computing and Electrical Engineering, IIT Mandi, Mandi, Himachal Pradesh, India

**Vimal Bhatia** Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, Madhya Pradesh, India

**Abhijeet Bishnu** Institute for Digital Communications, School of Engineering, The University of Edinburgh, Scotland, UK

**Umesh Chandra** School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Arijit Datta** Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, Madhya Pradesh, India

**Kuntal Deka** Indian Institute of Technology Goa, Farmagudi, Ponda, Goa, India

**Abhinaba Dey** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Dinh-Thuan Do** Wireless Communications Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

**Shubham Dwivedi** National Institute of Technology Silchar, Silchar, Assam, India

**Sourav Ghosh** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Tripti Goel** National Institute of Technology Silchar, Silchar, Assam, India

**Deepika Gupta** Department of Electronics and Communication Engineering, Dr S P M International Institute of Information Technology, Naya Raipur, Chhattisgarh, India

**Devendra Singh Gurjar** ECE, National Institute of Technology Silchar, Silchar, Assam, India

**Shanidul Hoque** Department of Electronics and Communication Engineering, Madanapalle Institute of Technology & Science, Madanapalle, India

**Amitkumar V. Jha** School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Ersan Kabalci** Faculty of Engineering and Architecture, Nevsehir Haci Bektas Veli University, Nevsehir, Turkey

**Yasin Kabalci** Electrical and Electronics Engineering, Faculty of Engineering, Nigde Omer Halisdemir University, Nigde-Turkey, Turkey

**Deepak Kumar** Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, India

**Mukesh Kumar Mishra** Samrat Ashok Technological Institute, Vidisha, India

**Prabhat Kumar Sharma** Department of ECE, Visvesvaraya National Institute of Technology, Nagpur, India

**Chi-Bao Le** Faculty of Electronics Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

**Manish Mandloi** Department of Electronics and Telecommunication Engineering, SVKM's NMIMS Shirpur Campus, Shirpur, Maharashtra, India

**Neelesh Mehra** Samrat Ashok Technological Institute, Vidisha, India

**Anshul Pandey** Department of Electronics and Communication Engineering, Indian Institute of Information Technology Allahabad, Prayagraj, India

**Prabina Pattanayak** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, Assam, India

**Ganesh Prasad** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Srinivas Ramavath** School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

**Jyotsna Rani** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Prasanna Raut** Department of ECE, Visvesvaraya National Institute of Technology, Nagpur, India

**Pankaj K. Sharma** Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha, India

**Sanjeev Sharma** Indian Institute of Technology (BHU) Varanasi, Varanasi, Uttar Pradesh, India

**Devendra Singh Gurjar** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**M. V. Swati** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Banani Talukdar** Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India

**Tu-Trinh Thi Nguyen** Faculty of Electronics Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

**Aditya Trivedi** Atal Bihari Vajpayee-Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior, India

**Shekhar Verma** Indian Institute of Information Technology, Allahabad, India

**Suneel Yadav** Department of Electronics and Communication Engineering, Indian Institute of Information Technology Allahabad, Prayagraj, India

**Budharam Yogesh** Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha, India

# Abbreviations

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project |
| 4G | Fourth Generation |
| 5G | Fifth Generation |
| ADC | Analog-To-Digital Converter |
| AE | Auto-Encoder |
| ARAIM | Advanced RAIM |
| ASE | Area Spectral Efficiency |
| ASIC | Active SI Cancelation |
| AWGN | Additive White Gaussian Noise |
| B5G | Beyond 5G |
| BC | Broadcast |
| BER | Bit Error Rate |
| BFDM | Biorthogonal Frequency-Division Multiplexing |
| BPP | Binomial Point Process |
| BS | Base Station |
| BSS | Blind Spectrum Sensing |
| CCDF | Complementary Cumulative Distribution Function |
| CD | Code Domain |
| CDF | Cumulative Distribution Function |
| CDR | Call Detail Record |
| CFDP | Cycle Frequency Domain Profile |
| CG | Conjugate Gradient |
| CHCN | Cognitive-Enabled Hybrid Cellular Network |
| CHD | Cumulative Handoff Delay |
| CNN | Convolutional Neural Network |
| Corr | Correlation |
| Cov | Covariance |
| CP | Cyclic Prefix |
| CPB | Channel Power-Based |
| CPBS | Cognitive Pico Base Station |

| | |
|---|---|
| CR | Cognitive Radio |
| CRN | Cognitive Radio Network |
| CSI | Channel State Information |
| CSS | Cooperative Spectrum Sensing |
| CST | Cyclo-Stationary Technique |
| CT | Continuous Time |
| CU | Cellular User |
| CVRN | Cooperative Vehicular Relay Network |
| D2D | Device-to-Device |
| DBN | Deep Belief Network |
| DF | Decode and Forward |
| DL | Downlink |
| DOT | Department Of Transportation |
| DRL | Deep Reinforcement Learning |
| DT | Discrete Time |
| DTV | Digital TV |
| DVB-S2X | Digital Video Broadcast-Satellite Second-Generation Extension |
| DWT | Discrete Wavelet Transform |
| EBG | Electromagnetic Band-Gap |
| ED | Energy Detection |
| EE | Energy Efficiency |
| EGC | Equal Gain Combing |
| Egnval | Eigenvalue |
| EH | Energy Harvesting |
| EH-CRN | Energy Harvesting Cognitive Radio Network |
| ELDRSCM | Energy with Logdet of Received Samples Covariance Matrix |
| eMBB | Enhanced Mobile Broadband |
| EME | Energy with Minimum Eigenvalue |
| eNB | Evolved Node Base Station |
| ENP | Estimated Noise Power |
| ESA | European Space Agency |
| ESC | Ergodic Secrecy Capacity |
| ETS | Extra-Terrestrial Source |
| EVM | Error Vector Magnitude |
| FBMC | Filter Bank Multicarrier |
| FC | Fusion Center |
| FCC | Federal Communications Commission |
| FD | Full Duplex |
| FDD | Frequency-Division Duplex |
| FMC | Filtered Multicarrier |
| F-OFDM | Filtered OFDM |
| FU | Far User |
| GA | Genetic Algorithm |
| GEO | Geostationary Earth Orbit |
| GFDM | Generalized Frequency-Division Multiplexing |

| | |
|---|---|
| GS | Gauss–Siedel |
| GSAE | Global Stacked Auto-Encoder |
| GSM | Generalized Spatial Modulation |
| GSM-MBM | Generalized Spatial Modulation-Media-Based Modulation |
| HAP | High-Altitude Platform |
| HCPP | Hard Core Point Process |
| HD | Half Duplex |
| HetNet | Heterogeneous Network |
| HMA | Hybrid Multiple Access |
| HMM | Hidden Markov Model |
| HPA | High-Power Amplifier |
| HPPP | Homogeneous Poisson Point Process |
| HSPPP | Homogeneous Spatial Poisson Point Process |
| i.i.d. | Independent and Identically Distributed |
| i.ni.d | Independent and Non-Identical Fading Distribution |
| ICI | Inter-Carrier Interference |
| IFFT-PPN | IFFT Polyphase Network |
| IM | Index Modulation |
| IoT | Internet of Things |
| ipSIC | Imperfect SIC |
| ISD | Iterative Sequential Detection |
| ISI | Intersymbol Interference |
| ITS | Intelligent Transportation System |
| ITU | International Telecommunication Union |
| IU | Interrupted User |
| IVC-CE | Inter-Vehicle Cooperation Channel Estimation |
| kNN | K-Nearest Neighbor |
| KPI | Key Performance Indicator |
| LDPC | Low-Density Parity Check |
| LDRSCM | Logdet of Received Samples Covariance Matrix |
| LDS | Low Density Spreading |
| LEO | Low Earth Orbit |
| LL | Low Latency |
| LNA | Low Noise Amplifier |
| LOS | Line Of Sight |
| LSAE | Local Stacked Auto-Encoder |
| LSTM | Long Short-Term Memory |
| LTE | Long-Term Evolution |
| LUN | Licensed User Network |
| MAC | Multiple Access Channel |
| MAP | Maximum A Posteriori |
| MBM | Media-Based Modulation |
| MBS | Macro-Base Station |
| MEO | Medium Earth Orbit |
| MIM | Multidimensional IM |

| | |
|---|---|
| MIMO | Multiple-Input Multiple-Output |
| MINLP | Mixed Integer Nonlinear Programming |
| ML | Maximum Likelihood |
| MLP | Multilayer Perceptron |
| MMS | Macro-Mobile Station |
| MMSE | Minimum Mean-Square Error |
| mMTC | Massive Machine-Type Communication |
| mm-wave | Millimeter Wave |
| mmWBSs | mm-Wave Base Stations |
| MPA | Message Passing Algorithm |
| MRC | Maximum Ratio Combining |
| MSE | Mean-Square Error |
| MUD | Multi-User Detection |
| Mu-LSTM | Multivariate LSTM |
| Multi-GNSS | Multi-Global Navigation Satellite System |
| MU-MIMO | Multi-User MIMO |
| MV | Mobile Vehicle |
| NBSS | Non-Blind Spectrum Sensing |
| NI | Newton Iteration |
| NLOS | Non-Line Of Sight |
| NOMA | Non-Orthogonal Multiple Access |
| NRAO | National Radio Astronomy Observatory |
| NSE | Neumann Series Expansion |
| NU | Near User |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| OFDM-IM | Orthogonal Frequency-Division Multiplexing with Index Modulation |
| OMA | Orthogonal Multiple Access |
| OMSTN | Overlay Multi-User Satellite-Terrestrial Network |
| OP | Outage Probability |
| OWR | One-Way Relaying |
| PAPR | Peak-to-Average Power Ratio |
| PC | Principal Component |
| PCP | Poisson Cluster Process |
| PD | Power Domain |
| PGFL | Probability Generating Functional |
| PLS | Physical Layer Security |
| PMS | Pico Mobile Station |
| PPP | Poisson Point Process |
| PPR | Primary Protection Region |
| PSD | Power Spectral Density |
| pSIC | Perfect Successive Interference Cancelation |
| PSIS | Passive SI Suppression |
| PT | Pilot Tone |
| PU | Primary User |

| | |
|---|---|
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RAIM | Receiver Autonomous Integrity Monitoring |
| RBG | Random Bipartite Graph |
| RF | Radio Frequency |
| RI | Richardson Iteration |
| RMSE | Root-Mean-Square Error |
| RN | Relay Node |
| RNN | Recurrent Neural Network |
| RSC | Resource Size Control |
| Rx | Receiving |
| SBAS | Satellite-Based Augmentation System |
| SBS | Small Base Station |
| SCMA | Sparse Code Multiple Access |
| SCN | Standard Condition Number |
| SD | Sphere Decoder |
| SE | Spectral Efficiency |
| SH | Spectrum Handoff |
| SI | Self-Interference |
| SIC | Successive Interference Cancelation |
| SINR | Signal-to-Interference-Plus-Noise Ratio |
| SIR | Signal-to-Interference Ratio |
| SISO | Single-Input Single-Output |
| SLE | Scaled Largest Eigenvalue |
| SM | Spatial Modulation |
| SM-MBM | Spatial Modulation-Media-Based Modulation |
| SOP | Secrecy Outage Probability |
| SPA | Smart Pilot Assignment |
| SS | Spectrum Sensing |
| SSK | Space Shift Keying |
| ST-IM | Space-Time IM |
| STN | Satellite-Terrestrial Network |
| SU | Secondary User |
| SU-MIMO | Single-User MIMO |
| SWIPT | Simultaneous Wireless Information and Power Transfer |
| TCP | Transmission Control Protocol |
| TDD | Time-Division Duplex |
| TDSC | Time-Domain Symbol Cross-Correlation |
| TS | Tabu Search |
| TW | Two-Way |
| TWR-NOMA | Two-Way Relay NOMA |
| Tx | Transmitting |
| UE | User Equipment |
| UFMC | Universal Filtered Multicarrier |
| UHF | Ultra-High Frequency |

| | |
|---|---|
| UL | Uplink |
| uRLLC | Ultra-Reliable Low-Latency Communication |
| UUN | Unlicensed User Network |
| V2I | Vehicle-to-Infrastructure |
| V2N | Vehicle-to-Network |
| V2P | Vehicle-to-Pedestrian |
| V2V | Vehicle-to-Vehicle |
| V2X | Vehicle-to-Everything |
| VANET | Vehicular Ad Hoc Network |
| VBS | Vehicular Base Station |
| VNI | Visual Network Index |
| WCS | Wireless Communication System |
| WGN | White Gaussian Noise |
| WiMAX | Wireless Interoperability for Microwave Access |
| WRAN | Wireless Regional Area Network |
| WSN | Wireless Sensor Network |
| ZF | Zero-Forcing |

# Chapter 1
# MIMO Antennas: A 5G Communication Perspective


**Gaurav Singh Baghel, M. V. Swati, and Sourav Ghosh**

**Abstract** The conventional single-input-single-output (SISO) system exhibits a major problem of having limited throughput, coverage and data rate. These drawbacks can be overcome by smart antenna technologies such as multiple-input-multiple-output (MIMO) technologies and beamforming antenna array. In MIMO systems, multiple antennas are used for both transmission and reception, which is the crucial architecture for emerging wireless communication systems. MIMO technologies increase the system throughput thanks to the utilization of multiple data streams for transmitter and receiver ends. In current fourth-generation (4G) communication systems, the $n \times n$ MIMO architecture has the number of antennas $n$ as 2 or 4. But in the emerging fifth-generation (5G) communication systems, the number of antennas increases to $n > 8$. This practice of using a greater number of antennas solves the problem of limited throughput, coverage and data rate. MIMO antenna design is not a minor task; rather, it has to be done carefully in order to uphold some performance metrics.

**Keywords** MIMO antenna · Channel capacity · System throughput · Electromagnetic band-gap (ebg) · Massive MIMO · mm-wave MIMO.

## 1.1 Introduction

With the rapid growth of fifth-generation (5G) communication systems, MIMO technology is one of the most promising technologies for future mobile communication.

G. S. Baghel (✉) · M. V. Swati · S. Ghosh
Department of Electronics and Communication Engineering,
National Institute of Technology Silchar, Silchar, India
e-mail: gauravsingh.5290@ieee.org

M. V. Swati
e-mail: swati.mv18@ieee.org

S. Ghosh
e-mail: souravg256@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_1

The demand of 5G system is increasing because of its high transmission rate, low latency compared to the current 4G system (Andrews et al. 2014). International Telecommunication Union (ITU) has reported the frequency bands of 3.4–3.6 GHz, 5–6 GHz, 24.25–27.5 GHz, 37–40.5 GHz and 66–76 GHz bands for 5G communication (Marcus 2020) and the Federal Communications Commission (FCC) has declared the spectrum of 27.5–28.35 GHz for 5G.

In MIMO systems, spatial multiplexing is used to split complete data into lower data rate streams to transmit using $n$-antennas. To distinguish these different data streams in receiver end, signal-to-noise ratio (SNR) should be very high which can be improved by using appropriate communication protocols. Full Duplex and Spatial Modulation techniques can easily mitigate the issues existing in conventional MIMO systems for 4G communications such as multiple RF chains involvement, synchronism among multiple radiating elements and inter-channel interference. To attain the multiplexing gain by transmitting the information simultaneously, multiple number of radiating elements are required at both receiving (Rx) and transmitting (Tx) sides (Rappaport et al. 2015). Therefore, MIMO antenna system is the prerequisite requirement for satisfying the large throughput with available bandwidth (Abdullah et al. 2019).

The MIMO system throughput (channel capacity) can be conveniently enhanced by increasing the number of radiating elements at the base station and mobile terminal simultaneously (Sanchez-Fernandez et al. 2008). MIMO provides the high quality of services for different multimedia applications for wireless communication system of future generation. Although the diversity and spatial multiplexing of the system can be improved, the complexity is increased due to a large number of radiating elements; hence, cost of the radio frequency (RF) chain is also increased. MIMO system with $N_r$ receiving antenna and $N_t$ transmitting antenna requires complete $N_t \times N_r$ number of RF chains which include low noise amplifier (LNA), down-converter and analog-to-digital converter (ADC). So the additional number of digital signal processing and large number of RF chains make the MIMO system complex and expensive (Molisch and Win 2004). This is the main drawback of MIMO system. To overcome this problem, one of the possible solutions is to implement different optimization techniques for selecting optimum number of antennas without degrading its system performance. Genetic algorithm (GA) is one of the most effective approaches to optimize the number of receiving and transmitting antennas in MIMO system with low computational complexity (Lu and Fang 2007).

Other characteristics to determine the performance of massive MIMO systems ($n > 8$) are mutual coupling. Electromagnetic band-gap (EBG) structures are used in the conventional MIMO antenna systems to overcome the disadvantage of mutual coupling in the system. The millimeter-wave MIMO technologies with high gain antennas are the promising candidate for high data rate applications such as high-quality video (8K) streaming with reduced path loss. Incorporating cognitive radio in such architectures enhances the spectrum utilization ability of a system which is a crucial need for 5G communication. Currently, 2.6 GHz frequency band is used in MIMO technology for 4G end users. For the sub-6 GHz 5G communication portable devices, the 3.6 GHz frequency band appears promising.

The key advantage of MIMO system is the compactness of incorporated radiating elements, which allow the higher frequency operation as well as reduction in interference without bandwidth expansion.

## 1.2 Single Element Versus Multiple Antenna System

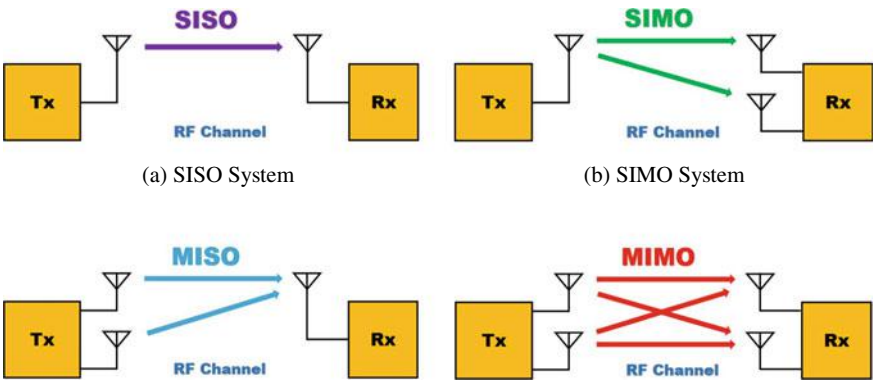
There are many challenges of SISO antenna system such as multi-path fading and low channel capacity ($C$). The channel capacity ($C$) of a communication system can be mathematically given as:

$$C = \log_2\left(1 + \frac{P_t}{\sigma_n^2}|h|^2\right) \tag{1.1}$$

where $\frac{P_t}{\sigma_n^2}|h|^2$ is signal-to-noise ratio (SNR), $P_t$ is the transmitted power, $\sigma_n^2$ is the noise power and $|h|$ is the channel matrix (Holter 2002).

General schematic of various antenna systems is illustrated in Fig. 1.1.

Due to the rising demand of wireless data traffic, millimeter-wave frequency band (30–300 GHz) is the obvious way for 5G communication system. It suffers huge propagation loss due to the high carrier frequency at millimeter-wave operational regime and high atmospheric losses due to the oxygen and water molecules present in the atmosphere (Niu et al. 2015; Rappaport et al. 2015). To mitigate these losses, highly directional antennas and steerable narrow beam antennas are required. By using a single antenna, it is very difficult to compensate these types of losses. So instead of using a single antenna, arrays of antenna can be used to produce high directive gain and to steer the main beam (Yang 2008). At high frequencies, size




(a) SISO System (b) SIMO System

(c) MISO System (d) MIMO System

**Fig. 1.1** Illustration of various antenna systems

of the antenna will be very small, so one single antenna is not sufficient to produce high directive gain in a particular direction. Hence, to overcome disadvantages of the SISO system, MIMO antenna system is being deployed for future 5G communication networks.

Figure 1.2 shows the graphical comparison of SISO (single element), SIMO (1 × 2 array), MISO (2 × 1 array) and MIMO (2 × 2 array) antenna systems on the basis of channel capacity as well as SNR value.

Figure 1.3 shows the graphical comparison of SISO and MIMO antenna systems on the basis of channel capacity as well as SNR value. It is evident from the plot that increase in the number of radiating elements results in the enhanced channel capacity for a fixed value of SNR.

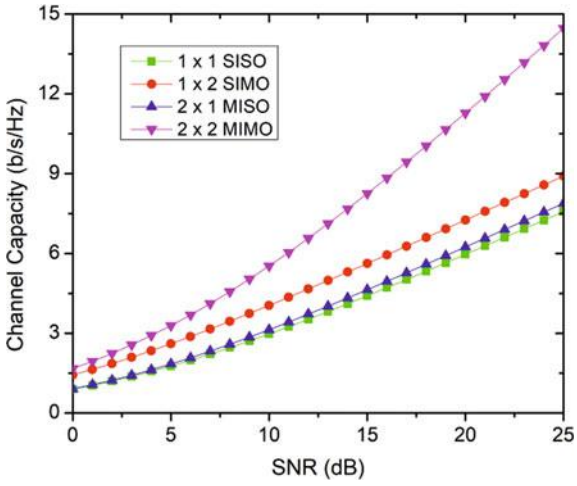



**Fig. 1.2** Channel capacity for various systems: (i) 1 × 1 SISO; (ii) 1 × 2 SIMO; (iii) 2 × 1 MISO; and (iv) 2 × 2 MIMO

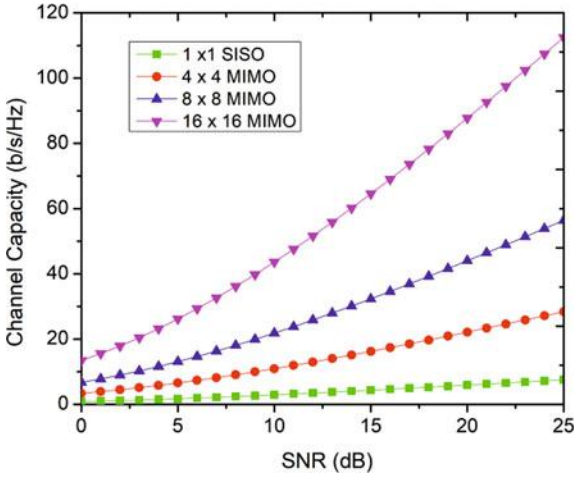



**Fig. 1.3** Channel capacity for various systems: (i) 1 × 1 SISO; (ii) 4 × 4 MIMO; (iii) 8 × 8 MIMO; and (iv) 16 × 16 MIMO

## 1.3 MIMO Antenna System

This section deals with chronological design methodology of a MIMO antenna system. Further, different antenna elements used in multiple element MIMO systems have also been briefly discussed. Figure 1.4 shows the general steps to be followed while designing a MIMO antenna system. These steps are coarsely divided into two parts: (i) physical design and (ii) performance analysis.

### *1.3.1 MIMO Antenna Design Methodology:*

1. **Frequency:**
   The very first step in the MIMO system design is to select proper frequency of operation. For the 5G communication system, antenna should operate in the complete frequency band allotted by FCC with optimum performance characteristics.
2. **Unit cell:**
   The receiving antennas will receive multiple transmitted signals via different paths, called diversity, with complex fading coefficient and noise. The path will be selected according to three ways—selection diversity method (path with maximum SNR), maximum ratio combining (MRC) method (depend on optimum linear combination of path signal) and equal gain combing (EGC) method (combine the path gain in same phase) (Brennan 1959). Transmitting antenna requires a feedback path from the receiver to the transmitter. A single transmitting antenna may be selected according to highest equivalent SNR values.
3. **Number of elements:**
   The gain and throughput of the system can be enhanced by deploying $(n \times n)$ antenna in MIMO system. The number $n$ is dependent on the above-mentioned performance parameters. On the other hand, while increasing $n$, mutual coupling between neighboring elements and bandwidth increases. Hence, one should carefully take the trade-off between these characteristics. For selecting multiple

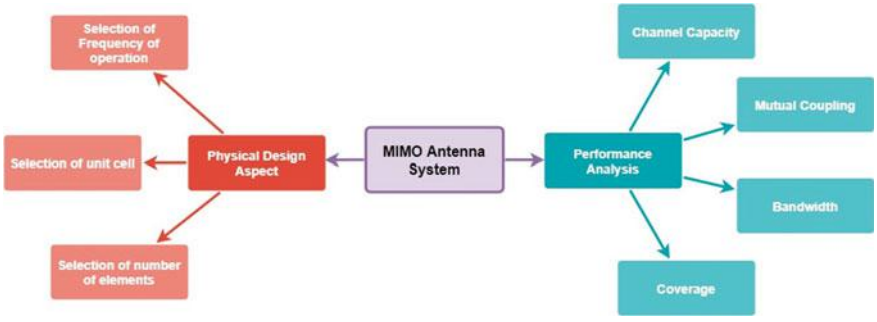



**Fig. 1.4** Design methodology of MIMO antenna systems

radiating elements at the transmitter end, single receiving antenna has to be fixed and most suitable $L_t$ number of antennas can be chosen from total $N_t$ number of antennas ($N_t > L_t$) at the transmitter end.

4. **Channel capacity:**
   Channel capacity of any communication system may be defined as its maximum attainable data rate. For an additive white Gaussian noise (AWGN) channel, it can be easily computed using (Rappaport et al. 2015):

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \tag{1.2}$$

   where $C$ is the capacity of the channel, $B$ is the bandwidth, $\frac{S}{N}$ is the signal-to-noise ratio of the system. The calculation of this parameter is important while designing the MIMO system.
5. **Mutual coupling:**
   With increase in the radiating elements to improve the capacity, neighboring elements start interfering with each other's radiation, this is known as mutual coupling between array elements. This unwanted coupling reduces the gain of antenna system and gives rise to the interference. To optimize the system performance, high port isolation is necessary between radiating elements to reduce mutual coupling.
6. **Bandwidth:**
   Bandwidth of any communication device can be defined as the largest working frequency band. Currently, FCC has allotted sub-6 GHz band and 22–28 GHz band (with center frequency as 24 GHz). The MIMO antenna system (unit cell as well as array) should perform optimally within the band. By increasing the number of radiating antenna elements to improve system throughput, bandwidth also increases. Hence, this trade-off should be estimated thoroughly.
7. **Coverage:**
   Another deciding factor for the deployment of 5G communication network is the coverage of the antenna. Basically, millimeter-wave communication is constrained by its range and obstruction penetration capabilities. As center frequency for 5G communication is around 24 GHz, the range is very small (500 m). Hence, for standalone 5G system, a large antenna array is needed for optimum performance (i.e., high data rate). System designer has to consider coverage constraints also.

### *1.3.2 Performance Enhancement of MIMO Systems*

Therefore, to meet the demand of 5G technology, one should design a MIMO antenna with high efficiency, high gain, wide bandwidth and compact size. It is very difficult to fabricate multiple number of antennas within a limited space, and it will increase

the hardware complexity and cost. So it is very important to select optimum number of radiating elements at the receiving and transmitting ends to alleviate the cost and complexity, as well as to maintain the advantages of the MIMO system. Recently, different MIMO antennas have been designed with different shapes for different frequency bands (Zhang et al. 2012; Li et al. 2016; Luo and Yang 2019; Wong et al. 2017). Generally, below 6 GHz band and mm-wave band are used for designing MIMO antenna to match the specification of 5G technology. It has been seen that in sub 6 GHz band MIMO antenna with large number of antenna elements receives higher transmission efficiency for 5G communication (Komandla et al. 2017; Li et al. 2016). But the main challenge is to mitigate the interference between radiating elements. Several techniques have been adopted to solve these issues (Chiu et al. 2007; Cihangir et al. 2014; Li et al. 2011; Zhang et al. 2012; Li et al. 2016; Su et al. 2011; Wang and Du 2013, 2014; Wong et al. 2017; Zhang et al. 2011; Zhao and Ren 2019; Zhu et al. 2009). In order to make design procedure simple, various decoupling methods are proposed to improve the isolation, such as neutralization line (Cihangir et al. 2014; Su et al. 2011; Wang and Du 2013, 2014), defected ground structure (Chiu et al. 2007; Zhu et al. 2009), etching slots (Zhang et al. 2011), connecting parasitic strips (Li et al. 2011), using self-coupling method (Zhao and Ren 2019), orthogonal polarization techniques (Zhang et al. 2012; Li et al. 2016) and asymmetrically mirror antenna (Wong et al. 2017).

## 1.4 Comparative Study of SISO, MIMO and Massive MIMO System

SISO systems utilize single radiating element at the transmitting (Tx) and receiving (Rx) ends, whereas in the MIMO system multiple number of elements are being deployed at both the ends. In comparison with SISO system, MIMO system has much better system throughput and reliability. Multiplexing gain, diversity gain and the array gain of MIMO antenna system are also much higher than SISO antenna system (Lu et al. 2014; Ordónez et al. 2011). The diversity gain of MIMO depends on the number of channels between Tx and Rx ends responsible for the diversity gain, which can be increased if more number of channels are present. On the other hand, maximum multiplexing gain of MIMO system can be attained while lesser number of antennas exist at Tx and Rx ends. Therefore, a trade-off always exists between these two gains, and it is very difficult to attain these parameters simultaneously (Ordónez et al. 2011). In massive MIMO system, larger numbers of radiating elements are used to compare conventional MIMO system (Larsson et al. 2014). The performance of the system can be evaluated under four situations—SISO antenna system, single-user MIMO (SU-MIMO) antenna system, multiuser MIMO (MU-MIMO) system and massive MIMO system. The communication model has $N_t$ number of transmitting antennas at the transmitter end and $M_r$ number of receiving antennas for each $K_u$ user at the receiver end.


**Fig. 1.5** Systematic representation of SISO systems

1. **SISO System** [$N_t = 1, M_r = 1$ **and** $K_u = 1$]:
   SISO system is comprised of single element in each transmitter and receiver module, which communicates using single channel as presented in Fig. 1.5.
   In SISO system, Tx and Rx have only one single antenna and the received signal $y$ is denoted as:

$$y = Hx + n \tag{1.3}$$

   where $H$ is the channel matrix, $x$ is the transmitting vector and $n$ is the noise vector which is additive white Gaussian noise (AWGN). Here, the noise is a complex normal distribution function denoted as $CN(0, \sigma)$, where mean is zero and standard deviation is $\sigma$. For SISO system, channel matrix will be scalar one dimensional, so the received signal ($y$) can be written as SISO channel capacity (bits/s/Hz) can be defined as (Rappaport et al. 2015):

$$C_{\text{SISO}} = \log_2(1 + \gamma) = \log_2\left(1 + h^2 \frac{P_t}{\sigma_n^2}\right) \tag{1.4}$$

   where $h$ is channel coefficient, $P_t$ is transmitted power, $\sigma_n^2$ is the noise power and $\gamma$ is the signal-to-noise ratio (SNR). So the channel capacity of a SISO system can be enhanced by increasing the SNR of the system.
2. **SU-MIMO System** [$N_t > 1, M_r > 1$, **and** $K_u = 1$]:
   SU-MIMO system is comprised of multiple elements in both transmitter and receiver modules, which communicate using multiple parallel channels with single user as represented in Fig. 1.6.

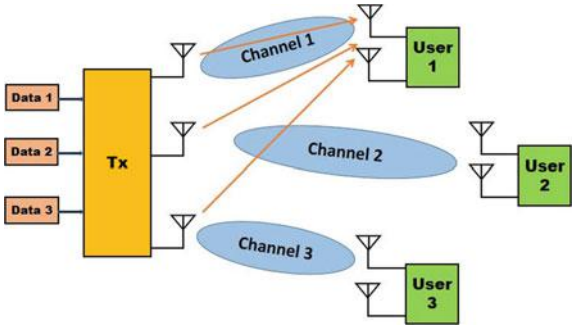



**Fig. 1.6** Systematic representation of SU-MIMO systems

In MIMO antenna system, the spectral efficiency and data rate are increased significantly without affecting the SNR or bandwidth due to the multiple number of antennas present at both Tx and Rx sides. The additional channel capacity of a MIMO system is generated by the reason of spatial multiplexing through multi-channel transmission from multiple radiating elements (Mattigiri and Warty 2013). The additional capacity of the channel comes at the expense of space (generally for mobile handset), cost for the deployment of multiple radiating elements and signal processing complexity (Goldsmith 2005). In SU-MIMO system, only one user will be served among $K$ number of users within a single transmission time interval and the received signal vector can be described as

$$y_m = \sqrt{\rho} H_{mn} x_n + n_m \tag{1.5}$$

where $x_n$ is transmit signal vector ($x_n \in C^{N_t \times 1}$), $n_m$ is the noise vector ($n_m \in C^{M_r \times 1}$) and n is the transmitting antennas ($n = 1, 2, 3 \ldots N_t$), $m$ is the receiving antennas ($m = 1, 2, 3 \ldots M_r$). $H_{mn}$ is the channel matrix, $\rho$ represents the normalized transmit power and the total power of the system is unity ($E \|x_n\|^2 = 1$). The channel capacity (bits/s/Hz) of a SU-MIMO system can be written as

$$C_{\text{SU-MISO}} = \log_2 \left| I + \frac{\rho}{N_t} HH^* \right| \tag{1.6}$$

Here $I$ is the identity matrix and the channel capacity of SU-MIMO system is bounded by:

$$\log_2(1 + \rho M_r) \leq C_{\text{SU-MISO}} \tag{1.7}$$

$$\log_2(1 + \rho M_r) \leq \min(N_t, M_r) \log_2 \left( 1 + \frac{\rho \max(N_t, M_r)}{N_t} \right) \tag{1.8}$$

3. **MU-MIMO System** [$N_t > 1, M_r > 1, M_{Ku} = 1$ **and** $K_u = 1$]:
MU-MIMO system is comprised of multiple elements in both transmitter and receiver modules, which communicate using multiple channels allotted to each user as illustrated in Fig. 1.7.
MU-MIMO has many advantages over the SU-MIMO system, such as (1) it attained multiuser diversity in spatial domain by managing the total resources to the multiple users for spatially correlated channels (Li et al. 2010); (2) MU-MIMO transmitting antennas can transmit the signal to the many users at the same time, and therefore, one single antenna can be used at the user end to reduce the cost of the system; (3) it is comparatively less sensitive for propagation (Kim and Malladi 2012). The received signal vector ($y_k \in C^{K_u \times 1}$) of a MU-MIMO system may be expressed as:

$$y_k = \sqrt{\rho} H_{k,n} x_n + n_k \tag{1.9}$$

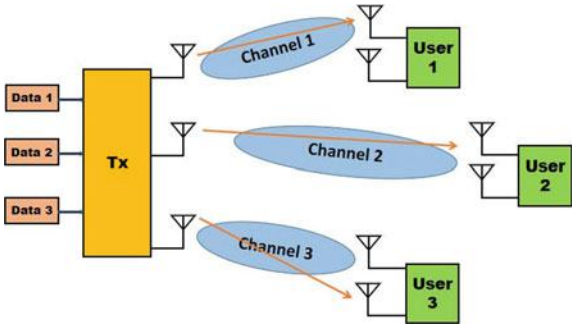



**Fig. 1.7** Systematic representation of MU-MIMO systems

where $H_{k,n}$ is the channel matrix ($H_{k,n} \in C^{K_u \times N_t}$), $x_n$ is the transmit signal vector ($x_n \in C^{N_t \times 1}$) and $n_k$ is the noise vector ($n_k \in C^{K_u \times 1}$) of MU-MIMO system. The capacity of a MU-MIMO system can be represented as:

$$C_{\text{MU-MISO}} = \max \log_2 |I + \rho HPH^*| \tag{1.10}$$

where $P$ is a diagonal matrix ($P = p_1, p_2, \ldots p_k$) with power allocation. In recent years, various MU-MIMO antenna systems have been developed because of its better system performance in terms of capacity and reliability (Li et al. 2010).

4. **Massive MIMO System** [$N_t \gg M_r$ and $N_t \to \infty$ or $M_r \gg N_t$ **and** $M_r \to \infty$]: Massive MIMO system is similar to MU-MIMO, but the number of radiation elements is very large in either transmitter or receiver module as illustrated in Fig. 1.8.
   In massive MIMO system, the spectral efficiency is enormously increased due to the presence of huge number of antennas (Larsson et al. 2014; Lu et al. 2014). For the first condition ($N_t \gg M_r$ and $N_t \to \infty$), the channel capacity of massive MIMO system can be written as

$$C_{\text{massive MIMO}} \approx M_r \log_2 (1 + \rho) \tag{1.11}$$

And for the second condition ($M_r \gg N_t$ and $M_r \to \infty$), the channel capacity becomes

$$C_{\text{massive MIMO}} \approx N_t \log_2 \left(1 + \frac{\rho M_r}{N_t}\right) \tag{1.12}$$

From these two equations, it is observed that channel capacity of a massive system increased linearly with the increment of number of antennas at both Tx and Rx ends.

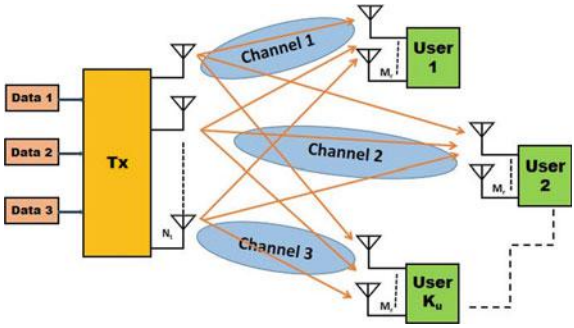



**Fig. 1.8** Systematic representation of Massive MIMO systems

## 1.5 mm-Wave MIMO

Till date, major civilian communication systems work below 30 GHz frequency including recent research on 5G band with center frequency around 28 GHz. Frequencies higher than 30 GHz are mainly exploited by military and space science application (Rappaport et al. 2017). In recent years, very popular mm-wave bands (30–300 GHz) open up new frontiers for 5G mobile communication because of its wide bandwidth which meets the demand of wireless data traffic (Elkashlan et al. 2014). It is a conception regarding using this mm-wave frequency for communication that the attenuation of signal increases adequately over frequency. But it is evident from many literatures available on behavioral analysis of mm-waves in different environments that if we consider the constant gain antenna (obviously with respect to frequency), then only Friis equation is valid. So mm-wave massive MIMO is one of the key technologies for future mobile communication (Mumtaz et al. 2016). Antenna arrays with different beamforming techniques are the heart of this technology. In mm-wave bands, the size of the antenna elements will be very small; therefore, it will enable massive MIMO antenna technology as large number of antennas can be incorporated at both Tx and Rx terminals (Swindlehurst et al. 2014). But the designing of antenna array in mm-wave bands is a very complicated job as it is very difficult to fabricate large number of antenna elements within a limited space. There are several other design challenges which include:

1. The compact arrangement of antenna elements causes mutual coupling between antenna elements which will degrade the system performance. Different approaches are already developed to minimize the mutual coupling (Chiu et al. 2007; Cihangir et al. 2014; Li et al. 2011; Zhang et al. 2012; Li et al. 2016; Su et al. 2011; Wang and Du 2013, 2014; Wong et al. 2017; Zhang et al. 2011; Zhao and Ren 2019; Zhu et al. 2009).
2. Inter-element spacing should be very small to attain the required scanning performance; otherwise, grating lobe may appear in the undesired direction (Niu et al. 2015).
3. In high frequencies, efficiency and sensitivity of the system reduced significantly due to high losses in the feed line connected to each antenna element and phase

shifter. So different feeding techniques have been proposed to minimize these types of losses.
4. In mm-wave, due to high propagation loss and different atmospheric losses, high gain antenna arrays are required. But with the increment in directivity, the beamwidth of the array gets narrower; therefore, antenna array beamforming technology should be adopted (Niu et al. 2015).

The challenges of mm-wave MIMO in terms of communication aspects are channel modeling. There are many channel models available for lower frequency spectrum such as COST and 3GPP models etc, but these are unable to exploit complete characteristics of mm-wave band. Covering complete spectrum of mm-wave is the key challenge in channel modeling (Hemadeh et al. 2018; Rappaport et al. 2017).

1. Unavailability of measured propagation characteristic in mm-wave region.
2. Pre-existed models work on a narrow band ($\approx$ 100MHz) and this cannot be directly scaled to mm-wave spectrum because of their wider bandwidth (Hemadeh et al. 2018).
3. Due to the multi-path channel capability, closely spaced users can receive same signal from same path and it should be incorporated while modeling the mm-wave channel (Adhikary et al. 2014).

In this mm-wave spectrum, MIMO elements and channels will be huge in terms of quantity. Hence, it is also described as mm-wave massive MIMO. Basically, at such high frequency, antenna element size is very small, and in the array form, separation between radiating elements should be at least 0.5λ. The formation can be of three major categories: (1) uniform linear, (2) uniform planar and (3) uniform circular array. These formations are shown in Fig. 1.9.

In current 5G communication scenario, uniform planar antenna array is being investigated thoroughly because this formation exhibits higher gain with proper space utilization. Array factor of the uniform planar antenna array can be expressed as:

$$AF(\theta, \phi) = \left[1, ..., e^{j[m_X \phi_X + n_Y \phi_Y]}, ...\right]^{\mathrm{T}}, \tag{1.13}$$

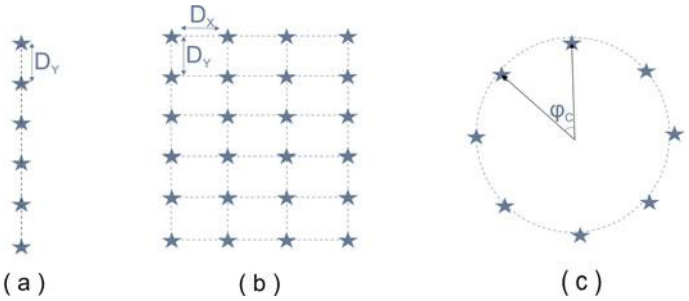



**Fig. 1.9** Different array formations: **a** uniform linear; **b** uniform planar; **c** uniform circular

where

$$\begin{cases} \phi_X = \frac{2\pi}{\lambda} D_X sin(\theta)\cos(\varphi) \\ \phi_Y = \frac{2\pi}{\lambda} D_Y sin(\theta)\sin(\varphi) \end{cases}$$

where total number of antenna elements, $N = N_X M_Y$, with indices $m_X$ and $n_Y$ be $1 \leq m_X \leq N_X - 1$ and $1 \leq n_Y \leq N_Y - 1$. $D_X$ and $D_Y$ are the separation between two elements in $x$- and $y$-axis, respectively.

## 1.6 Antenna Array Beamforming:

The throughput of a MIMO system can be improved by different beamforming techniques. The beamforming is a technique that can be performed by regulating the phase of the transmitted signal from each radiating element to obtain the main beam in particular direction (Gao et al. 2017; Roh et al. 2014). The architecture of beamforming can be categorized into three techniques: analog beamforming technique (performed in analog domain), digital beamforming technique (performed in discrete domain) and hybrid beamforming technique (performed in both analog and digital domain) (Kutty and Sen 2015). The first method can be applicable for single-user system only, whereas remaining two techniques can be employed for both, single and multiuser systems (Gao et al. 2017).

1. **Analog beamforming:**
   In analog beamforming, the RF front end of the antenna array is analog baseband and only the phase of the signal with single data stream is generally used to control the optimum gain and SNR. In this method, the beam can be formed using a single RF chain at both Tx and Rx terminals based on the available channel as presented in Fig. 1.10. Not only beamforming, beam steering is also an important factor for antenna array. The analog beam-former vector f and analog combiner vector w can be selected according to the following condition (Busari et al. 2017):

$$\left(w_i^{\text{opt}}, f^{\text{opt}} = \arg\max |w * Hf_l|^2\right) \tag{1.14}$$

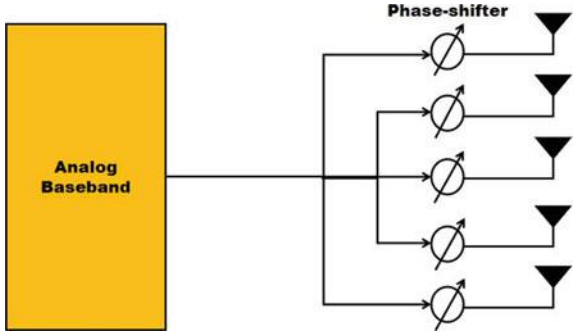



**Fig. 1.10** Analog beamforming architecture

where $w_i = \sqrt{N_t^{-1}} \exp j\phi_i$ for all $i$, and $f_l = \sqrt{M_r^{-1}} \exp j\phi_l$ for all $l$. But in systems where a large number of radiating elements are used (i.e., MIMO system), it is very difficult to find perfect channel state information (CSI). Thus, to improve the performance of MIMO system, beam training is also necessary, where Tx and Rx jointly selected the pair of best beamforming antennas from predefined codebook (Gao et al. 2017). However, in massive MIMO system, the codebook size is huge; thus, different beam training approaches were developed to optimize the overhead. Wang et al. (Wang et al. 200) developed a beamforming protocol based on discrete phase-shift codebook for high-speed mm-wave communication. Although the hardware cost for analog beamforming method is very less due to the single RF chain, the performance of the system will be degraded as only phase of transmitted signal can be controlled using this method. Apart from this, analog beamforming method is also not useful for multiuser system, so it is insignificant for massive MIMO technology (Busari et al. 2017).

2. **Digital beamforming:**
   In digital beamforming technique, the phase and amplitude of the transmitted signal can be regulated for single-user as well as multiuser system. The basic architecture of a digital beamforming system is illustrated in Fig. 1.11. For single user, $N_t$ transmitting antenna can transmit multiple data streams to the receiver connected with $M_r$ receiving antenna ($N_t \times M_r$). In this method, D numbers of digital precoder and $N_t$ RF chain are used for data transmission. On the other hand, for multiuser system, $N_t$ transmitting antenna can transmit multiple data streams to $K_u$ number of users each having $M_r$ receiving antenna, and the size of RF chain will be $N_t$ and digital precoder size will be ($N_t \times M_r K_u$). such that $N_t = M_r K_u$. The $k$th user has ($N_t \times M_r$) digital precoder $D_k = [D_1, D_2, D_3 \ldots D_{K_u}]$; then, the signal received by $k$th user can be written as (Busari et al. 2017)

$$y_k = H_k \sum_{n=1}^{K_u} D_n x_n + n_k \tag{1.15}$$

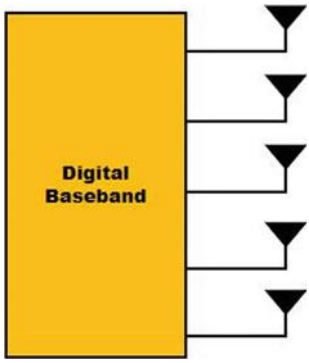



**Fig. 1.11** Digital beamforming architecture

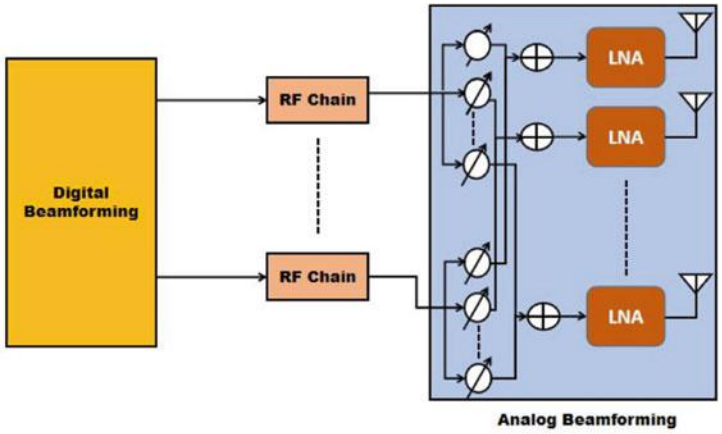
where $x_n$ is the transmitted signal vector before precoding, $D_n$ is denoted as precoder, $H_k$ is the channel matrix of size $(N_t \times M_r)$ and $n_k$ is noise vector. To avoid interference at the receiving terminal precoder, $D_n$ may be selected so that it can assure the condition $(H_k D_n = 0)$ is satisfied (Shen et al. 2006). As digital beamforming technique can be capable of controlling both phase and amplitude of the transmitted signal, so it has better system performance compared to analog beamforming technique. The main challenge of digital beamforming technique is the requirement of separate RF chain for each radiating element. Hence, the hardware cost and power utilization of the system will be very high, so it will not be useful for mm-wave massive MIMO systems.

3. **Hybrid beamforming:** This technique is one of the promising techniques which can enhance the system performance compared to analog beamforming method as well as reduce the complexity of digital beamforming technique (Alkhateeb et al. 2015; Kutty and Sen 2015; Molisch et al. 2017; Zhang et al. 2017). In hybrid beamforming method, the optimal system performance may be attained by reducing the number of RF chains, so the cost and power consumption of the system will be reduced significantly (Zhang et al. 2011). For mm-wave MIMO system, hybrid beamforming technique can be realized using two stages: In first stage, digital precoders are generally used with minimum number of RF chains which reduce the coupling, and in the second stage, analog beamformer is used with large number of phase shifters (no RF chain) to enhance the antenna gain (Alkhateeb et al. 2015).

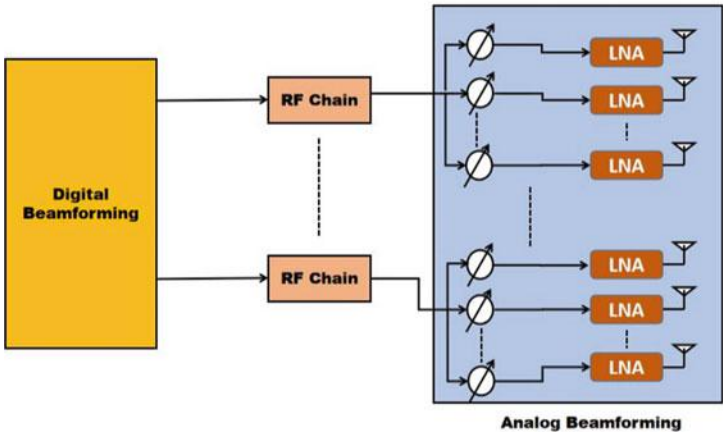
   The performance of hybrid precoder can be measured for single-user as well as multiuser system. For single-user system, it is categorized into two distinct groups: (i) spatially sparse architecture, where all the transmitting antennas are associated with each RF chain through phase shifters as illustrated in Fig. 1.12a, and (ii) successive interference cancellation (SIC) architecture, where the group of transmitting radiating element is associated with each RF chain as represented in Fig. 1.12b. For multiuser system, the combination of analog and digital beamforming technique has been adopted (Molisch et al. 2017). In this type of system, $N_t$ transmitting antennas can transmit the multistream data through $N_{tRF}$ RF chain in such a way that the condition $N_{tRF} \leq N_t$ is maintained, and each $K_u$ user is connected with $M_r$ receiving antennas along a single RF chain. The received signal vector of a multiuser hybrid precoder can be expressed as (Busari et al. 2017):

$$r_k = H_k \sum_{n=1}^{K_u} A_n D_n s_n + n_k \tag{1.16}$$

   where digital precoder $D_k = [D_1, D_2, D_3 \ldots D_{K_u}]$ is followed by analog precoder $A_k = [A_1, A_2, A_3 \ldots A_{K_u}]$. After multiplying with analog combiner $(w_k)$, the received signal $(y_k)$ can be written as (Busari et al. 2017):




(a) Fully-connected hybrid beamforming architecture




(b) Sub-connected hybrid beamforming architecture

**Fig. 1.12** Beamforming architecture for mm-wave MIMO system

$$y_k = w_k^H r_k = w_k^H H_k \sum_{n=1}^{K_u} A_n D_n s_n + w_k^H n_k \tag{1.17}$$

From the comparative analysis of these three beamforming techniques, it is found that the performance of hybrid beamforming method is superior for mm-wave massive MIMO system (Molisch et al. 2017; Zhang et al. 2017).

## 1.7 Conclusion

This chapter gives the basic idea about different MIMO antenna systems for futuristic 5G communication. Firstly, single element and multiple element antenna systems have been compared on the basis of their respective performance characteristics. It is evident from the analysis that performance of the system is enhanced with increase in number of radiating elements. The design methodology of the MIMO antenna system has been discussed on the basis of physical design as well as performance analysis followed by various techniques of system performance enhancement. A comparative investigation of SISO and MIMO antenna systems has been presented with respect to their architecture and channel capacity. The mm-wave MIMO system has been discussed as an evolving and more efficient technique to be used in future 5G communication systems. Antenna array beamforming methods are very useful to enhance the system throughput with minimum mutual coupling. These methods have been proved to be effective in both single and multiple user scenarios.

## References

Abdullah M, Kiani SH, Iqbal A (2019) Eight element multiple-input multiple-output (MIMO) antenna for 5G mobile applications. IEEE Access 7:134488–134495

Adhikary A, Al Safadi E, Samimi MK, Wang R, Caire G, Rappaport TS, Molisch AF (2014) Joint spatial division and multiplexing for mm-wave channels. IEEE J Sel Areas Commun 32(6): 1239–1255

Alkhateeb A, Leus G, Heath RW (2015) Limited feedback hybrid precoding for multi-user millimeter wave systems. IEEE Trans Wireless Commun 14(11):6481–6494

Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong AC, Zhang JC (2014) What will 5G be? IEEE J Sel Areas Commun 32(6):1065–1082

Brennan DG (1959) Linear diversity combining techniques. Proc IRE 47(6):1075–1102

Busari SA, Huq KMS, Mumtaz S, Dai L, Rodriguez J (2017) Millimeter-wave massive MIMO communication for future wireless systems: A survey. IEEE Commun Surveys Tutorials 20(2): 836–869

Chiu C-Y, Cheng C-H, Murch RD, Rowell CR (2007) Reduction of mutual coupling between closely-packed antenna elements. IEEE Trans Antennas Propag 55(6):1732–1738

Cihangir A, Ferrero F, Jacquemod G, Brachat P, Luxey C (2014) Neutralized coupling elements for mimo operation in 4G mobile terminals. IEEE Antennas Wireless Propag Lett 13:141–144

Elkashlan M, Duong TQ, Chen H-H (2014) Millimeter-wave communications for 5G: Fundamentals: Part i [guest editorial]. IEEE Commun Mag 52(9):52–54

Gao X, Dai L, Gao Z, Xie T, Wang Z (2017) Precoding for mm wave massive MIMO. In: mmWave Massive MIMO, pp 79–111. Elsevier

Goldsmith A (2005) Wireless Communications. Cambridge University Press

Hemadeh IA, Satyanarayana K, El-Hajjar M, Hanzo L (2018) Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget. IEEE Commun Surveys Tutorials 20(2):870–913

Holter B (2002) Capacity of multiple-input multiple-output (MIMO) systems in wireless communications. Norwegian University of Science and Technology, Department of Telecommunications

Kim BH, Malladi DP (2012) Approach to a unified SU-MIMO/MU-MIMO operation, 18 Sept 2012. US Patent 8,271,043

Komandla MV, Mishra G, Sharma SK (2017) Investigations on dual slant polarized cavity-backed massive MIMO antenna panel with beamforming. IEEE Trans Antennas Propag 65(12): 6794–6799

Kutty S, Sen D (2015) Beamforming for millimeter wave communications: An inclusive survey. IEEE Commun Surveys Tutorials 18(2):949–973

Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2):186–195

Li Q, Li G, Lee W, Lee M-I, Mazzarese D, Clerckx B, Li Z (2010) MIMO techniques in WiMAX and LTE: A feature overview. IEEE Commun Mag 48(5):86–92

Li Z, Du Z, Takahashi M, Saito K, Ito K (2011) Reducing mutual coupling of MIMO antennas with parasitic elements for mobile terminals. IEEE Trans Antennas Propag 60(2):473–481

Li Y, Zhang Z, Zheng J, Feng Z (2012) Compact azimuthal omnidirectional dual-polarized antenna using highly isolated colocated slots. IEEE Trans Antennas Propag 60(9):4037–4045

Li M-Y, Ban Y-L, Xu Z-Q, Wu G, Kang K, Yu Z-F et al (2016) Eight-port orthogonally dual-polarized antenna array for 5G smartphone applications. IEEE Trans Antennas Propag 64(9):3820–3830

Li Y, Luo Y, Yang G et al (2019) High-isolation 3.5 GHz eight-antenna mimo array using balanced open-slot antenna element for 5G smartphones. IEEE Trans Antennas Propag 67(6):3820–3830

Lu H-Y, Fang W-H (2007) Joint transmit/receive antenna selection in MIMO systems based on the priority-based genetic algorithm. IEEE Antennas Wireless Propag Lett 6:588–591

Lu L, Li GY, Swindlehurst AL, Ashikhmin A, Zhang R (2014) An overview of massive MIMO: benefits and challenges. IEEE J Sel Top Signal Process 8(5):742–758

Marcus MJ (2015) 5G and "IMT for 2020 and beyond" [spectrum policy and regulatory issues]. IEEE Wireless Commun 22(4):2–3

Mattigiri S, Warty C (2013) A study of fundamental limitations of small antennas: MIMO approach. In: 2013 IEEE aerospace conference, pp 1–8. IEEE

Molisch AF, Win MZ (2004) MIMO systems with antenna selection. IEEE Microwave Mag 5(1): 46–56

Molisch AF, Ratnam VV, Han S, Li Z, Nguyen SLH, Li L, Haneda K (2017) Hybrid beamforming for massive MIMO: A survey. IEEE Commun Mag 55(9):134–141

Mumtaz S, Rodriguez J, Dai L (2016) MmWave massive MIMO: a paradigm for 5G. Academic Press

Niu Y, Li Y, Jin D, Su L, Vasilakos AV (2015) A survey of millimeter wave communications (mmwave) for 5G: Opportunities and challenges. Wireless Netw 21(8):2657–2676

Ordónez LG, Palomar DP, Fonollosa JR (2011) Fundamental diversity, multiplexing, and array gain tradeoff under different mimo channel models. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)

Rappaport TS, Heath RW Jr, Daniels RC, Murdock JN (2015) Millimeter wave wireless communications. Pearson Education

Rappaport TS, Xing Y, MacCartney GR, Molisch AF, Mellios E, Zhang J (2017) Overview of millimeter wave communications for fifth-generation (5G) wireless networks-with a focus on propagation models. IEEE Trans Antennas Propag 65(12):6213–6230

Roh W, Seol J-Y, Park J, Lee B, Lee J, Kim Y, Cho J, Cheun K, Aryanfar F (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results. IEEE Commun Mag 52(2):106–113

Sanchez-Fernandez M, Rajo-Iglesias E, Quevedo-Teruel O, Pablo-Gonzalez ML (2008) Spectral efficiency in MIMO systems using space and pattern diversities under compactness constraints. IEEE Trans Veh Technol 57(3):1637–1645

Shen Z, Chen R, Andrews JG, Heath RW, Evans BL (2006) Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. IEEE Trans Signal Proces 54(9):3658–3663

Su S-W, Lee C-T, Chang F-S (2011) Printed MIMO-antenna system using neutralization-line technique for wireless usb-dongle applications. IEEE Trans Antennas Propag 60(2):456–463

Swindlehurst AL, Ayanoglu E, Heydari P, Capolino F (2014) Millimeter-wave massive MIMO: The next wireless revolution? IEEE Commun Mag 52(9):56–62

Wang Y, Du Z (2013) A wideband printed dual-antenna system with a novel neutralization line for mobile terminals. IEEE Antennas Wireless Propag Lett 12:1428–1431

Wang S, Du Z (2014) Decoupled dual-antenna system using crossed neutralization lines for LTE/WWAN smartphone applications. IEEE Antennas Wireless Propag Lett 14:523–526

Wang J, Lan Z, Pyo C-W, Baykas T, Sum C-S, Rahman MA, Gao J, Funada R, Kojima F, Harada H et al (2009) Beam codebook based beamforming protocol for multi-gbps millimeter-wave WPAN systems. IEEE J Sel Areas Commun 27(8):1390–1399

Wong K-L, Tsai C-Y, Lu J-Y (2017) Two asymmetrically mirrored gap-coupled loop antennas as a compact building block for eight-antenna MIMO array in the future smartphone. IEEE Trans Antennas Propag 65(4):1765–1778

Yang H, Herben MH, Akkermans IJ, Smulders PF (2008) Impact analysis of directional antennas and multiantenna beamformers on radio transmission. IEEE Trans Veh Technol 57(3):1695–1707

Zhang S, Lau BK, Tan Y, Ying Z, He S (2011) Mutual coupling reduction of two pifas with a t-shape slot impedance transformer for MIMO mobile terminals. IEEE Trans Antennas Propag 60(3):1521–1531

Zhang D, Wang Y, Li X, Xiang W (2017) Hybridly connected structure for hybrid beamforming in mmwave massive MIMO systems. IEEE Trans Commun 66(2):662–674

Zhao A, Ren Z (2019) Multiple-input and multiple-output antenna system with self-isolated antenna element for fifth-generation mobile terminals. Microwave Opt Technol Lett 61(1):20–27

Zhu F-G, Xu J-D, Xu Q (2009) Reduction of mutual coupling between closely-packed antenna elements using defected ground structure. Electron Lett 45(12):601–602

# Chapter 2
# Pilot Contamination in Massive MIMO Communications


**Abhinaba Dey, Prabina Pattanayak, and Devendra Singh Gurjar**

**Abstract** Massive multiple-input multiple-output (MIMO) has been considered as one of the most promising technologies for addressing high data rates and capacity for cellular networks in the fifth generation and beyond. In massive MIMO, the base station (BS) requires complete and accurate channel state information (CSI) for realizing the benefits of massive MIMO technology. The CSI can be obtained either through feedback from users or channel reciprocity technique. The use of non-orthogonal pilot sequences, sent by the users for determining CSI, has been the main source of pilot contamination in the uplink data communications. Various pilot contamination scenarios and their possible solutions have been extensively studied in the last few years. In this chapter, recent works in the area of mitigating the effects of pilot contamination have been presented and analyzed. Some of the higher throughput resulting algorithms and their capabilities in eliminating the effects of pilot contamination have been studied in detail.

**Keywords** Massive MIMO · Channel state information · Time/frequency division duplex · Mutually orthogonal pilot sequences · Signal to interference plus noise ratio (SINR) · Spectral efficiency

A. Dey · P. Pattanayak (✉) · D. Singh Gurjar
Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India
e-mail: prabina.pattanayak@ieee.org

A. Dey
e-mail: abhinabadey@ieee.org

D. Singh Gurjar
e-mail: devendra.gurjar@ieee.org



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_2

## 2.1 Introduction

As the world is moving toward fifth generation, the demand for increased data rate, connectivity and enhanced network services call for the development of new advanced technologies as the existing technologies are unable to meet these requirements (Elijah et al. 2016). It is being calculated that data rate and mobile traffic demands are going to increase thousand folds in the coming decade. The research community is of the opinion that there is an urgent requirement of developing newer and more advanced cellular network technologies (Li et al. 2014) to achieve this target. Many technologies like multiple-input multiple-output (MIMO), massive MIMO (Larsson et al. 2014), device-to-device communications, machine-to-machine communications, millimeter (mm) wave communications, etc. are being introduced in the 5G cellular network deployments. These technologies possess the capability and are expected to meet the increased demands of data rate in the upcoming years (Hossain and Hasan 2015).

The available fourth-generation technologies such as the third- generation partnership project (3GPP) LTE-Advanced as well as IEEE 802.16m have the capacity to achieve a peak spectral efficiency of around 15 bps/Hz, cell average spectral efficiency of around 2 bps/Hz and effective bandwidth of 100 MHz (Prasad et al. 2013). But the projected data usage and data rate growth in 5G are far beyond the capabilities of existing 4G technologies. Combining the researches done till now on the advancements to be perceived by the upcoming technologies, eight main technology requirements can be summarized as follows,

- Very high data rate in the real-time networks: The projected data rate of 1–10 Gbps is almost a 10 times increase in data rate as compared to the 4G LTE networks having a 150 Mbps peak data rate.
- Low round trip latency: It will have a round trip latency of around 1 ms which is a tenfold reduction from traditional 4G technologies having 10 ms.
- Unit area bandwidth requirement:A lot of devices will be connected in a specified area with large bandwidth requirements for a longer time period.
- Huge number of connected devices: The newer technologies must have the capability to handle thousands of connected devices at a time.
- Very high perceived availability: The vision of 5G is to make the network available to the connected devices almost all the times (theoretically 99.999%).
- Full coverage irrespective of location: 5G technologies plans to provide full coverage to every connected devices all the time irrespective of the device locations.
- Energy usage reduction: Due to huge increase in data rate and connectivity requirements, high energy consumption comes into picture and researches are done toward reducing the same. The advent of green communication opened up a way of reducing the energy usage by almost 90%.
- Higher battery life: Due to reduced power consumption by the connected devices with the aid of green technologies, the battery life of the devices has increased quite a bit.

Hence, the need for newer technologies came into picture. 5G is expected to come forth with many new technologies including massive MIMO, D2D communication, mm wave technology and many more to meet the increased demands (Chávez-Santiago et al. 2015).

Massive MIMO is a type of communication system wherein the base station (BS) has large antenna arrays in the order of hundreds which serves users in the order of tens in the same time-frequency resource (Larsson et al. 2014). The advantage of such system is that the BS with multiple antennas transmit independent data streams to multiple users in the same time-frequency resource. This helps in achieving higher data rates. Also, use of massive MIMO enhances link reliability, increases coverage area and improves energy efficiency. Extensive research works have been discussed in the literature for achieving the spatial multiplexing in various single carrier MIMO communication systems (Chu et al. 2014; Min et al. 2013; Pattanayak and Kumar 2016; Pattanayak et al. 2015; Wang et al. 2013; Xu et al. 2013). Orthogonal frequency-division multiplexing (OFDM) is generally used along with MIMO for diminishing the impairments of wide-band MIMO channels (Stuber et al. 2004). High throughput resulting schemes have also been discussed for MIMO-OFDM systems in the literature (Chen et al. 2007; Eslami and Krzymień 2009; Eslami and Krzymien 2011; Fakhereddin et al. 2009; Jorswieck et al. 2008; Pattanayak and Kumar 2017, 2018, 2019, 2016; Pattanayak et al. 2017; Svedman et al. 2007). Moreover, soft computing techniques have also been explored for achieving higher system sum-rate with low computational complexity (Naeem and Lee 2014; Pattanayak and Kumar 2015, 2020). Massive MIMO has gained the momentum in last decade where the advantages of MIMO have been increased by many folds.

Some of the basic advantages provided by massive MIMO includes:

- Spatial Multiplexing Gain: Sending independent data streams to users by utilizing the large antenna array at the BS helps in increasing the data rate by more than tenfold (Larsson et al. 2014).
- Energy Efficiency: The large antenna array at The BS helps in significantly reducing the uplink (UL) and downlink (DL) transmission power due to the coherent combining. It has been found that UL transmit power can be reduced inversely as the number of BS antennas (Li et al. 2014; Ngo et al. 2013).
- Increased Reliability: The large antenna array at the BS allows for higher diversity gains, which, in turn, decreases error reception and increases the reliability. It has also been shown that large BS antenna arrays diminish the effects of uncorrelated noise and intra-cellular interference as the number of BS antennas grow without limit (Marzetta 2010; Rusek et al. 2013).
- Cost Reduction: As the overall power consumption is reduced due to large BS antenna array, hence it allows the deployment of low-cost RF amplifiers, which reduces the overall cost (Rusek et al. 2013).

Massive MIMO, although promising, does have certain practical implementation challenges such as channel estimation, pilot designing and scheduling, antenna calibration, etc (Kurras 2013). In order to fully utilize the benefits of massive MIMO, the BS needs to have the accurate estimation of the channel between itself and the

served users (Adhikary et al. 2013). This can be done either by channel reciprocity or feedback schemes (Ashikhmin and Marzetta 2012; Yin et al. 2013). Any cellular network is either time-division duplex (TDD) or frequency-division duplex (FDD). TDD is considered more advantageous then FDD because the estimation needs to be done in one direction only (either uplink or downlink) and used for both the directions using channel reciprocity. On the other hand, FDD requires estimation in both directions along with simultaneous feedback (Larsson et al. 2014; Lu et al. 2014; Rusek et al. 2013). As a result, TDD has been quoted as a better option for massive MIMO due to efficient use of radio resources (Osseiran et al. 2014).

Having the knowledge of entire channel matrix is advantageous for transmit and receive signal processing. A quantity known as Frobenius-squared norm or FSN of the channel matrix can be used for getting acquainted with channel gain and can be used for scheduling and rate adaptation. This quantity (FSN) is estimated using the channel matrix estimation or a minimum mean-squared error (mmse) estimator. In case of Rayleigh or Ricean fading channels, however, the former method faces a limitation for various SINR values as compared to the latter.

In TDD, channel reciprocity and pilot usage for channel estimation are the key features (Appaiah et al. 2010). Applying reciprocity, the channel needs to be estimated in the uplink only after which the downlink channel is assumed to be the transpose of the uplink channel (Marzetta 2006; Wang et al. 2013). This estimation is done by sending pilots from UTs to the BS. But in practical scenarios, an antenna calibration scheme is required either at the transmitter or receiver side due to difference in the characteristics of transmitter/receiver RF chains (Hou et al. 2012; Medard 2000). In most of the works done till now, the number of different pilot sequences has been considered to be equal to the number of users. Also, it is assumed that the pilots used by all the users are of the same size (Marzetta 2006).

In case of cellular networks and other wireless communication systems, spectral efficiency has been considered to be an important performance metric for enhancing the overall throughput (Jose et al. 2011). However, in a massive MIMO TDD system, the pilots sent for channel estimation get corrupted due to reuse of non-orthogonal sequences in other cells in a multi-cellular system (Peng Xu et al. 2013). As a result, it causes inter-cell interference and hence limits the spectral efficiency of the system (Boccardi et al. 2014; Mueller et al. 2013; Zhang and Zeng 2014).

This problem of pilot contamination can be reduced by coordination among the various cells present in the cellular system, by making use of a low-rate communication during channel estimation phase. This coordination makes use of the statistical second-order information of the participating user channels which creates a selectivity among the interfering users having strongly correlated pilot sequences also. As a result of this, the massive MIMO systems tend to minimize or even remove the pilot contamination completely for specific channel covariances.

Several works have been done on eliminating the effects of pilot contamination and improving the spectral efficiency. It has been found that in the limiting case of massive MIMO where the number of BS antennas tends to infinity, although the intra-cell interference and uncorrelated noise effects vanish (Gopalakrishnan and Jindal 2011; Nam et al. 2012; Ngo et al. 2011), the effect of inter-cell interference still looms

**Table 2.1** Notations

| Notation | Description |
|---|---|
| $I_N$ | Identity matrix of size $N$ |
| $(-)^{\mathrm{H}}$ | Hermitian of a matrix |
| $(-)^{\mathrm{H}}$ | Transpose of a matrix |
| $E\|\|-\|\|$ | Expected value of a variable |
| tr(-) | Trace of a matrix |
| var(-) | Variance of a variable |
| $\|\|-\|\|$ | Double norm |

large due to pilot reuse (Fernandes et al. 2012, 2013; Marzetta 2010). Many studies have been conducted for reducing the effects of inter-cellular interference on channel estimation. In most of these studies, it has been assumed that the CSI is available at the BS. However, when practically implemented, it has been found that estimation of the channel is of utmost importance. For the assumption where we have unlimited antenna at the BS, it has been found that all kinds of interferences do not vanish because of reusing the orthogonal pilot sequences in adjacent cells. This invoked several studies where the authors have developed various methods of mitigating this interference. Although most of them have quoted the inter-cell interference as the only source of pilot contamination, later studies have found many other reasons of the same.

As we are well aware of the importance of pilot sequences in massive MIMO scenario, hence we analyze here some of the pilot scheduling algorithms developed by various researchers that has helped in mitigating the effects of inter-cell interference to satisfactory levels (Jin et al. 2015; Yue et al. 2015; Zhou and Wang 2016; Zhu et al. 2015).

Table 2.1 shows the list of variables and operators that will be used in the rest of the chapter.

## 2.2 System Model

Here, a cellular network having number of cells $L > 1$ has been considered. Each cell consists of a central BS having $M$ antennas and $K$ single-antenna users, where $M \gg K$. Here, it is assumed that the $K$ users use the same time-frequency resources and are uniformly distributed over the entire cell, excluding a central disc of radius $r_h$. The $k$th user in the $i$th cell is represented as $<k, i>$.

The channel propagation factor between the $n$th antenna of the $j$th BS and the user $<k, i>$ is given by

$$g_{ijkn} = h_{ijkn}\beta_{ijk}^{1/2} \tag{2.1}$$

where $h_{ijkn}$ belongs to $CN(0, 1)$, and represents the small scale fading coefficients. On the other hand, $\beta_{ijk}$ represents the large-scale fading coefficients and is given by

$$\beta_{ijk} = z_{ijk}/r_{ijk}^{\gamma} \tag{2.2}$$

where $z_{ijk}$ denotes a shadow fading and follows the log-normal distribution which means $10\log_{10}(z_{ijk})$ follows a normal distribution with zero mean and standard deviation $\sigma_{\text{shadow}}$. $r_{ijk}$ represents the distance between the $j$th BS and the user $<k, i>$. $\gamma$ represents the path-loss exponent.

Here, it is assumed that the number of pilots is equal to the number of users. The number of pilots is $K$ and the same set of $K$ pilots is reused in $L$ cells. This means that the users in each cell use the same set of mutually orthogonal pilot sequences. During the training phase in uplink, the users in each cell send these pilots to the BS for estimating the channel condition. However, users in other adjacent cells use the same pilot sequences which are not mutually orthogonal, and hence interfere causing pilot contamination leading to erroneous channel estimation.

Let $\zeta_k$ forms the set of all the users assigned to $k$th pilot and is given by

$$\zeta_k = (Z_1^k, Z_2^k, Z_3^k, \ldots, Z_L^k) \tag{2.3}$$

where $Z_l^k$ is the user of cell $l$ assigned to $k$th pilot. Then the matrix $P$ used for assigning the pilots is given by

$$P = (\zeta_1, \zeta_2, \zeta_3, \ldots, \zeta_K) \tag{2.4}$$

The number of cases for $\zeta_k$ and $P$ is $K^L$ and $(K!)^{L-1}$ respectively. Pilot scheduling algorithms thus assign the users in various groups in such a way so that the interference between them is minimized, thereby maximizing the overall system performance or throughput.

In this chapter, a system with unlimited BS antennas has been considered, that is $M \to \infty$. With $M$ growing without bound, the effective uplink SINR of user $<k, i>$ is given by

$$\text{SINR}_{ik}^{\text{up}} = \frac{\beta_{iki}^2}{\sum\limits_{l \neq i} \beta_{ikl}^2} \tag{2.5}$$

The uplink SINR is a random quantity and is dependent upon the random user positions and large-scale fading values. The system achievable sum-rate is given by

$$\text{Rate}_{\text{sum}}^{\text{up}} = \sum_{l=1}^{L} \sum_{k=1}^{K} \log_2(1 + \text{SINR}_{ik}^{\text{up}}) \tag{2.6}$$

Hence, the system achievable sum-rate in (2.6) is used as the performance metric that needs to be maximized by applying the pilot scheduling algorithms. It is worth stating

here that the pilot scheduling algorithms are basically global optimization problems. This can be solved using brute-force methods but it raises the complexity on account of high values of $K$ and $L$. The scheduling algorithms incorporates various methods so as to reduce the complexity and achieve near optimum performance in minimum time or iterations. In this chapter, five sub-optimal pilot scheduling algorithms have been explained that has successfully maximized the system sum-rate without increasing the complexity much.

## 2.3 The Pilot Scheduling Algorithms

In this section, various pilot scheduling algorithms are discussed elaborately. Here, greedy-based pilot scheduling, tabu search-based pilot scheduling, sorting-based pilot scheduling, smart pilot assignment and the water-filling-based pilot scheduling algorithms are presented.

### *2.3.1 Greedy Pilot Scheduling Method*

The sum-rate of $\zeta_k$ user group is given by

$$\text{Rate}(\zeta_k) = \sum_{l=1}^{L} \log_2(1 + \text{SINR}^{\text{up}}_{i\kappa_l^k}) \tag{2.7}$$

The variable $\kappa_l^k$ represents the $k$th user in $l$th belonging to the user group $\zeta_k$. Basically, the pilot scheduling algorithm does as follows

$$\begin{aligned} P_{\text{opt}}^{\text{greedy}} &= (\zeta_1^{\text{opt}}, \zeta_2^{\text{opt}}, \zeta_3^{\text{opt}}, \ldots, \zeta_K^{\text{opt}}) \\ &= \arg\max \sum_{k=1}^{K} \text{Rate}(\zeta_k) \end{aligned} \tag{2.8}$$

The main idea of greedy algorithm is selecting a user group having maximum sum-rate in each turn till $K$ groups. The detailed steps are provided as follows.

*Step 1*: Find out the large-scale fading factors between all the users from all cells and BS of all cells, that is $\beta_{ikl}\ \forall\ k \in [1, 2, 3, \ldots, K]$ and $1 \in [1, 2, 3, \ldots, L]$.

*Step 2*: After that, form a table $G$ of all $K^L$ possible user groups and find out the group rate of all the corresponding user groups in the table.

*Step 3*: Find out the maximum group rate value and the corresponding user group from the table. This user group represents $\zeta_1^{\text{opt}}$. That is

$$\zeta_1^{\text{opt}} = \arg\max \text{Rate}(\kappa_1^k, \kappa_2^k, \kappa_3^k, \ldots, \kappa_L^k) \tag{2.9}$$

*Step 4:* Update $G$ by replacing the row corresponding to user group obtained in $\zeta_1^{\text{opt}}$. Also, update $G$ by replacing the column corresponding to user group obtained in $\zeta_1^{\text{opt}}$. Then, find the maximum group rate and corresponding user group from the remaining $(K-1)^L$ possible cases in the table. This user group represents $\zeta_2^{\text{opt}}$. That is

$$\zeta_2^{\text{opt}} = \underset{\kappa_l^k \in G - \zeta_1^{\text{opt}}}{\arg\max}\, \text{Rate}(\kappa_1^k, \kappa_2^k, \kappa_3^k, \ldots, \kappa_L^k) \tag{2.10}$$

This step is repeated until all the K user groups are not obtained, by nullifying the corresponding rows and columns of user groups already obtained in each iteration. The obtained solution after the process is completed is given by (2.4).

This algorithm gives fruitful result but is very inefficient for high values of $K$ and $L$, as the number of possible user groups increases manifold.

### *2.3.2 Tabu Search Pilot Scheduling Method*

The tabu search (TS) method is a local neighborhood low-complexity scheme. The scheme starts with a random initial vector and defines a set of neighborhood vectors corresponding to the initial one. The scheme then forwards to the best neighborhood vector among those present in the list, even when the new vector is worse than the old one. The algorithm is run for a total of $N_{\text{iter}}$ iterations after which the best result obtained till then is taken as the final solution. While defining the set of neighborhood vectors for the present vector, the scheme averts looping by maintaining a tabu list of length $L_{\text{tab}}$ in which the solution vectors of last few iterations are stored for checking. This insures efficient traversal of the search region.

The problem statement remains similar to that of the greedy scheme. We need to find $K$ user groups from matrix $G$ having $K^L$ possible permutations of users so as to maximize the overall system throughput. The sum-rate of the $l$th cell is given by

$$\text{Rate}(\rho_l) = \sum_{k=1}^{K} \log_2(1 + \text{SINR}_{l\kappa_l^k}^{\text{up}}) \tag{2.11}$$

And the final optimized solution that needs to be obtained is given by (4).

Before understanding the actual algorithm, let us understand some salient features of the scheme.

- *Neighborhood*: Given a vector $V$ of length $L$, we define a set $S(V)$ representing the neighborhood of $V$, in which the components are obtained by exchanging any two elements of $V$. Thus, a vector $V$ of length $L$ has $C_L^2$ neighborhood vectors.
- *Tabu List*: The scheme averts cycling or looping by maintaining a list of vectors that have already been obtained in a list known as tabu list of length $L_{\text{tab}}$. When the list is full, the scheme pushes out the initially added vectors from the list one

at a time which once again becomes a candidate of the search space. It basically functions like a queue having first-in-first-out feature.

- *Stopping Criterion*: The algorithm is stopped once the maximum iterations $N_{\text{iter}}$ is achieved.
- *Optimization Problem:* The fitness function for the optimization problem is given by

$$f(P) = \text{Rate}_{\text{sum}}^{\text{up}}(P) \tag{2.12}$$

- *Aspiration Criterion*: During the process, a variable $A$ is used to store the best value obtained till now and the corresponding vector $V$ is denoted as $V^*$. That is,

$$A = f(P|_{V=V^*}) \tag{2.13}$$

The entire flow of the TS algorithm is shown in Fig. 2.1. Thus, the steps of the algorithm are explained below.

*Step 1*: Find out the large-scale fading factors between all the users from all cells and BS of all cells, that is $\beta_{ikl}$ $\forall$ k $\in$ [1,2,3, …,K] and l $\in$ [1, 2, 3, …, $L$].

*Step 2*: The pilot assignment matrix $P$ is initialized randomly and further the cell index $l$ is initialized to zero.

*Step 3*: $l$ is then incremented. If the value of $l$ is greater than $L$, terminate the process. The largest value in the past is retained and the optimal pilot matrix $P_{\text{opt}}^{\text{TS}} = P$.

*Step 4*: The current iteration vector is taken as $V = V_l$, the historical best vector $V^* = V$, the best value so far is $A = f(P)$, the tabu list is $T_l = \phi$ and the counter used for iteration is $c = 0$.

*Step 5*: Increment $c$. If $c > N_{\text{iter}}$, $P(l, :) = V^*$ and move to 3rd step, and otherwise move to 6th step.

*Step 6*: Select all possible neighborhood vector of $V$ and form the set $S(V)$. $\forall$ $V^c \in S(V)$, find $P^c = P|_{V_l=V^c}$. Find the best vector

$$V^c = \arg\max f(P^c) \tag{2.14}$$

Now, $A^c = f(P^c)$. If $A^c > A$, the aspiration criteria is satisfied. We update $V = V^c$, $V^* = V^c$, insert $V^c$ to $T_l$ and go back to 5th step. Otherwise, we move forward to 7th step.

*Step 7:* Here, the best among the candidate vector absent in the given tabu list is selected.

$$V^c = \arg\max f(P^c),\ V^c \in S(V)/T_l \tag{2.15}$$

Updation is done similar to step 6 and returned to step 5.

The TS scheme is very much advantageous than Greedy scheme, but the rate of convergence to global optimum depends upon the selection of the initial vector. If the chosen initial vector is not near optimum, the scheme will take more number of iterations to converge.

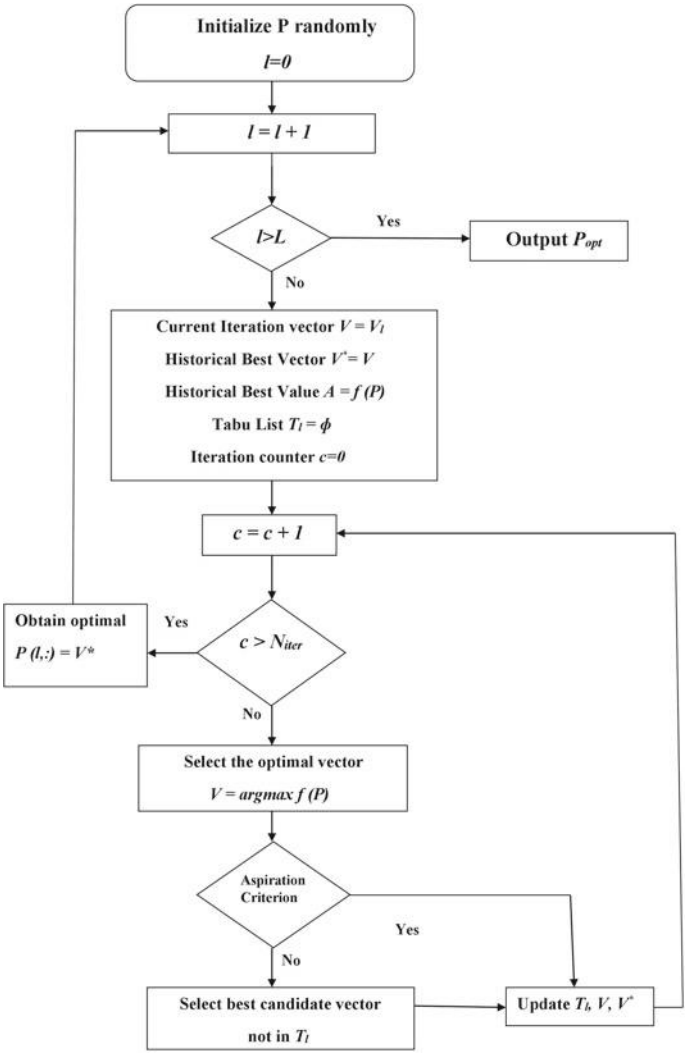



**Fig. 2.1** The TS scheme flowchart

### *2.3.3 The Sorting-Based Scheme*

It has already been learned that pilot contamination is one of the major limitations of the massive MIMO systems. It adds to errors in the estimation of channel by BS. As a consequence of which the system throughput becomes saturated. This algorithm is based on the mean squared error (MSE) of the estimated channel at the BS. The basis of this algorithm is the channel estimation mean square error (MSE) at the BS. Depending upon MSE, it alternatively minimizes the MSE in channel estimation for

users having poor channel conditions. It assumes the large-scale fading factors of all users to be known at the BS. The BS then allocates the pilots to its serving users for channel estimation process, in accordance with the following steps, that allocates the pilots with lesser interference to users with poorer channels. The target cell is assumed to be cell 1. The steps are:

*Step 1*: Given the target cell, the users in the cell are sorted in ascending order on the basis of their large-scale fading coefficients $\beta_{11k}$. This gives a idea to the BS how good are its own users channels. The pilots are then allocated sequentially according to the user's order as 1,2,3, …, $K$.

*Step 2*: Starting from $l$th cell where $l$ = [2, 3, …, $L$], the users in each cell are sorted in ascending order of their large-scale fading coefficients $\beta_{l1k}$. This provides an idea of how much interference the users in other cells will have toward the users of the target cell. The pilots are then assigned sequentially to the sorted users in each cell as done in the target cell.

*Step 3*: To be fair toward every cell, any particular cell can be taken as the target cell by following some specific rules.

The advantage of this algorithm is that in addition to increasing the overall throughput of the system, it also increases the user rate of users having poorer channel conditions.

### 2.3.4 The Smart Pilot Assignment (SPA) Scheme

This scheme was developed in order to improve the performance of users in a cell having high pilot contamination. After finding out the large-scale fading coefficients, the BS finds the inter-cell interference or extent of contamination caused by users using non-orthogonal pilot sequences in adjacent cells. After that, rather than assigning the pilots randomly, the SPA scheme assigns the pilots with lower contamination levels to users having poorer channel conditions in an attempt to enhance their performance. The scheme basically aims at maximizing the minimum uplink SINR which is formulated as an optimization problem shown below.

$$\wp : \max_{B_s} \min_{\forall k} \frac{|g_{iib_s^k}^H g_{iib_s^k}|^2}{\sum_{j \neq i} |g_{ijb_s^k}^H g_{ijb_s^k}|^2 + \dfrac{|\xi_{ib_s^k}^u|^2}{\rho_u}} \tag{2.16}$$

where $\rho_{\mathrm{u}}$ denotes the uplink transmission power and $\xi_{ib_s^k}^{\mathrm{u}}$ denotes the additions of uncorrelated noise and intra-cell interference. $(B_s{:}s = 1, \ldots, K!)$ represents all possible $K!$ pilot assignments. In the limiting case where we consider unlimited BS antennas, the optimization problem simplifies as below.

$$\wp \xrightarrow{M\to\infty} \hat{\wp} : \max_{B_s} \min_{\forall k} \frac{\beta^2_{iib^k_s}}{\sum_{j\neq i} \beta^2_{ijb^k_s}} \tag{2.17}$$

The SPA scheme solves the optimization problem in (17) in a greedy manner with low complexity. The procedure followed by the scheme is given below.

At first a target cell is chosen randomly. The large-scale fading coefficients of all the $K$ users in the cell are calculated and stored in a separate variable to get an idea of their channel qualities.

$$\alpha_k = \beta^2_{iik} \forall k \in (1, 2, 3, \ldots, K) \tag{2.18}$$

After that, the large-scale fading coefficients of users from adjacent cells are calculated and stored in a variable in order to get an idea about the inter-cell interference affecting the target cell's users.

$$\gamma_k = \sum_{j\neq i} \beta^2_{ijk} \forall k \in (1, 2, 3, \ldots, K) \tag{2.19}$$

For maximizing the minimum SINR, it is necessary that users with poorer channel conditions are not assigned pilots with high interference levels. Based on this fact, the SPA scheme preferably allocates the pilot sequence having least inter-cell interference to the user with poorest channel quality in a sequential manner. Mathematically, we first sort the pilots based on their extent of interference in descending order, that is

$$B_{\mathrm{p}} = (\phi_{b^1_{\mathrm{p}}}, \phi_{b^2_{\mathrm{p}}}, \phi_{b^3_{\mathrm{p}}}, \ldots, \phi_{b^K_{\mathrm{p}}}) \tag{2.20}$$

where $\gamma_{b^1_p} \geq \gamma_{b^2_p} \geq \gamma_{b^3_p} \geq \cdots \geq \gamma_{b^K_p} > 0$.

Next, we sort the users in target cell based on their channel qualities in descending order, as follows.

$$B_{\mathrm{u}} = (U_{b^1_{\mathrm{u}}}, U_{b^2_{\mathrm{u}}}, U_{b^3_{\mathrm{u}}}, \ldots, U_{b^K_{\mathrm{u}}}) \tag{2.21}$$

where $\alpha_{b^1_{\mathrm{u}}} \geq \alpha_{b^2_{\mathrm{u}}} \geq \alpha_{b^3_{\mathrm{u}}} \geq \cdots \geq \alpha_{b^K_{\mathrm{u}}} > 0$. Then, the pilot assignment is done by assigning the pilot $\phi_{b^K_{\mathrm{p}}}$ to the user $U_{b^K_{\mathrm{q}}}$ for all $k \in (1, 2, 3, \ldots, K)$.

It has been found that the SPA scheme indeed enhances the minimum SINR by good margin.

### *2.3.5 The Water-filling Based Pilot Assignment Scheme*

Similar to the SPA scheme, this scheme also takes into account the channel conditions of various users and the interfering nature of different pilot sequences. This is a low-complexity scheduling scheme inspired from the water-filling algorithm. However, this scheme assigns the less interfering pilot sequences to users with better channel

conditions, much similar to the water-filling approach where the BS assigns more power to users with good channels. The SINR expression is shown as follows.

$$\text{SINR}_{ik} = \frac{\beta_{iik}^2}{\sum\limits_{j \neq i} \beta_{ijk}^2} \tag{2.22}$$

After that, we obtain the achievable sum-rate expression as:

$$\text{Rate}_{\text{sum}} = \sum_{i=1}^{L} \sum_{k=1}^{K} \log_2(1 + \text{SINR}_{ik}) \tag{2.23}$$

The $K$ pilots in one cell are defined as $P = (\phi_1, \phi_2, \ldots, \phi_K)$ and the $K$ users are $U = (u_1, u_2, \ldots, u_K)$. The total number of user-pilot combination $< P, U >$ is $K!$, and hence optimum allocation is not possible to be found exhaustively when number of users is large. This algorithm reduces the complexity of searching by allocating pilots having less interference to users with better channel qualities in a sequential manner. The quality of channels is calculated using large-scale fading coefficients and stored in an array $(Q_k)_{k=1}^{K}$.

$$Q_k = \beta_{iik}^2 \forall k \in (1, 2, \ldots, K) \tag{2.24}$$

Another array is taken to store the interference levels of the $K$ pilots to be assigned to the users as $(R_k)_{k=1}^{K}$, where

$$R_k = \sum_{j \neq i} \beta_{ijk}^2 \forall k \in (1, 2, \ldots, K) \tag{2.25}$$

The scheme then sorts the $Q_k$ matrix in ascending order and the $R_k$ matrix in descending order. After that, it assigns the $i$th pilot with interference level $R_i$ to the $i$th user having channel quality $Q_i$.

This scheme is different from SPA in the sense that while SPA attempts at maximizing the minimum SINR, the water-filling approach described here attempts to maximize the overall system SINR.

## 2.4 Results and Discussions

In this section, the performance of the pilot scheduling algorithms mentioned in the previous section has been analyzed along with the simulation results.

### 2.4.1 Greedy and TS Scheme

The difference cumulative distribution function (CDF) between the average rate of greedy scheme with random pilot allocation has been plotted in Figs. 2.2 and 2.3 for various values of user numbers ($K = 2, 4, 6$ and 8) and showed their performance for two values of cell numbers ($L = 2$ and 3). It can be observed from the results that for lower values of K in both cases, such as for $K = 2$, the scheme may perform badly with respect to the random allotment due to the fact that the greedy scheme cannot ensure global optimization with so less number of choices. But as the value of $K$ and $L$ increases, it is visible from the plots that the greedy scheme performs much better than the random allocation of pilots.

In Fig. 2.4, the CDF of difference in mean achievable rate between TS and random pilot allocation scheme per UT has been shown for the cases of ($L = 2$) and ($K = 2, 4, 6$ and 8). From the figure, it can be seen that as K increases, the performance of TS scheme gets better as compared to the random allotment. Also, for higher values of K and L, the TS scheme outperforms the greedy scheme in terms of complexity but has lower performance than the greedy scheme.

Thus, it can be said that when the value of K and L is small, the greedy scheme can be used because of its better performance in terms of throughput, but while moving toward large cellular networks where the number of cells and users increases, it is required to shift to the TS scheme because the greedy scheme becomes much complex for realization.

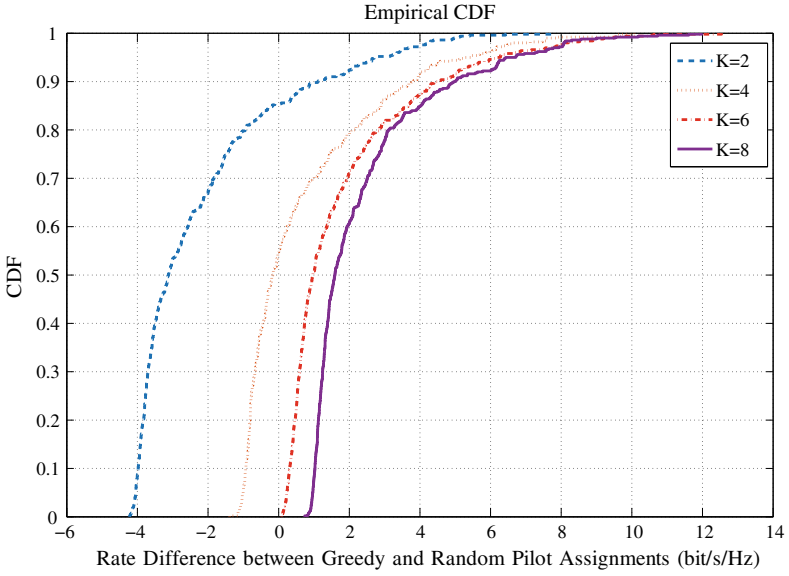



**Fig. 2.2** CDF of the difference between the average rate per UT of greedy and random pilot allocation schemes. In this case, $L = 2$ and $K = 2, 4, 6$, and 8.

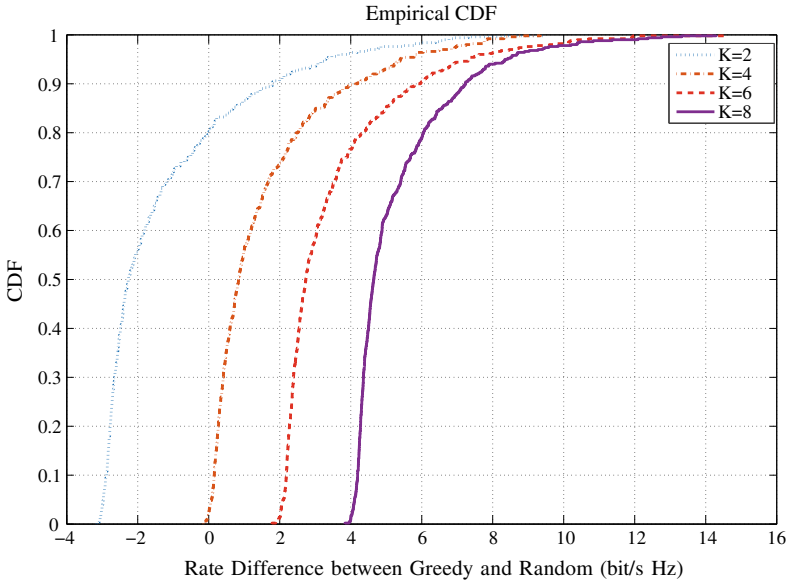



**Fig. 2.3** CDF of the difference between the average rate per UT of greedy and random pilot allocation schemes. In this case, $L = 3$ and $K = 2, 4, 6$, and 8

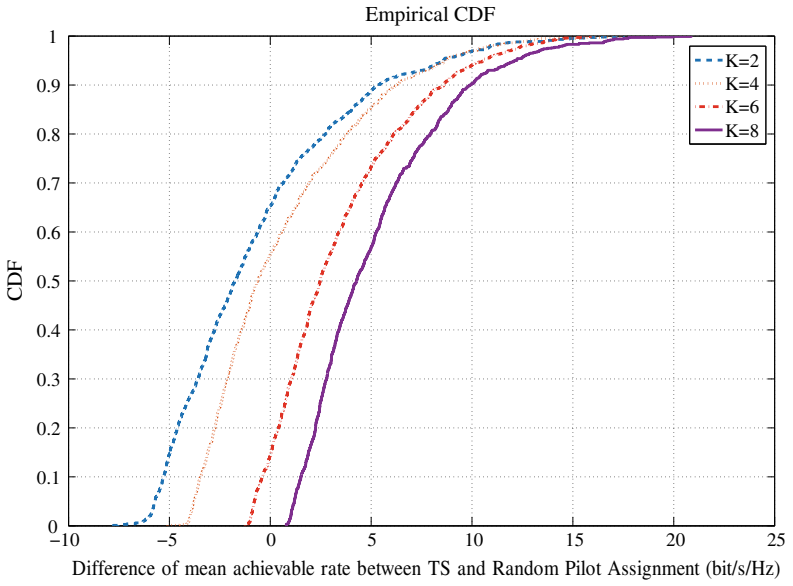



**Fig. 2.4** CDF of the difference between the average rate per UT of TS and random pilot allocation schemes. In this case, $L = 2$ and $K = 2, 4, 6$, and 8

### 2.4.2 *Sorting-Based Scheme*

In this case, a basic cellular system with 2 cells has been considered with each of the cell consisting of four single-antenna users $K$ and a BS having $M$ antennas. Users present in the cell are distributed randomly and all the users in each cell reuse the same set of orthogonal pilot sequences. The cell radius is taken as 500 meters. The pilot transmission power of users is taken as 10 dB and transmitting power of data is taken as 20 dB in general, without told otherwise.

Figure 2.5 shows the system sum-rate plots with respect to number of BS antennas ($M$) for two cases, with and without pilot contamination. For the case of pilot allocation without any contamination, the users in every cell are assigned different orthogonal pilots. For instance, in our system with 2 cells having four users each, total number of mutually orthogonal pilot sequences are taken as 8. It can be observed from the plot that pilot contamination highly degrades the performance of cellular network and decreases its throughput.

Another plot is shown in Fig. 2.6 where the system sum-rate is observed by changing the pilot transmission power gradually. As observed from the plot, increasing the number of BS antenna significantly enhances the overall system performance. But increasing the pilot transmission power only has effect in very low value ranges. As we keep on increasing the power gradually, the system performance does not increase

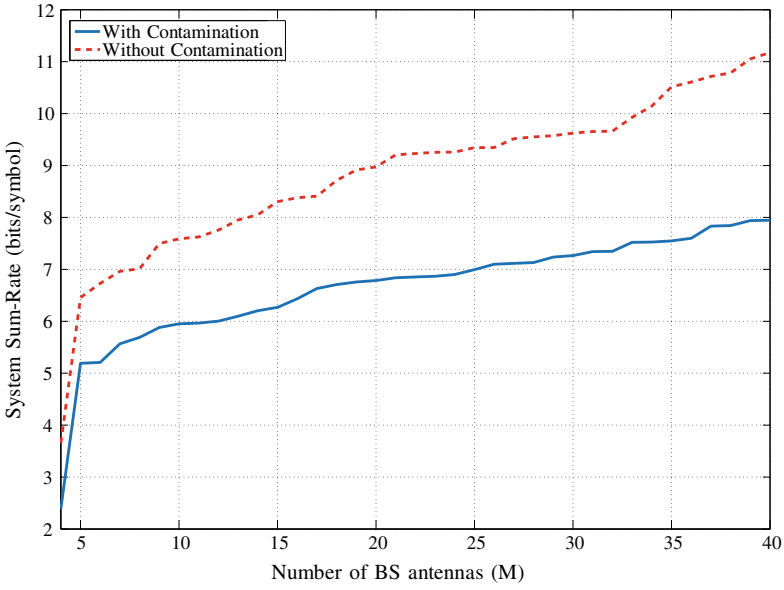



**Fig. 2.5** System sum-rate considering cases including and excluding pilot contamination

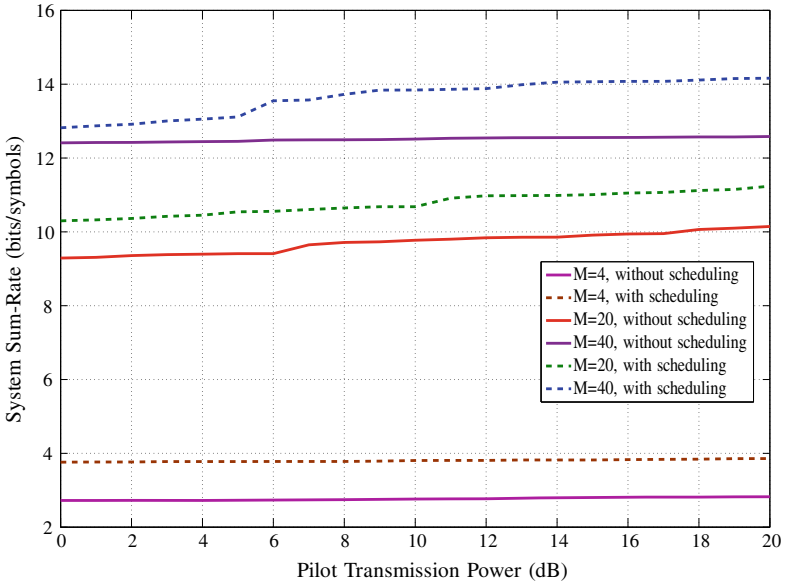



**Fig. 2.6** System sum-rate with respect to pilot transmission power in cases with various numbers of BS antennas

much. Also, from the plot, it is clear that the algorithm indeed helps in improving the overall system capacity with respect to the scenario where no scheduling is done and pilots are assigned randomly.

### *2.4.3 The SPA scheme and Water-filling Pilot Allocation*

Here, the results of the smart pilot assignment (SPA) scheme and the water-filling pilot allocation are being discussed by showing the simulation results. The proposed schemes are being tested for different values of parameters involved for verifying the effectiveness of the algorithms. The simulations are run for iterations of 1000 to eradicate any errors that may have occurred during the computations of large-scale fading coefficients. The number of cells $L$ is taken as 7, with each cell having $K = 5$ number of users. The cell radius is taken as 2000 meters with a central disk of 100 meters around the BS.

Figure 2.7 shows a comparison between both the algorithms in terms of overall system sum-rate. From the figure, it is clear that the water-filling pilot allocation approach performs better than the SPA in terms of system sum-rate because the ultimate objective of the water-filling approach is to optimize the pilot allocation for better system capacity by assigning the pilots fairly to the users in terms of their channel qualities.

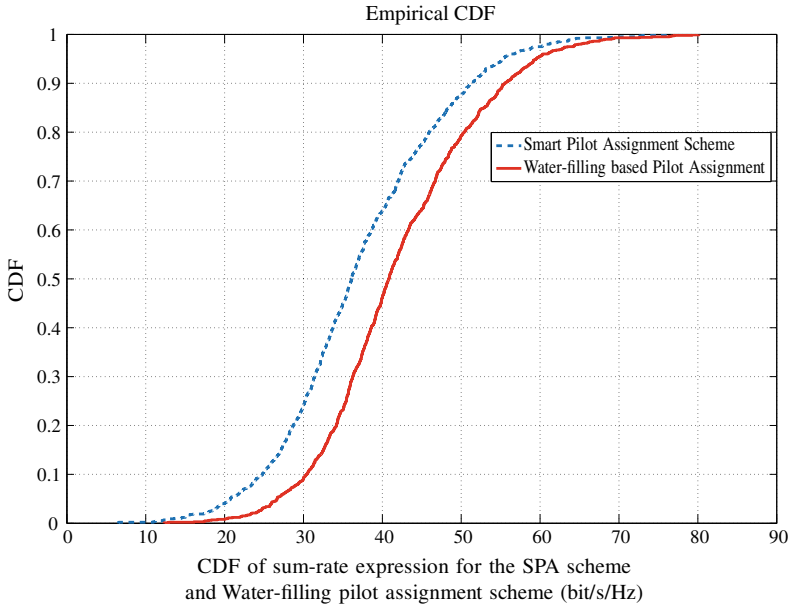



**Fig. 2.7** Comparison of system sum-rate for the two schemes

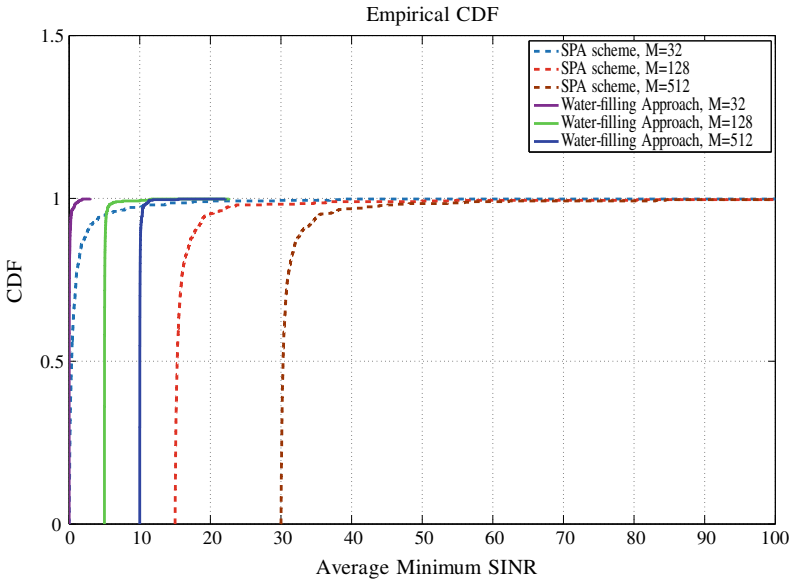



**Fig. 2.8** CDF of average minimum SINR for the two schemes

Figure 2.8 compares both the algorithms in terms of average minimum SINR per user in the system. From the figure, it is clear that the SPA has better result than water-filling approach for maximizing the minimum SINR because the objective of SPA is the same and it allocates the pilots by considering the users with weak channels to be served with better pilots having less interference. Also, the minimum average SINR shows an increasing trend with the number of BS antenna.

In terms of complexity, both the schemes have better performance than the greedy scheme. Mathematically speaking, the computational complexity of the SPA and water-filling approach is of the order of $O(KLlogK)$ which is much better than the greedy scheme having complexity of the order of $O(K^L)$ for higher values of $K$ and $L$.

## 2.5 Conclusion

In this chapter, an overview of massive MIMO TDD system has been given. The effects of pilot contamination in channel estimation and subsequent data transmission in TDD cellular systems have been analyzed. Different procedures of mitigating and eliminating these effects of pilot contamination have been discussed. Several algorithms proposed in the literature for mitigating the pilot contamination have been discussed thoroughly in this chapter. Their effectiveness has been verified using simulation results. A comparison of various algorithms has also been done to show the benefits and shortcomings of each of them. Still the research community has different open research opportunities in the domain of effective pilot scheduling to be explored for achieving better system throughput by the massive MIMO communications.

## References

Adhikary A, Nam J, Ahn J, Caire G (2013) Joint spatial division and multiplexing-the large-scale array regime. IEEE Trans Inf Theory 59(10):6441–6463. https://doi.org/10.1109/TIT.2013.2269476

Appaiah K, Ashikhmin A, Marzetta TL (2010) Pilot contamination reduction in multi-user TDD systems. In: 2010 IEEE international conference on communications, pp 1–5. https://doi.org/10.1109/ICC.2010.5502810

Ashikhmin A, Marzetta T (2012) Pilot contamination precoding in multi-cell large scale antenna systems. In: 2012 IEEE International symposium on information theory proceedings, pp 1137–1141. https://doi.org/10.1109/ISIT.2012.6283031

Boccardi F, Heath RW, Lozano A, Marzetta TL, Popovski P (2014) Five disruptive technology directions for 5g. IEEE Commun Mag 52(2):74–80. https://doi.org/10.1109/MCOM.2014.6736746

Chávez-Santiago R, Szydełko M, Kliks A, Foukalas F, Haddad Y, Nolan KE, Kelly MY, Masonta MT, Balasingham I (2015) 5G: The convergence of wireless communications. Wirel Pers Commun 83(3):1617–1642

Chen J, Berry RA, Honig ML (2007) Performance of limited feedback schemes for downlink OFDMA with finite coherence time. In: 2007 IEEE international symposium on information theory, pp 2751–2755. https://doi.org/10.1109/ISIT.2007.4557186

Chu S, Wang X, Yang Y (2014) Adaptive scheduling in MIMO-based heterogeneous ad hoc networks. IEEE Trans Mobile Comput 13:964–978

Elijah O, Leow CY, Rahman TA, Nunoo S, Iliya SZ (2016) A comprehensive survey of pilot contamination in massive MIMO-5G system. IEEE Commun Surv Tutor 18(2):905–923. https://doi.org/10.1109/COMST.2015.2504379

Eslami M, Krzymień WA (2009) Limited-feedback multiuser MIMO-OFDM downlink with spatial multiplexing and per-chunk/per-antenna user scheduling. Springer, Netherlands, Dordrecht, pp 269–278

Eslami M, Krzymien WA (2011) Net throughput maximization of per-chunk user scheduling for MIMO-OFDM downlink. IEEE Trans Veh Technol 60:4338–4348

Fakhereddin M, Sharif M, Hassibi B (2009) Reduced feedback and random beamforming for OFDM MIMO broadcast channels. IEEE Trans Commun 57:3827–3835

Fernandes F, Ashikhmin A, Marzetta T (2012) Interference reduction on cellular networks with large antenna arrays. In: IEEE international conference on communications (ICC), pp 1–5

Fernandes F, Ashikhmin A, Marzetta TL (2013) Inter-cell interference in noncooperative tdd large scale antenna systems. IEEE J Sel Areas Commun 31(2):192–201. https://doi.org/10.1109/JSAC.2013.130208

Gopalakrishnan B, Jindal N (2011) An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas. In: 2011 IEEE 12th international workshop on signal processing advances in wireless communications, pp 381–385. https://doi.org/10.1109/SPAWC.2011.5990435

Hossain E, Hasan M (2015) 5G cellular: Key enabling technologies and research challenges. IEEE Instrum Measur Mag 18(3):11–21. https://doi.org/10.1109/MIM.2015.7108393

Hou X, Harada A, Suda H (2012) Experimental study of advanced MU-MIMO scheme with antenna calibration for the evolving LTE TDD system. In: 2012 IEEE 23rd International symposium on personal, indoor and mobile radio communications (PIMRC), pp 2443–2448. https://doi.org/10.1109/PIMRC.2012.6362767

Jin S, Li M, Huang Y, Du Y, Gao X (2015) Pilot scheduling schemes for multi-cell massive multiple-input multiple-output transmission. IET Commun 9(5):689–700. https://doi.org/10.1049/iet-com.2014.0842

Jorswieck E, Sezgin A, Ottersten B, Paulraj A (2008) Feedback reduction in uplink MIMO OFDM systems by chunk optimization. EURASIP J Adv Sig Process 1–14. https://doi.org/10.1155/2008/597072

Jose J, Ashikhmin A, Marzetta TL, Vishwanath S (2011) Pilot contamination and precoding in multi-cell tdd systems. IEEE Trans Wirel Commun 10(8):2640–2651. https://doi.org/10.1109/TWC.2011.060711.101155

Kurras M (2013) Positioning of multi-node/multi-antenna transmission technologies. https://doi.org/10.13140/RG.2.1.1888.9201

Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2):186–195. https://doi.org/10.1109/MCOM.2014.6736761

Li Q, Niu H, Papathanassiou A, Wu G (2014) 5g network capacity: Key elements and technologies. Veh Technol Mag 9:71–78 (IEEE). https://doi.org/10.1109/MVT.2013.2295070

Lu L, Li GY, Swindlehurst AL, Ashikhmin A, Zhang R (2014) An overview of massive mimo: benefits and challenges. IEEE J Sel Top Sig Process 8(5):742–758. https://doi.org/10.1109/JSTSP.2014.2317671

Marzetta TL (2006) How much training is required for multiuser mimo? In: 2006 Fortieth Asilomar conference on signals, systems and computers, pp 359–363. https://doi.org/10.1109/ACSSC.2006.354768

Marzetta TL (2010) Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun 9(11):3590–3600. https://doi.org/10.1109/TWC.2010.092810.091092

Medard M (2000) The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. IEEE Trans Inf Theory 46(3):933–946. https://doi.org/10.1109/18.841172

Min M, Kim D, Kim H, Im G (2013) Opportunistic two-stage feedback and scheduling for MIMO downlink systems. IEEE Trans Commun 61:312–324

Mueller RR, Vehkaperae M, Cottatellucci L (2013) Blind pilot decontamination. In: WSA 2013; 17th International ITG workshop on smart antennas, pp 1–6

Naeem M, Lee D (2014) A joint antenna and user selection scheme for multiuser MIMO system. Elsevier App Soft Comput 23:366–374

Nam J, Ahn J, Adhikary A, Caire G (2012) Joint spatial division and multiplexing: realizing massive MIMO gains with limited channel state information. In: 2012 46th Annual conference on information sciences and systems (CISS), pp 1–6. https://doi.org/10.1109/CISS.2012.6310934

Ngo HQ, Marzetta TL, Larsson EG (2011) Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 3464–3467. https://doi.org/10.1109/ICASSP.2011.5947131

Ngo HQ, Larsson EG, Marzetta TL (2013) Energy and spectral efficiency of very large multiuser mimo systems. IEEE Trans Commun 61(4):1436–1449. https://doi.org/10.1109/TCOMM.2013.020413.110848

Osseiran A, Boccardi F, Braun V, Kusume K, Marsch P, Maternia M, Queseth O, Schellmann M, Schotten H, Taoka H, Tullberg H, Uusitalo MA, Timus B, Fallgren M (2014) Scenarios for 5g mobile and wireless communications: the vision of the metis project. IEEE Commun Mag 52(5):26–35. https://doi.org/10.1109/MCOM.2014.6815890

Pattanayak P, Kumar P (2015) A computationally efficient genetic algorithm for MIMO broadcast scheduling. Elsevier Appl Comput 37:545–553

Pattanayak P, Kumar P (2016) Quantized feedback MIMO scheduling for heterogeneous broadcast networks. Springer Wireless Networks, pp 1–18

Pattanayak P, Kumar P (2017) Quantized feedback scheduling for MIMO-OFDM broadcast networks with subcarrier clustering. Elsevier Ad Hoc Netw 65:26–37

Pattanayak P, Kumar P (2018) Combined user and antenna scheduling scheme for MIMO-OFDM networks. Telecommun Syst. https://doi.org/10.1007/s11235-018-0462-0

Pattanayak P, Kumar P (2019) An efficient scheduling scheme for MIMO-OFDM broadcast networks. AEU Int J Electron Commun 101:15–26. 10.1016/j.aeue.2019.01.017, http://www.sciencedirect.com/science/article/pii/S1434841118300359

Pattanayak P, Kumar P (2020) Computationally efficient scheduling schemes for multiple antenna systems using evolutionary algorithms and swarm optimization. In: Gandomi AH, Emrouznejad A, Jamshidi MM, Deb K, Rahimi I (eds) Evolutionary computation in scheduling. Wiley, chap 5, pp 105–135

Pattanayak P, Kumar P (2016) SINR based limited feedback scheduling for MIMO-OFDM heterogeneous broadcast networks. In: Proceedings of the IEEE twenty second national conference on communication (NCC), Mar 2016

Pattanayak P, Roy KM, Kumar P (2015) Analysis of a new MIMO broadcast channel limited feedback scheduling algorithm with user grouping. Springer Wirel Pers Commun 80:1079–1094

Pattanayak P, Trivedi VK, Chakraborty S, Kumar P (2017) BER performance of multi user scheduling for MIMO-STBC and MIMO-OFDM broadcast network with imperfect CSI. In: 2017 4th International conference on signal processing and integrated networks (SPIN), pp 66–70. https://doi.org/10.1109/SPIN.2017.8049917

Peng Xu, Wang J, Wang J (2013) Effect of pilot contamination on channel estimation in massive MIMO systems. In: 2013 International conference on wireless communications and signal processing, pp 1–6. https://doi.org/10.1109/WCSP.2013.6677112

Prasad N, Zhang H, Zhu H, Rangarajan S (2013) Multi-user MIMO scheduling in the fourth generation cellular uplink. IEEE Trans Wirel Commun 12(9):4272–4285. https://doi.org/10.1109/TWC.2013.072513.120734

Rusek F, Persson D, Lau BK, Larsson EG, Marzetta TL, Edfors O, Tufvesson F (2013) Scaling up mimo: opportunities and challenges with very large arrays. IEEE Sig Process Mag 30(1):40–60. https://doi.org/10.1109/MSP.2011.2178495

Stuber GL, Barry JR, McLaughlin SW, Li Y, Ingram MA, Pratt TG (2004) Broadband MIMO-OFDM wireless communications. Proc IEEE 92(2):271–294. https://doi.org/10.1109/JPROC.2003.821912

Svedman P, Wilson S, Cimini L, Ottersten B (2007) Opportunistic beamforming and scheduling for OFDMA systems. IEEE Trans Commun 55:941–952

Wang F, Bialkowski ME, Bialkowski KS (2013) Successive user scheduling schemes for a multiuser MIMO system employing generalized channel inversion. Springer Wirel Pers Commun 68:401–416

Wang H, Huang Y, Jin S, Yu F, Yang L (2013) Performance analysis on precoding and pilot scheduling in very large MIMO multi-cell systems. In: 2013 IEEE wireless communications and networking conference (WCNC), pp 2722–2726. https://doi.org/10.1109/WCNC.2013.6554992

Xu W, Cao Y, Zhao C (2013) Capacity analysis of MIMO group-broadcast channels with time-division scheduling. Springer Wirel Pers Commun 68:1331–1350

Yin H, Gesbert D, Filippou M, Liu Y (2013) A coordinated approach to channel estimation in large-scale multiple-antenna systems. IEEE J Sel Areas Commun 31(2):264–273. https://doi.org/10.1109/JSAC.2013.130214

Yue S, Liu J, Zhai C, Wang Q (2015) User pilot scheduling in massive MIMO systems. In: 2015 International conference on wireless communications signal processing (WCSP), pp 1–5. https://doi.org/10.1109/WCSP.2015.7341058

Zhang C, Zeng G (2014) Pilot contamination reduction scheme in massive MIMO multi-cell TDD systems. J Comput Syst Sci 10(15):6721–6729

Zhou Z, Wang D (2016) Pilot scheduling based on water-filling algorithm in massive MIMO. In: 2016 6th International conference on electronics information and emergency communication (ICEIEC), pp 89–92. https://doi.org/10.1109/ICEIEC.2016.7589694

Zhu X, Wang Z, Dai L, Qian C (2015) Smart pilot assignment for massive MIMO. IEEE Commun Lett 19(9):1644–1647. https://doi.org/10.1109/LCOMM.2015.2409176

# Chapter 3
# Detection Techniques in Uplink Massive MIMO Systems


**Arijit Datta, Manish Mandloi, and Vimal Bhatia**

**Abstract** Devising low-complexity data detection techniques is one of the fundamental challenges in the uplink of massive multiple-input multiple-output (MIMO) wireless systems. Linear detection techniques such as zero-forcing (ZF) and minimum mean square error (MMSE) are shown to achieve near-optimal bit-error-rate (BER) performance in such systems. However, ZF/MMSE technique requires inversion of large-dimensional matrices which makes them practically infeasible. This motivates the development of alternate low-complexity inversionless detection techniques which are capable of achieving BER performance close to that of the ZF/MMSE detectors with comparatively less computations. Recently, there is an upsurge in research toward solving this crucial issue in massive MIMO systems. In particular, several detection algorithms have been proposed in the literature, which provides a better trade-off between BER performance and computational complexity. This chapter discusses the fundamentals of massive MIMO detection and also provides an overview of some of the recent state-of-the-art detection techniques. Simulation results on BER performance and computational complexity of these algorithms are also compared to draw useful insights. Furthermore, research scopes in massive MIMO detection are also discussed to provide possible research directions in the field.

**Keywords** Massive MIMO · Bit-error rate · Low-complexity · Matrix inversion · Minimum mean square error

A. Datta
Discipline of Electrical Engineering,
Indian Institute of Technology Indore, Indore, Madhya Pradesh 453552, India
e-mail: phd1601102003@iiti.ac.in

M. Mandloi (✉)
Department of Electronics and Telecommunication Engineering,
SVKM's NMIMS Shirpur Campus, Shirpur, Maharashtra 425405, India
e-mail: manish.mandloi@nmims.edu

V. Bhatia
Discipline of Electrical Engineering,
Indian Institute of Technology Indore, Indore, Madhya Pradesh 453552, India
e-mail: vbhatia@iiti.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_3

## 3.1 Introduction

Multiple-input multiple-output (MIMO) has been a successful technique for achieving improved spectral efficiency in the fourth-generation (4G) wireless standards such as long-term evolution (LTE), LTE-Advanced, wireless interoperability for microwave access (WiMAX) and 802.11 (WiFi) (Lu et al. 2014). The underlying idea in MIMO systems is that it facilitates the transmission of multiple data streams simultaneously from source to destination, which enhances the diversity and multiplexing gains in wireless systems. Moreover, the introduction of multiple antenna wireless systems has opened up new research dimensions such as space-time block coding, index modulation and antenna switching for more reliable and efficient communication. With the unprecedented growth in data starving wireless applications such as Internet of things (IoT), wireless sensor networks (WSN), device-to-device communications (D2D) and vehicle to everything (V2X) communications, high spectral and energy-efficient wireless techniques are of particular interest (Andrews 2014). Massive multiple-input multiple-output (MIMO) is a key enabling technology for realizing the demand for excessive data rates, spectral efficiency and energy efficiency in the fifth generation (5G) and beyond wireless communication systems (Andrews 2014; Larsson et al. 2014). Massive MIMO systems have attracted a lot of research attention due to the fact that such systems, when used in the existing cellular/communication networks, could provide enormous advantages in terms of data rate, quality of service and number of users serviced simultaneously (Andrews 2014; Larsson et al. 2014; Lu et al. 2014). In massive MIMO systems, base station (BS) employs hundreds to thousands of antennas to serve few tens of the users. Large BS antennas in massive MIMO enhance the diversity and the multiplexing gain which in turn leads to the improvement in both the reliability of service and the rate of information transfer, respectively (Larsson et al. 2014). However, practically, there are several challenging issues such as architectural complexity, power requirements, latency in signal processing and size of antenna array (Rusek 2013). Amongst these challenging issues, computational complexity for reliable detection of the transmitted information at the receiver side is extremely crucial for robust receiver architectures in 5G and beyond systems (Larsson 2009; Wu et al. 2014). Moreover, the data rate requirement of recent wireless applications such as IoT, WSN, D2D and V2X necessitates the development of ultra-low latency and highly reliable signal processing algorithms and efficient architecture designs.

This chapter introduces the problem of symbol detection in the uplink of massive MIMO systems. The chapter starts with a brief review of the detection techniques in the literature, followed by the introduction of the massive MIMO system model and formulation of the detection problems. Some of the state-of-the-art detection techniques are discussed in detail with their simulation result on BER performance and the computational complexity. Finally, the chapter concludes with the possible future directions for the design of efficient detection techniques for practical realization of massive MIMO system in 5G and beyond systems.

### 3.1.1 Literature Overview and Challenges in Massive MIMO Detection

Optimal detection in MIMO systems requires an exhaustive search over all the possible transmit symbol vectors. Specifically, the symbol vector which maximizes the *a posteriori* probability is selected as the optimal solution and the technique is known maximum a posteriori (MAP) detection. It is also called as maximum likelihood (ML) detection when the transmit symbols are equally likely. ML exhibits exponential computational complexity in terms of the number of users with modulation order as the base. Which makes it impractical in large-scale and massive MIMO systems (Chockalingam and Rajan 2014). Another detector which achieves performance close to the ML detector is sphere decoder (SD) (Viterbo and Boutros 1999). However, SD is practical only for small-scale MIMO systems (Barbero and Thompson 2008). Over the last decade, different techniques have been adopted for devising low-complexity detection algorithms for the conventional MIMO and large MIMO systems. These techniques include neighborhood search-based detector (Sah and Chaturvedi 2017; Silva et al. 2018), lattice reduction-based detector (Zhou and Ma 2013), sparsity-based MIMO detection (Fadlallah et al. 2015; Peng et al. 2015), Gibbs sampling-based detector (Mandloi and Bhatia 2017b; Mussi and Abrão 2018) and interference cancellation-based detector (Li et al. 2011; Mandloi et al. 2017). The algorithms in Sah and Chaturvedi (2017) to Yang and Hanzo (2015) overcome the computational issues with a good BER performance in the conventional and large MIMO systems. However, these algorithms cannot be directly applied to the massive MIMO systems due to complexity constraints. Interestingly, due to the excessive BS antennas, the wireless fading channel exhibits an important property termed as *channel hardening*. Due to channel hardening, the columns of the MIMO channel matrix tend to be nearly orthogonal with the increase in the number of BS antennas (of the order of hundreds to thousands). Therefore, simple linear detectors such as zero-forcing (ZF) and minimum mean-squared error (MMSE) achieve near-optimal BER performance (Rusek 2013). Despite of their BER performance, ZF/MMSE detectors involve inversion of matrices with large dimensions which require $O(n^3)$ computations, feasible design of linear detectors in massive MIMO systems for 5G and beyond systems are not practical (Wu et al. 2014). In other words, in massive MIMO systems, matrices involved in ZF/MMSE become large in dimension, and therefore, computing inverse of such large-dimensional matrices is computationally challenging as it introduces a significant delay in processing, cost of receiver design and space required for the receiver architecture. This motivates for the design of new detection techniques which can achieve near-optimal BER performance with comparatively low-computational complexity. The objective of further research in massive MIMO detection is to solve linear detection (ZF/MMSE) problem using alternate techniques such as approximate inverse and iterative search methods for massive MIMO systems.

To this end, several researchers have proposed various detection techniques for the uplink of massive MIMO systems, where the following two issues are considered

and solved up to some extent, i.e., (a) to have low complexity in terms of number of computations and (b) to achieve near-optimal BER performance. The important idea behind such developments is either the introduction of approximate matrix inversion techniques and/or iterative techniques for detecting information symbols in massive MIMO systems. Recently, there has been an increased research attention toward devising techniques for massive MIMO systems. Detection techniques developed in the literature for massive MIMO systems can be broadly classified into two types: (1) approximate inverse techniques which involve methods for finding an approximation of the exact inverse of the ZF/MMSE filter matrix for estimating the transmitted information vector, and (2) iterative detection techniques where a low-complexity initial solution is refined iteratively. These two detector types differ in the way they find the estimates of the transmitted information symbol while incurring low computations in the processing.

The first type of massive MIMO detection techniques is based on finding an approximate inverse of the ZF/MMSE filter matrix by using either series expansion method or iterative matrix inverse update technique. Neumann series expansion method (Yin et al. 2013; Zhu et al. 2015) is a well-known series expansion method which is based on Taylor series expansion of the inverse of a matrix. The series expansion consists of a large number of terms; however, to reduce the number of computations, only first few terms are considered which results in an approximate inverse of the matrix. NS-based detection is mostly used as an initial block in various other iterative detection schemes. On the other hand, Newton iteration (NI)-based approximate inversion technique is an iterative inverse update technique where an initial estimate of the inverse of a matrix is refined iteratively (Tang et al. 2016; Wu et al. 2014). These approximate inverse techniques provide good estimates of the inverse of ZF/MMSE filter matrix but are limited by their applications to the massive MIMO systems having very less number of users for, e.g., number of base station antennas = 128 and number of users = 8. In order to improve the performance of NI-based detection with low complexity, the concept of band matrix is introduced in the literature in Tang et al. (2016) where the band of off-diagonal elements is considered in the initialization part of the NI detector. Another important idea of stair matrix is proposed recently in the literature which considers the stair-like elements in the band diagonal matrix (Jiang et al. 2018). Stair matrix-based detection technique is shown to achieve superior BER performance over the NI-based detectors. Further, in Minango and Almeida (2018), the Newton–Schultz iteration for finding approximate inverse is proposed which obtains a near exact matrix inverse with an accelerated convergence rate over the conventional NI-based matrix inversion. The approximate matrix inversion technique is limited to the massive MIMO systems with comparatively less number of users. Therefore, low-complexity iterative algorithms have been proposed in the literature wherein first the detection problem is modeled as solving a set of the linear system of equations, and next, the iterative refinements are performed over the initial solution. The iterative methods which are being used recently for massive MIMO detection include Richardson method (Gao et al. 2014), conjugate gradient method (Hu et al. 2014), join steepest descent and Jacobi method (Jiang et al. 2017; Qin et al. 2015), Gauss–seidel method (Dai et al.

2015; Wu et al. 2016), iterative sequential detection (Mandloi and Bhatia 2016), error recovery-based detector (Mandloi and Bhatia 2017a), message passing-based detection (Zeng et al. 2018) and hybrid pseudo-stationary detector (Datta et al. 2018).

Though the existing detection algorithms for massive MIMO systems are able to achieve near MMSE performance with comparatively less complexity, they lack in the consistency of maintaining the good BER performance when the number of users scales-up in massive MIMO (Fukuda and Abrao 2019). Therefore, it is highly required to look further on the problem and search novel techniques which do not involve computationally hard matrix inversions of large-dimensional matrices and also that the techniques are capable of achieving near-optimal BER performance when the number of users is scaling up in massive MIMO systems.

## 3.2 System Model and Problem Formulation

In this section, we briefly discuss the mathematical model of uplink massive MIMO system. The problem formulation for symbol detection is also carried in this section along with the ML and MMSE detection techniques. In order to build the motivation for approximate inversion and iterative detection techniques, challenges with the MMSE detectors are also discussed later in the section.

### *3.2.1 Mathematical Model*

We consider $K$ single antenna users and a base station with $M$ antennas ($M \gg K$ for e.g., $M = 128$, $K = 16$ ) as depicted in Fig. 3.1. Each user transmits its information symbol say $x_i$, for $i = 1, 2, \ldots, K$, to the base station simultaneously which is also known as uplink transmission. Each symbol $x_i$ is assumed to be drawn from a given constellation set say $\mathbb{A}$, for example, 16-QAM, 64-QAM, etc. Without loss of generality, the users are assumed to be synchronized in time during their transmission to the base station. For simplicity, we consider a Rayleigh flat fading channel between the $i$th user and $j$th receive antenna denoted as $h_{ji}$, for $j = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, K$, distributed as complex normal with mean zero and variance unity, i.e., $\sim \mathcal{CN}(0, 1)$. The received vector $\mathbf{y}$ after demodulation and sampling at the receiver end can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{n}$ is the additive white Gaussian noise (AWGN) vector with each of its elements $n_j$, for $j = 1, 2, \ldots, M$, is distributed as complex normal with mean zero and variance $\sigma^2$, i.e., $\sim \mathcal{CN}(0, \sigma^2)$. The average received SNR at the receiver can be computed as $\frac{KE_x}{\sigma^2}$. In (3.1), each element $y_j$ of $\mathbf{y}$, for $j = 1, 2, \ldots, M$, is the received information at the $j$th receive antenna which consists of combination of symbols

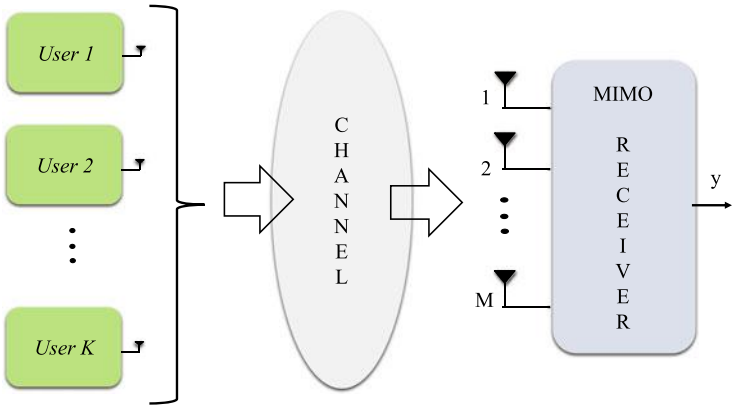



**Fig. 3.1** Massive MIMO system

from each transmit antenna modified by the channel state $h_{ji}$ and the AWGN noise $n_j$ as

$$y_j = \sum_{i=1}^{K} h_{ji} x_i + n_i, \quad \forall\, j = 1, 2, \ldots, N. \tag{3.2}$$

Due to the interference of symbols from different users at each receive antenna and the presence of additive noise, it becomes tough to separate and detect the symbol from each user at the receiver. Therefore, reliable symbol detection in the multiple antenna system is a challenging issue. Next, we discuss the mathematical formulation of the detection problem in massive MIMO systems.

### *3.2.2 Problem Formulation*

The symbol detection problem in massive MIMO systems is to find the best symbol vector $\mathbf{x}$ given the received vector $\mathbf{y}$ and the channel state information (CSI) $\mathbf{H}$. To begin with, let us assume that the elements $x_i$'s of symbol vector $\mathbf{x}$ are drawn from an $N$-QAM constellation set, and therefore, the total number of possible transmit symbol vectors is $(\log_2 N)^K$. To achieve optimal performance, one needs to perform joint search over all the possible combinations of the transmitted symbols. Maximum likelihood (ML) detection is one such technique where an exhaustive search is performed over the set, say $\mathbb{S}$, of $(\log_2 N)^K$ all the possible symbol vectors and achieves optimal BER performance. Basically, the symbol vector which minimizes the ML cost metric is selected to be the optimal solution. The ML solution can be written as

$$\widehat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathbf{x}\in\mathbb{S}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{3.3}$$

Though ML detection achieves optimal BER performance, its exponential nature of computational complexity makes it practically infeasible for the systems with a large number of antennas. Sphere decoding (SD) is another such technique which achieves close to ML solution but is practically limited to a specific number of transmit antennas.

In massive MIMO systems, due to excessive base station antennas, linear detectors such as ZF/MMSE achieve near-optimal BER performance as a consequence of channel hardening. Therefore, the research focus in such systems is toward developing or modifying the linear detection techniques. Next, we discuss the ZF detector and the MMSE detector, and also, the challenges associated with them.

**Zero-Forcing Detector**

In ZF detection, a filter matrix $\mathbf{W}$ is determined such that it removes the impairments due to the channel from the received vector $\mathbf{y}$. Basically, it is the least square solution which minimizes $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ without performing an exhaustive search over all the possible transmit symbol vectors. The ZF filter $\mathbf{W} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ which is applied over the received vector $\mathbf{y}$ in order to estimate the transmitted symbols. The ZF solution can be written as

$$\mathbf{x}_{\text{ZF}} = Q\left[\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{y}\right], \tag{3.4}$$

where $(\cdot)^H$ and $(\cdot)^{-1}$ denotes the hermitian and inverse of a matrix, respectively. ZF suffers from the problem of noise enhancement which effects its BER performance. MMSE detector which is discussed in next section overcomes the noise enhancement issue and achieve better BER performance over ZF detector.

**Minimum Mean-Squared Error Detector**

The MMSE detection as the name suggests is based on minimizing the mean-squared error between the transmit vector $\mathbf{x}$ and a linear transformation of the received vector $\mathbf{y}$. The linear transformation matrix say $\mathbf{W}$ can be determined as

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}} \|\mathbf{x} - \mathbf{W}\mathbf{y}\|^2. \tag{3.5}$$

The solution to (3.5) is

$$\widehat{\mathbf{W}} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma^2}{E_x}\mathbf{I}_K\right)^{-1}\mathbf{H}^H. \tag{3.6}$$

The MMSE solution can be computed as

$$\mathbf{x}_{\text{MMSE}} = Q\left[\left(\mathbf{H}^H\mathbf{H} + \frac{\sigma^2}{E_x}\mathbf{I}_K\right)^{-1}\mathbf{H}^H\mathbf{y}\right]. \tag{3.7}$$

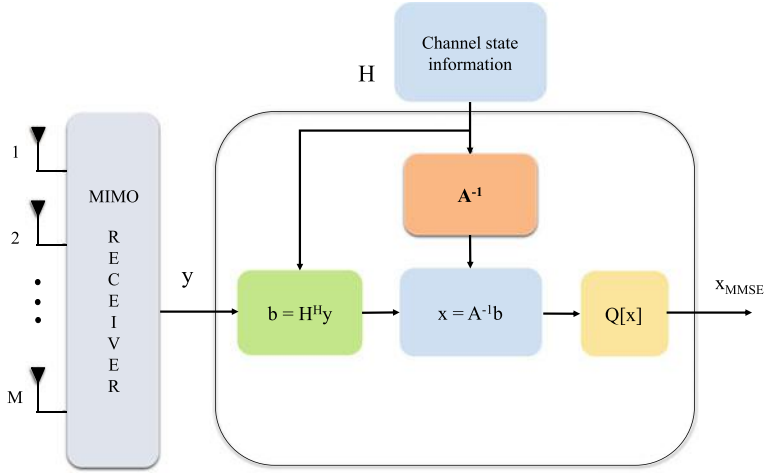



**Fig. 3.2** MMSE detector

where $Q[\cdot]$ is the quantization operation which maps the soft values to the nearest point in the constellation set $\mathbb{S}$. The MMSE solution can also be written as $\mathbf{x}_{\text{MMSE}} = \mathbf{A}^{-1}\mathbf{b}$, where $\mathbf{A} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma^2}{E_x}\mathbf{I}_K\right)$ and $\mathbf{b} = \mathbf{H}^H\mathbf{y}$ as depicted in Fig. 3.2.

**Algorithm 1** MMSE detector

1: **Input:** $\mathbf{y}, \mathbf{H}, K, M, \sigma^2$
2: Compute matrix $\mathbf{A} = \mathbf{G} + \frac{\sigma^2}{E_x}\mathbf{I}_K$ where $\mathbf{G} = \mathbf{H}^H\mathbf{H}$
3: Compute the inverse of matrix $\mathbf{A}$ i.e. $\mathbf{A}^{-1}$
4: Compute the MF output of $\mathbf{y}$ : $\mathbf{b} = \mathbf{H}^H\mathbf{y}$
5: Compute the MMSE estimate $\mathbf{x}_{\text{MMSE}} = Q\left[\mathbf{A}^{-1}\mathbf{b}\right]$
6: **Output:** $\mathbf{x}_{\text{MMSE}}$

### Challenges with MMSE

The inversion of a matrix $\mathbf{A}$ makes the MMSE detection challenging as it requires cubic complexity with respect to the number of users which ultimately restricts the reduction in architectural complexity and the signal processing delay required for possible application in 5G and beyond systems. It can easily be observed that the problem of finding MMSE solution is nothing but solving a set of linear equations given by $\mathbf{Ax} = \mathbf{b}$. Therefore, several alternate techniques which do not involve the inverse of matrix such as Neumann series expansion, Newton iteration, Jacobi iteration, etc., are available in the linear algebra literature, which are successfully applied for massive MIMO detection. These techniques are discussed later in Sect. 3.4 in detail.

## 3.3 Massive MIMO Detection Techniques

The symbol detection problem in massive MIMO is to find the solution to the expression $\mathbf{Ax} = \mathbf{b}$ without performing the inverse of a matrix $\mathbf{A}$. To find a feasible solution, there are two types of techniques which can be broadly classified as (a) approximate matrix inversion methods and (b) iterative methods.

### *3.3.1 Approximate Matrix Inversion Methods*

In these method, approximate matrix inverse approaches are discussed, wherein an approximation of regularised Gram matrix $\mathbf{A}$ is computed and used for symbol detection for massive MIMO systems.

**Neumann Series Expansion-Based Detection**

This technique is based on the concept of *polynomial expansion* where the inverse of a matrix is expanded as a sum of infinite number of terms. Each term involves either matrix-matrix multiplications or matrix-vector multiplications or both. The $P$th order Neumann series approximation (consists of only first $P$ terms of an infinite series) of the inverse of a matrix can be written as

$$\mathbf{A}_P^{-1} = \sum_{m=0}^{P-1} \left(\mathbf{W}\left(\mathbf{W}^{-1} - \mathbf{A}\right)\right)^m \mathbf{W} \tag{3.8}$$

where $\mathbf{W}$ denotes an initial low-complexity estimate of the inverse. The series in (3.8) converges to the exact inverse with $P \to \infty$ with $\mathbf{W}$ satisfying the condition

$$\lim_{m\to\infty} (\mathbf{I} - \mathbf{WA})^m = \mathbf{O}. \tag{3.9}$$

The application of Neumann series expansion for massive MIMO detection is based on the fact that in massive MIMO , due to channel hardening, the regularised Gram matrix $\mathbf{A}$ is a diagonally dominant matrix, i.e., the diagonal elements are significantly higher than the off-diagonal elements. Therefore, the matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{D} + \mathbf{E}$, where $\mathbf{D}$ is the diagonal matrix and $\mathbf{E}$ is the hollow matrix. The matrix $\mathbf{D}^{-1}$ which is inverse of the diagonal elements of $\mathbf{D}$ is used as a low-complexity initial estimate of the inverse in Neumann series expansion-based detection. The $P$th order series expansion can be written as

$$\mathbf{A}_P^{-1} = \sum_{m=0}^{P-1} \left(-\mathbf{D}^{-1}\mathbf{E}\right)^m \mathbf{D}^{-1}. \tag{3.10}$$

Using $\mathbf{A}_P^{-1}$, the symbols can be detected as

$$\widehat{\mathbf{x}}_{NS} = Q\left[\mathbf{A}_P^{-1}\mathbf{b}\right] = Q\left[\sum_{m=0}^{P-1}\left(-\mathbf{D}^{-1}\mathbf{E}\right)^m \mathbf{D}^{-1}\mathbf{b}\right]. \tag{3.11}$$

Algorithm 2 shows the pseudo-code of Neumann series expansion-based massive MIMO detection.

**Algorithm 2** Neumann series-based detector

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $K$, $M$, $\sigma^2$
2: Compute matrix $\mathbf{A} = \mathbf{G} + \frac{\sigma^2}{E_x}\mathbf{I}_K$ where $\mathbf{G} = \mathbf{H}^H\mathbf{H}$
3: Compute the inverse of matrix $\mathbf{A}$ i.e. $\mathbf{A}^{-1}$
4: Compute the MF output of $\mathbf{y}$ : $\mathbf{b} = \mathbf{H}^H\mathbf{y}$
5: Compute the MMSE estimate $\mathbf{x}_{\text{MMSE}} = Q\left[\mathbf{A}^{-1}\mathbf{b}\right]$
6: **Output:**$\mathbf{x}_{\text{MMSE}}$

**Newton Iteration-Based Detection**

In Newton iteration-based approximate inverse method, an initial estimate of the inverse of a matrix is refined iteratively. The iterative update of inverse is given by

$$\mathbf{A}_{m+1}^{-1} = \mathbf{A}_m^{-1}\left(2\mathbf{I} - \mathbf{A}\mathbf{A}_m^{-1}\right). \tag{3.12}$$

The approximate inverse in (3.12) converges quadratically to the exact inverse if the initial estimate satisfies

$$\|\mathbf{I} - \mathbf{A}\mathbf{A}_0^{-1}\| \leq 1. \tag{3.13}$$

Similar to the Neumann series expansion, the initial estimate of inverse is $\mathbf{D}^{-1}$ which is refined iteratively. The detected symbols using $L$ iterations can be written as

$$\widehat{\mathbf{x}}_{NI} = Q\left[\mathbf{A}_L^{-1}\mathbf{b}\right]. \tag{3.14}$$

The pseudo-code of the Newton iteration-based massive MIMO detection is shown in Algorithm 3.

**Algorithm 3** Newton iteration-based detector

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $K$, $M$, $\sigma^2$
2: Compute matrix $\mathbf{A} = \mathbf{G} + \frac{\sigma^2}{E_x}\mathbf{I}_K$ where $\mathbf{G} = \mathbf{H}^H\mathbf{H}$
3: Compute the inverse of matrix $\mathbf{A}$ i.e. $\mathbf{A}^{-1}$
4: Compute the MF output of $\mathbf{y}$ : $\mathbf{b} = \mathbf{H}^H\mathbf{y}$
5: Compute the MMSE estimate $\mathbf{x}_{\text{MMSE}} = Q\left[\mathbf{A}^{-1}\mathbf{b}\right]$
6: **Output:**$\mathbf{x}_{\text{MMSE}}$

### 3.3.2 Iterative Methods

In this section, we discuss the iterative methods for symbol detection in massive MIMO systems. The key idea behind iterative methods is that a low-complexity initial solution is refined iteratively for a few iterations to obtain a better solution. The initial solution is refined in such a way that the final solution achieves superior BER performance without much efforts in the computational complexity. There are two main components in the iterative methods which are

i Initial solution
ii Iterative update

The pseudo-code of each iterative detection technique is more or less similar and can be shown as Algorithm 4. A good initial solution determines how fast the algorithm

**Algorithm 4** Iterative detection

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: Compute initial solution $\mathbf{x}_0$
3: Update the initial solution iteratively
4: **Output:**Final solution after $L$ iterations $\mathbf{x}_L$

converges to the final solution. However, the convergence of the algorithm highly depends on the iterative update scheme. Different algorithms in the literature consider different update mechanism and initial solution which is discussed next.

**Richardson Iteration-Based Detection**
In this technique, the special property of the regularised Gram matrix which is positive semi-definiteness is exploited for performing Richardson iterations in order to detect the symbols efficiently. The Richardson iterations can be written as

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \rho\left(\mathbf{b} - \mathbf{A}\mathbf{x}_m\right), \tag{3.15}$$

where $\rho$ is the relaxation parameter in Richardson iterations. For convergence, the relaxation parameter $\rho$ should be between $0 < \rho < 2/\lambda_{max}$, where $\lambda_{max}$ is the maximum eigenvalue of the matrix $\mathbf{A}$. The performance of Richardson method also depends on the initial solution, and therefore, a low-complexity initial solution is obtained as

$$\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}. \tag{3.16}$$

The refined solution after $L$ iterations is considered as the output of the detector. The pseudo-code of the Richardson-based detection algorithm is discussed in Algorithm 5.

**Jacobi Iteration-Based Detection** In this technique, Jacobi iterations are used for refining the initial solution. The Jacobi iterations can be written as

**Algorithm 5** Richardson iteration-based detection

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: Compute initial solution : $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}$
3: **for** $(m = 1, m++, m \leq L)$ **do**
4: Update the initial solution iteratively as : $\mathbf{x}_{m+1} = \mathbf{x}_m + \rho\,(\mathbf{b} - \mathbf{A}\mathbf{x}_m)$
5: **end for**
6: **Output:**Final solution after $L$ iterations $Q[\mathbf{x}_L]$

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \mathbf{D}^{-1}\,(\mathbf{b} - \mathbf{A}\mathbf{x}_m)\,. \tag{3.17}$$

In order to accelerate the convergence of the algorithm, an iteration of steepest descent is hybridised with Jacobi iteration which is given as

$$\mathbf{x}_2 = \mathbf{x}_0 + v\mathbf{r}_0 + \mathbf{D}^{-1}(\mathbf{r}_0 - v\mathbf{g}_0), \tag{3.18}$$

where $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, $v = \frac{\mathbf{r}_0^H \mathbf{r}_0}{(\mathbf{A}\mathbf{r}_0)^H \mathbf{r}_0}$ and $\mathbf{g}_0 = \mathbf{A}\mathbf{r}_0$. The pseudo-code of hybrid Jacobi iteration and steepest descent iteration is given in Algorithm 6.

**Algorithm 6** Hybrid Jacobi iteration and steepest descent-based detection

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: Compute initial solution : $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}$
3: Perform the first joint iteration as $\mathbf{x}_2 = \mathbf{x}_0 + v\mathbf{r}_0 + \mathbf{D}^{-1}(\mathbf{r}_0 - v\mathbf{g}_0)$ where $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, $v = \frac{\mathbf{r}_0^H \mathbf{r}_0}{(\mathbf{A}\mathbf{r}_0)^H \mathbf{r}_0}$ and $\mathbf{g}_0 = \mathbf{A}\mathbf{r}_0$
4: **for** $(m = 3, m++, m \leq L)$ **do**
5: Update the initial solution iteratively as : $\mathbf{x}_{m+1} = \mathbf{x}_m + \mathbf{D}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_m)$
6: **end for**
7: **Output:**Final solution after $L$ iterations $Q[\mathbf{x}_L]$

### Gauss–Seidel-Based Detection

This is another iterative technique proposed in the literature for solving the set of linear equations. As discussed earlier in Sect. 3.2, MMSE detection in massive MIMO is nothing but finding a solution to the set of linear equation with noisy measurements. However, due to channel hardening, the MMSE solution turns out to be a near-optimal one. Therefore, the techniques which can find the solution without explicitly computing the matrix inverse are of high interest. Gauss–Seidel method is one such technique which computes the solution in an iterative manner. Similar to other iterative techniques such as Richardson and Jacobi methods, this technique also exploits the positive semi-definite property of regularised Gram matrix $\mathbf{A}$. The key idea behind the Gauss–Seidel technique is to represent the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{L}^H, \tag{3.19}$$

where $\mathbf{D}$, $\mathbf{L}$ and $\mathbf{L}^H$ denote the diagonal matrix, strictly lower triangular matrix and strictly upper triangular matrix of $\mathbf{A}$, respectively. The Gauss–Seidel iterations are then used for refining the initial solution as

$$\mathbf{x}_{m+1} = (\mathbf{D} + \mathbf{L})^{-1} \left(\mathbf{b} - \mathbf{L}^H \mathbf{x}_m\right). \tag{3.20}$$

The pseudo-code of Gauss–Seidel-based detection is shown in Algorithm 7.

**Algorithm 7** Gauss–Seidel iteration-based detector

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: Compute initial solution : $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}$
3: **for** $(m = 1, m++, m \leq L)$ **do**
4: Update the initial solution iteratively as : $\mathbf{x}_{m+1} = (\mathbf{D} + \mathbf{L})^{-1} \left(\mathbf{b} - \mathbf{L}^H \mathbf{x}_m\right)$
5: **end for**
6: **Output:**Final solution after $L$ iterations $Q[\mathbf{x}_L]$

### Conjugate Gradient-Based Detection

Conjugate gradient (CG) is another way of solving a set of linear equations (Hestenes and Stiefel 1952) where the search for the solution is performed in the conjugate direction with a step movement toward the better solution iteratively. In CG-based massive MIMO detection (Yin 2014), the search direction and the step size for the movement are determined first and the solution is updated by moving a step in the search direction next. The detailed steps involved in the CG method are shown in Algorithm 8.

**Algorithm 8** Conjugate gradient-based detection

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: **Initialization:** $\mathbf{b} = \mathbf{H}^H\mathbf{y}$, $\mathbf{v}_0 = \mathbf{O}$, $\mathbf{r}_0 = \mathbf{b}$, $\mathbf{p}_0 = \mathbf{r}_0$
3: **for** $(m = 1, m++, m \leq L)$ **do**
4: $\mathbf{e}_{m-1} = \mathbf{A}\mathbf{p}_m - 1$
5: $\alpha_m = \|\mathbf{r}_{m-1}/\mathbf{p}_{m-1}^H \mathbf{e}_{m-1}\|$
6: $\mathbf{v_m} = \mathbf{v}_{m-1} + \alpha_m \mathbf{p}_{m-1}$
7: $\mathbf{r}_m = \mathbf{r}_{m-1} - \alpha_m \mathbf{e}_{m-1}$
8: $\beta_m = \|\mathbf{r}_m\|^2 / \|\mathbf{r}_{m-1}\|^2$
9: $\mathbf{p}_m = \mathbf{r}_m + \beta_m \mathbf{p}_{m-1}$
10: **end for**
11: **Output:**Final solution after $L$ iterations $\mathbf{x}_L = Q[\mathbf{v}_L]$

### Iterative Sequential Detection

The iterative detection techniques discussed so far are based on the conventional iterative methods which solve the set of linear equations without computing matrix inversion. On the other hand, ISD is inspired from the successive interference cancelation (SIC)-based MIMO detection where symbol transmitted from each user is detected successively. The key idea in ISD is to detect the symbol corresponding to

each user while nulling the interference from all the users. The ISD algorithm starts with an initial solution $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}$. Let $x_m^{(i)}$ denote the symbol detected, corresponding to the $i$th user in the $m$th iteration. Now, let us consider the case of detecting the symbol corresponding to the $j$th user in the $(m+1)$th iteration which can be written as

$$\mathbf{r}_{m+1}^{(j)} = \mathbf{y} - \mathbf{H}\phi_m + \mathbf{h}_j x_m^{(j)} \tag{3.21}$$

where $\phi_{m+1} = (x_{m+1}^{(1)}, x_{m+1}^{(2)}, \ldots, x_{m+1}^{(j-1)}, x_m^{(j)}, x_m^{(j+1)}, \ldots, x_{m+1}^{(K)})$ is the symbol update vector and $\mathbf{h}_j$ is the $j$th column of channel matrix $\mathbf{H}$. To detect the symbol for the $j$th user, matched filtering of (3.21) is performed followed by quantization operation as

$$\mathbf{x}_{m+1}^{(j)} = Q\left[\frac{\mathbf{h}_j^H \mathbf{r}_{m+1}^{(j)}}{\|\mathbf{h}_j\|}\right]. \tag{3.22}$$

The detected symbol is then updated in the symbol update vector $\phi_{m+1}$. This operations are performed iterative for each user till the number of iterations not completed. Algorithm 9 describes the pseudo-code of ISD algorithm.

**Algorithm 9** Iterative sequential detection

1: **Input:** $\mathbf{y}$, $\mathbf{H}$, $\mathbf{A}$, $\mathbf{G}$ $K$, $M$, $\sigma^2$
2: Compute initial solution : $\mathbf{x}_0 = \mathbf{D}^{-1}\mathbf{b}$
3: **for** $(m = 1, m++, m \leq L)$ **do**
4: **for** $j = 1, j++, j \leq K$ **do**
5: Perform interference nulling : $\mathbf{r}_{m+1}^{(j)} = \mathbf{y} - \mathbf{H}\phi_m + \mathbf{h}_j x_m^{(j)}$
6: Detect symbol : $\mathbf{x}_{m+1}^{(j)} = Q\left[\frac{\mathbf{h}_j^H \mathbf{r}_{m+1}^{(j)}}{\|\mathbf{h}_j\|}\right]$
7: Update symbol vector : $\phi_{m+1} = (x_{m+1}^{(1)}, x_{m+1}^{(2)}, \ldots, x_{m+1}^{(j-1)}, x_m^{(j)}, x_m^{(j+1)}, \ldots, x_{m+1}^{(K)})$
8: **end for**
9: **end for**
10: **Output:**Final solution after $L$ iterations $\mathbf{x}_L = \phi_L$

## 3.4 Simulation Results and Discussions

In this section, simulation results on BER performance versus average received SNR and the computational complexity (in terms of total number of real time operations) are discussed for different detection techniques. For drawing useful insights, simulations are carried for various massive MIMO environments. The simulations for plotting BER curve are carried out in MATLAB, and upto 1000 errors are counted for averaging the BER over random noise and channel samples (Fig. 3.3).

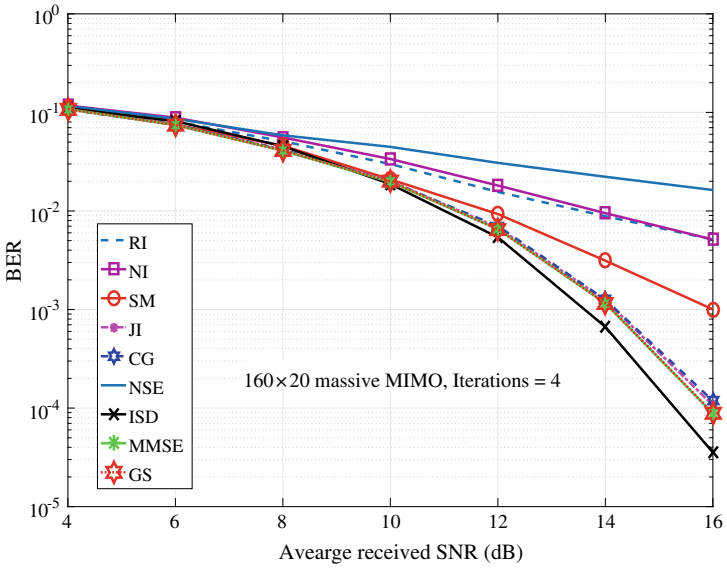



**Fig. 3.3** BER performance comparison of different detection techniques for 160 × 20 massive MIMO system with 64-QAM modulation

### *3.4.1 Bit Error Rate Comparison*

Figures 3.1 and 3.2 present the BER performance comparison of various massive MIMO systems with 64-QAM modulation. The detection techniques such as Neumann series expansion (NSE), Newton iteration (NI), Stair matrix (SM)-based detection, Conjugate gradient (CG), Richardson iteration (RI), Gauss–Siedel (GS), MMSE and iterative sequential detection (ISD) are used for comparison. Further, during simulations, two iterations are performed for NSE, NI and SM, and four iterations are performed for RI, CG, GS and ISD. In Fig. 3.1, 160 × 20 massive MIMO system is used for simulating the BER versus average received SNR. It is observed that the ISD algorithm achieves best BER performance among all the detectors, which is mainly due to the interference cancelation-based detection of symbols corresponding to each user. Moreover, ISD also exploits the channel hardening property while detecting the symbols. CG, JI and GS achieve performance close to the MMSE performance, whereas NI, NSE, RI and SM are far inferior as compared to the MMSE and ISD detectors. Another reason behind the detectors other than ISD are limited in performance is that these algorithms are devised aiming to solve the linear system of equations without matrix inversion or using approximate inversion techniques (Fig. 3.4).

In Fig. 3.2, 160 × 40 massive MIMO system is used for simulating the BER performance. Similar observations as in Fig. 3.1 can be drawn from Fig. 3.2 on the BER performance. However, one key observation from Fig. 3.2 is that the performance

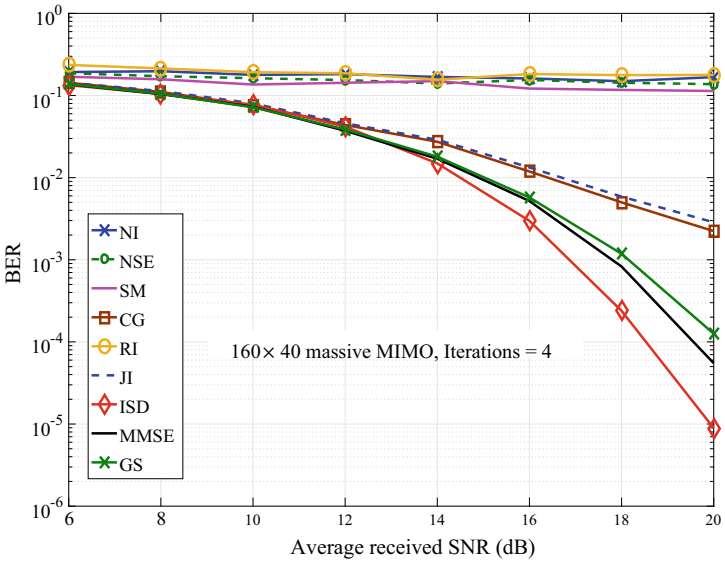



**Fig. 3.4** BER performance comparison of different detection techniques for 160 × 40 massive MIMO system with 64-QAM modulation

of NI, NSE, RI and SM degrades with the increase in the number of users in the system. It is because NI, NSE and SM involve approximate inverse of the matrix and are limited by the number of iterations due to the computational complexity of matrix-matrix multiplications. The performance of CG, JI, and GS become inferior as compared to the MMSE detector when the number of users is increased from 20 to 40 users. The performance of ISD is still superior as compared with the MMSE and other detectors, and therefore, ISD can be used for systems with comparatively large number of users.

### *3.4.2 Computational Complexity*

In this section, we discuss the computational complexity of various detection techniques in terms of the total number of real-valued operations, i.e., number of additions and multiplications. For the comparison of computational complexity, we ignore the computations of common terms such as **A**, **G** and **b**. The number of iterations for NS and NI is assumed to be two iterations as the matrix-matrix multiplications dominate for iterations greater than two which leads to the computations to be of the order of $M^3$. For other detection techniques such as RI, JI, GS, CG and ISD, $p$ iterations are considered for computing the additions and multiplications. From Table 3.1, it can easily be observed that the complexity of all the iterative algorithms is approximately

**Table 3.1** Computational complexity of different detection techniques

| Detection techniques | Multiplications | Additions |
|---|---|---|
| Neumann series (2 iterations) | $12M^2 + 2M$ | $8M^2 - 2M$ |
| Newton iteration (2 iterations) | $12M^2 + 2M$ | $8M^2 - 2M$ |
| Richardson iteration ($p$ iterations) | $4pM^2 + 2pM$ | $4pM^2 + 2pM$ |
| Jacobi iteration ($p$ iterations) | $4(p+1)M^2 + 2(p+4)M$ | $4pM^2 + 4pM$ |
| Gauss–Siedel ($p$ iterations) | $4pM^2 + 2pM$ | $10pM^2 - 3pM$ |
| Conjugate gradient ($p$ iterations) | $4pM^2 + 8pM$ | $4pM^2 + 6pM$ |
| Iterative sequential detection ($p$ iterations) | $4pM^2 + 2(2p+1)M$ | $4pM^2 + 2pM$ |

the same. However, ISD dominates in terms of the BER performance, and therefore, ISD achieves superior performance-complexity trade-off over the other detection techniques.

## 3.5 Conclusion and Future Scope

In this chapter, we presented a crucial issue of symbol detection in the uplink of massive MIMO systems for its practical feasibility in 5G and beyond systems. The chapter briefly discussed the motivation behind research in massive MIMO detection. Problem formulation for massive MIMO detection and various detection techniques such techniques based on Neumann series expansion, Newton iteration, Jacobi iteration, Richardson iteration, Gauss–Seidel iteration, Conjugate gradient, and iterative sequential detection are discussed in detail. Moreover, the simulation results on BER performance and computational complexity are also discussed for drawing some useful insights.

Though the existing techniques achieve close-to-MMSE or superior (in case of ISD) BER performance, there is still a lot of scope for further research in the field. Most of the existing detection techniques for massive MIMO systems are based on solving the set of linear equations through approximate inverse or iterative methods which limits their achievable BER performance. Moreover, when the number of users increases in the system, i.e., when the ratio of number of base station antennas to the number of users becomes less than 8, the BER performance of the existing techniques starts degrading. Therefore, there is a need to develop nonlinear heuristic techniques such as bio-inspired optimization, neighborhood search algorithms, message passing techniques in the context of massive MIMO detection where computational complexity plays an important role in practical design. Interestingly, when the loading factor decreases beyond five, there are no such techniques which can provide an optimal solution (the lower bound on BER is still an open research problem). This results in strong motivation for further research in the field.

**Acknowledgements** The authors would like to thank Science and Engineering Research Board (SERB) (File no. SRG/2019/000654) of Department of Science and Technology (DST), Government of India, for supporting the research work. Authors would also like to thank Visvesvaraya Ph.D. scheme of Ministry of Electronics and Information Technology, Government of India, for their support.

## References

Andrews JG et al (2014) What will 5G be? IEEE J Sel Areas Commun 32(6):1065–1082

Barbero LG, Thompson JS (2008) Fixing the complexity of the sphere decoder for MIMO detection. IEEE Trans Wireless Commun 7:2131–2142

Chockalingam A, Rajan BS (2014) Large MIMO systems. Cambridge University Press, New York

Dai L, Gao X, Su X, Han S, Wang Z (2015) Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems. IEEE Trans Veh Technol 64(10):4839–4845

Datta A, Mandloi M, Bhatia V (2018) Hybrid pseudo-stationary iterative detection algorithm for uplink massive MIMO systems. In: 2018 IEEE international conference on advanced networks and telecommunications systems (ANTS), pp 1–6

Fadlallah Y, Aïssa-El-Bey A, Amis K, Pastor D, Pyndiah R (2015) New iterative detector of MIMO transmission using sparse decomposition. IEEE Trans Veh Technol 64(8):3458–3464

Fukuda RM, Abrao T (2019) Linear, quadratic, and semidefinite programming massive MIMO detectors: reliability and complexity. IEEE Access 7:29506–29519

Gao X, Dai L, Ma Y, Wang Z (2014) Low-complexity near-optimal signal detection for uplink large-scale MIMO systems. Electron Lett 50(18):1326–1328

Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. J Res Natl Bur Stan 49(6):409–436

Hu Y, Wang Z, Gaol X, Ning J (2014) Low-complexity signal detection using CG method for uplink large-scale MIMO systems. In: 2014 IEEE international conference on communication systems, pp 477–481

Jiang F, Li C, Gong Z (2017) A low complexity soft-output data detection scheme based on Jacobi method for massive MIMO uplink transmission. In: 2017 IEEE international conference on communications (ICC), pp 1–5

Jiang F, Li C, Gong Z, Su R (2018) Stair matrix and its applications to massive MIMO uplink data detection. IEEE Trans Commun 66(6):2437–2455

Larsson EG (2009) MIMO detection methods: how they work [lecture notes]. IEEE Sig Process Mag 26(3):91–95

Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2):186–195

Li P, De Lamare RC, Fa R (2011) Multiple feedback successive interference cancellation detection for multiuser MIMO systems. IEEE Trans Wireless Commun 10(8):2434–2439

Lu L, Li GY, Swindlehurst AL, Ashikhmin A, Zhang R (2014) An overview of massive MIMO: benefits and challenges. IEEE J Sel Top Sig Process 8(5):742–758

Mandloi M, Bhatia V (2017b) Layered gibbs sampling algorithm for near-optimal detection in large-MIMO systems. In: IEEE wireless communications and networking conference (WCNC). San Francisco, CA, pp 1–6

Mandloi M, Bhatia V (2016) Low-complexity near-optimal iterative sequential detection for uplink massive MIMO systems. IEEE Commun Lett 21(3):568–571

Mandloi M, Bhatia V (2017a) Error recovery based low-complexity detection for uplink massive MIMO systems. IEEE Wireless Commun Lett 6(3):302–305

Mandloi M, Hussain MA, Bhatia V (2017) Improved multiple feedback successive interference cancellation algorithms for near-optimal MIMO detection. IET Commun 11(1):150–159
Minango J, de Almeida C (2018) Low complexity zero forcing detector based on Newton-Schultz iterative algorithm for massive MIMO systems. IEEE Trans Veh Technol 67(12):11759–11766
Mussi AM, Abrão T (2018) Multiple restarts mixed Gibbs sampling detector for large-scale antenna systems. IET Signal Proc 13(3):273–285
Peng X, Wu W, Sun J, Liu Y (2015) Sparsity-boosted detection for large MIMO systems. IEEE Commun Lett 19(2):191–194
Qin X, Yan Z, He G (2015) A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems. IEEE Commun Lett 20(2):276–279
Rusek F et al (2013) Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Sig Process Mag 30(1):40–60
Sah AK, Chaturvedi AK (2017) An unconstrained likelihood ascent based detection algorithm for large MIMO systems. IEEE Trans Wireless Commun 16(4):2262–2273
Silva GMF, Marinello Filho JC, Abrao T (2018) Sequential likelihood ascent search detector for massive MIMO systems. AEU Int J Electron Commun 96:30–39
Tang C, Liu C, Yuan L, Xing Z (2016) High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO. IEEE Commun Lett 20(3):490–493
Viterbo E, Boutros J (1999) A universal lattice code decoder for fading channels. IEEE Trans Inf Theory 45:1639–1642
Wu M, Yin B, Wang G, Dick C (2014) Large-scale MIMO detection for 3GPP LTE: algorithms and FPGA implementations. IEEE J Sel Top Sig Process 8(5):916–929
Wu Z, Zhang C, Xue Y, Xu S, You X (2016) Efficient architecture for soft-output massive MIMO detection with Gauss-Seidel method. In: 2016 IEEE international symposium on circuits and systems (ISCAS), pp 1886–1889
Yang S, Hanzo L (2015) Fifty years of MIMO detection: the road to large-scale MIMOs. IEEE Commun Surv Tuts 17(4):1941–1988
Yin B, Wu M, Cavallaro JR, Studer C (2014) Conjugate gradient-based soft output detection and precoding in massive MIMO systems. In: 2014 IEEE global conference on communications (Globecom), pp 3696–3701
Yin B, Wu M, Studer C, Cavallaro JR, Dick C, (2013) Implementation trade-offs for linear detection in large-scale MIMO systems. In: 2013 IEEE international conference on acoustics, speech and signal processing, pp 2679–2683
Zeng J, Lin J, Wang Z (2018) Low complexity message passing detection algorithm for large-scale MIMO systems. IEEE Wirel Commun Lett 7(5):708–711
Zhou Q, Ma X (2013) Element-based lattice reduction algorithms for large MIMO detection. IEEE J Sel Areas Commun 31(2):274–286
Zhu D, Li B, Liang P (2015) On the matrix inversion approximation based on Neumann series in massive MIMO systems. In: 2015 IEEE international conference on communications (ICC), June 2015, pp 1763–1769

# Chapter 4
# Index Modulation Techniques for 5G and Beyond Wireless Systems


**Manish Mandloi, Arijit Datta, and Vimal Bhatia**

**Abstract** Index modulation (IM) is one of the emerging techniques for enabling beyond 5G (B5G) wireless communications. It has the potential to meet the stringent energy efficiency (EE) and spectral efficiency (SE) requirements of B5G systems with better error rate performance over the conventional techniques. IM relaxes the need for activating all the resources at the transmitter to transmit the information, thereby allowing low-complexity transmitter architecture designs. The key idea behind IM is to encode the information in the indices of the available resources at the transmitter such as antenna, sub-carriers in orthogonal frequency division multiplexing, time-slots, and radio frequency (RF) mirrors. Massive-MIMO is another such promising technique which provides unprecedented growth in both EE and SE for B5G wireless systems. Spatial multiplexed massive-MIMO is shown to achieve the unbelievable capacity gains over the conventional MIMO systems. However, achieving such gains requires dedicated signal processing resources for each antenna which increase the cost, area, and power-requirement at the transmitter. Interestingly, through IM in massive-MIMO, exceptionally high EE and SE can be achieved with minimal use of the available resources. Recently, multi-dimensional IM (MIM), wherein multiple resources are indexed simultaneously during transmission to enhance the SE further, has attracted researchers and experts from both academia and industry. In this chapter, we discuss different IM and MIM techniques, their representations, and advantages in B5G communications. In particular, we discuss spatial modulation, generalized spatial modulation, media-based modulation, and their possible combi-

M. Mandloi (✉)
Department of Electronics and Telecommunication Engineering,
SVKM's NMIMS Shirpur Campus, Shirpur, Maharashtra 425405, India
e-mail: manish.mandloi@nmims.edu

A. Datta · V. Bhatia
Discipline of Electrical Engineering,
Indian Institute of Technology Indore, Indore, Madhya Pradesh 453552, India
e-mail: phd1601102003@iiti.ac.in

V. Bhatia
e-mail: vbhatia@iiti.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_4

nations as MIM in detail. We also shed the light on the maximum-likelihood detection of information symbols in such systems.

**Keywords** Index modulation · Massive-MIMO · Radio frequency mirrors · Maximum-likelihood detection · Bit error rate · Spectral efficiency

## 4.1 Introduction

The exponential increase in the number of mobile subscribers across the globe with advanced data thirsty applications at each user terminal demands fully connected, reliable, and high data rate wireless services (Boccardi 2014). Over the past two decades, the wireless community has seen an upshift of two generations over the second generation (2G) wireless standards to overcome the data rate requirements with an increased number of mobile users serviced simultaneously. With each generation shift (i.e., from 2G to 3G and from 3G to 4G), the overall spectral efficiency is enhanced almost 1000 times. The next generation of wireless systems (5G) are expected to roll out in the market by end 2020 with three-dimensional objectives identified by the international telecommunication union (ITU): (1) enhanced mobile broadband (eMBB), (2) massive machine type communication, and (3) ultra-reliable low latency communication (uRLLC) (Andrews 2014; Boccardi 2014). It is anticipated that the successful implementation of 5G networks will provide the wireless service between any-thing, any-time, and any-where. However, to achieve the key objectives of 5G and beyond wireless systems is a challenging task. To meet the fiber rate communication with high quality of service (QoS), wireless researchers and engineers have come up with multiple physical layer (PHY layer) concepts such as massive multiple-input multiple-output (MIMO) (Lu et al. 2014; Rusek 2013), non-orthogonal multiple access (NOMA) (Dai 2018; Ding 2017), and index modulation (IM) (Basar 2016; Wen et al. 2017). Moreover, research efforts are being put in place to utilize these concepts efficiently in wireless systems in order to achieve enhanced spectral efficiency (SE) and energy efficiency (EE).

Massive-MIMO is one of the promising techniques for achieving high SE and EE in wireless systems (Lu et al. 2014). In massive-MIMO, the large antenna array is used at the base station (BS) which serves comparatively less (few tens) number of users. This results in a substantial processing gain which provides enhanced performance gain over the conventional MIMO systems (Rusek 2013). In spite of its several advantages, the practical feasibility of massive-MIMO is challenging due to the fact that each transmit antenna at the base station as well as at multi-antenna users requires dedicated signal processing resources such as RF chains, power amplifiers, and other circuitry. This increases the power consumption and architectural complexity of the transmitter. Furthermore, reliable detection of spatially multiplexed information streams at the receiver is another crucial factor due to inter-channel interference (interference between multiple streams transmitted from the antennas of the same user/BS) for realizing uRLLC (Yang and Hanzo 2015).

On the other hand, NOMA (Ding 2017) is considered as another such technique which enhances the SE of wireless systems by relaxing the orthogonality of resources for multiple access (termed as orthogonal multiple access (OMA)). In contrast to OMA, NOMA shares the same resources, such as time and frequency for transmitting independent information streams to multiple users simultaneously. NOMA can be broadly classified as code domain (CD) NOMA and power domain (PD) NOMA (Dai 2018). Among these, PD-NOMA is interestingly explored by the researchers for its potential applications in B5G systems. In PD-NOMA, the information of multiple users is superimposed at the transmitter with different power gains which depend on the channel condition of each user. However, at the receiver end, users employ successive interference cancellation (SIC) to decode their information which is a complex process and is severely affected by error propagation (Dai 2018; Ding 2017).

Recently, IM techniques (Basar 2016; Wen et al. 2017) are getting increased attention from both academia and industry as a promising solution for spectrally efficient, low-power, and low-complexity architecture designs. IM relaxes the need for dedicated signal processing resources in wireless systems by encoding the extra information across the resources such as antenna, sub-carriers in OFDM, time-slots, and parasitic elements (radio frequency (RF) mirrors) available at the transmitter. Based on the indexing of different resources, IM techniques can be classified as:

- Spatial modulation (SM);
- Generalized spatial modulation (GSM);
- Space shift keying (SSK);
- OFDM sub-carrier index modulation (OFDM-IM);
- Media-based modulation (MBM).

In IM, extra information bits are encoded across the indices of the available resources at the transmitter along with transmitting information bits using the conventional modulation, i.e., through a symbol selected from the modulation alphabet (Basar 2016; Wen et al. 2017). Basically, there are two types of mappings involved in IM, namely index mapping and symbol mapping. Through index mapping, a part of incoming bits is allocated for activating the resources which yield activation pattern whereas the other part of the incoming bits is used for selecting a constellation point from the modulation alphabet as depicted in Fig. 4.1.

The concept of IM is first introduced in the name of *spatial modulation* in (Di Renzo et al. 2011; Mesleh 2008). In SM, extra information bits are encoded across the index of one active antenna element which is used for transmission of information by selecting a symbol from the conventional modulation alphabet. The increase in SE due to index mapping in SM depends on the total number of combinations of one active antenna. It turns out that, in SM the increase in SE is logarithmic with respect to the number of transmit antennas (discussed in detail later in the chapter). Although SM has limited SE due to only one active antenna, some of its key advantages include the complete elimination of inter-channel interference, requirement of only one radio frequency (RF) chain, and no constraints on the number of the receive antennas. The generalization of SM termed as GSM is proposed in Younis et al.

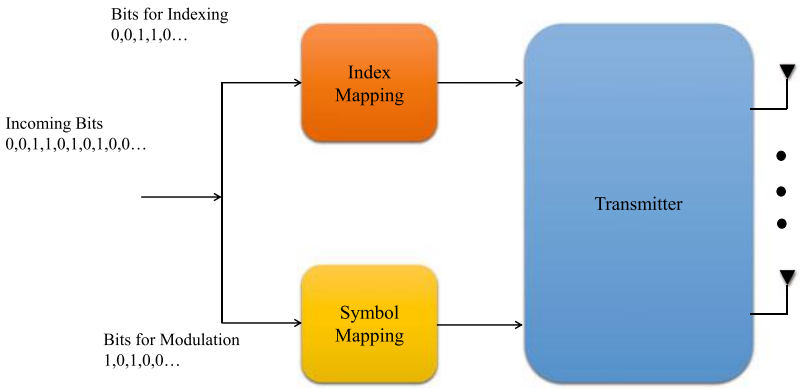



**Fig. 4.1** Index and symbol mapping

(2010) where more than one antenna elements are active which helps in embedding more information bits. It is worth noting that the total number of combinations of active antennas increases in GSM which results in higher SE. Though GSM has higher SE over SM, it suffers from the inter-channel interference which is due to multiple active antennas. Reliable detection of information in SM and GSM is a challenging task due to embedded information in index as well as the transmitted symbols which involve significant complexity in detection. To avoid the complexity of detectors in SM and GSM, a simple modulation scheme called as SSK is proposed in Jeganathan et al. (2008, 2009). In SSK, only one antenna is active, which transmits the symbol 1 rather than transmitting a symbol from the modulation alphabet in contrast to SM and GSM. Therefore, SSK relaxes the need for detection of the transmitted symbol from the active antenna yielding a low-complexity detector design. OFDM-IM (Basar 2013, 2015; Ozturk 2016) is another technique where indexing is performed across the sub-carriers in an OFDM transmission system. In OFDM-IM, the combinations of active sub-carriers are large due to which not all the possible combinations are used for transmitting the information. On the other hand, MBM (Khandani 2013, 2014) is a recently proposed IM scheme where multiple RF mirrors are placed near the transmit antenna which creates different channel fade realizations for transmission of information. The indexing in MBM is performed across these channel fade realizations by activating the RF mirrors (Naresh and Chockalingam 2017). The key advantage of MBM is that it offers enhanced spectral efficiency over the other IM techniques such as SM, GSM, and SSK.

IM with massive-MIMO serves as a potential solution to provide better SE and EE in futuristic wireless systems. IM with massive-MIMO has the capability of enabling low-power and ultra-reliable communication which are key objectives of B5G wireless systems. Moreover, further research is being carried out over generalizing the IM concept by combining one or more IM techniques, which are termed as multi-dimensional IM (MIM) (Shamasundar 2017). In MIM, multiple IM concepts are integrated to transmit more information bits through multiple indexing as compared

with the conventional IM. Therefore, based on the different possible combinations of IM techniques, MIM can be classified as:

- Space-time IM (ST-IM);
- Spatial modulated media-based modulation (SM-MBM);
- Generalized spatial modulated media-based modulation (GSM-MBM).

In all the aforementioned MIM schemes, indexing is performed across multiple resources simultaneously to encode additional information bits. In ST-IM, antenna and time-slots are indexed together which increase sparsity in the transmitted symbol vector. However, it lacks significantly in the spectral efficiency due to the use of multiple time-slots and activating only one of them. SM-MBM and GSM-MBM are the two most promising MIM techniques where antenna indexing and RF mirrors activity pattern are used for encoding the information. It provides a significant increase in both the SE and the sparsity in the transmitted symbol vector. However, in such MIM techniques, the complexity of the detector increases for achieving the target bit error rate performance. Although IM and MIM have several advantages over the conventional transmission scheme, symbol detection is one of the practical challenges in such systems. IM and MIM detection algorithms involve two key steps: i) detecting the indices used for transmission and ii) detecting the transmitted symbols (Gao et al. 2017; Narasimhan et al. 2015). Finally, the detected indices and symbols are combined to obtain the transmitted information bits through demapping as depicted in Fig. 4.2. To this end, there are several detection techniques proposed in the literature to detect the symbols reliably. In this chapter, we discuss ST-IM and SM-MBM schemes with their mathematical formulation and ML detection in detail.

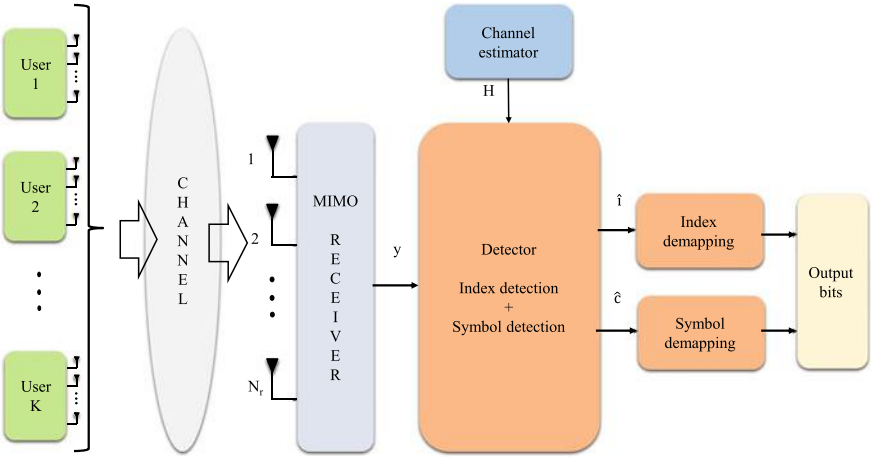



**Fig. 4.2** Detector design and demapping

*Notations* Boldface uppercase and lowercase alphabets represent column vectors and matrices, respectively. $(.)^T$ and $()^{-1}$, respectively, denote transpose and inverse operations. $|\ .\ |$ is the cardinality of a set. $||\ .\ ||_2$ denotes L2-norm. The absolute value of an element is represented as $|\ .\ |$.

## 4.2 System Model

In this section, we briefly discuss the mathematical formulation of various terminologies in massive-MIMO system which is very much useful in developing the mathematical model for IM massive-MIMO systems in later sections of the chapter. We also discuss the optimal detection scheme in massive-MIMO systems which is termed as maximum-likelihood (ML) detection for detecting the transmitted information from the received symbol vector. Let us consider a massive-MIMO system with $K$ users with each user having $N_{\mathrm{t}}$ transmit antennas, and a base station (BS) having $N_{\mathrm{r}}$ receive antennas to serve the users ($N_{\mathrm{r}} \times KN_t$) (Lu et al. 2014). We specifically consider the uplink scenario to introduce different IM and MIM schemes in the chapter. User $i$ is assumed to transmit symbol vector $\mathbf{x}_i$ to the BS in each channel use. Each element $x_{(i,k)}$ of $\mathbf{x}_i$ for $k = 1, 2, \ldots, N_t$ is assumed to be taken from a modulation alphabet say $\mathcal{A}$ ( e.g., $\mathcal{A} = \{-1 - 1i, -1 + 1i, 1 - 1i, 1 + 1i\}$ for 4-QAM). The wireless channel between the $i$th user and the BS is represented by $\mathbf{H}_i$ with dimension $N_r \times N_t$ as shown in Fig. 4.3. Note that the dimensions of the channel matrix may change during the discussion of different IM schemes due to different considerations in each scheme. With these assumptions and considerations, the received vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{H_1 x_1} + \mathbf{H_2 x_2} + \cdots + \mathbf{H_K x_K} + \mathbf{n}, \tag{4.1}$$

where $\mathbf{n}$ is the additive white Gaussian noise (AWGN) at the receiver, with each element having the Gaussian distribution with mean zero and variance $\sigma^2$, i.e., $\sim \mathcal{CN}(0, \sigma^2)$. Eq. (4.1) can also be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.2}$$

where $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_K]$ and $\mathbf{x} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_K^T\right]^T$ are the composite channel matrix and the transmitted symbol vector, respectively.

### *4.2.1 Maximum-Likelihood Detection*

The optimal detection of information from the received symbol vector can be obtained by performing an exhaustive search over all the possible combinations of the transmit symbol vectors (Rusek 2013). Since each element $x_{(i,k)}$ for $i = 1, 2, \ldots, K$ and $k =$

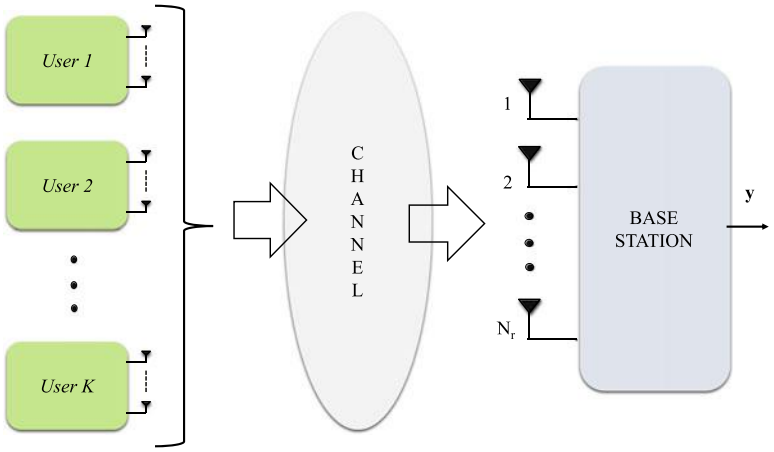



**Fig. 4.3** Massive-MIMO system

$1, 2 \ldots, N_t$ is taken from a modulation alphabet of size $|\mathcal{A}|$, and therefore, the set of all the possible combinations of composite transmitted symbol vector $\mathbf{x}$ consists of $|\mathcal{A}|^{KN_t}$ symbol vectors. Thus, to obtain the best possible symbol vector, ML detector performs search over a set consisting of possible transmit $|\mathcal{A}|^{KN_t}$ vectors using the maximum-likelihood (ML) cost metric. The ML cost metric is given by

$$C_{\text{ML}} = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \tag{4.3}$$

and the ML solution is given by

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{x} \in \mathcal{A}^{KN_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \tag{4.4}$$

This can further be simplified as

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{x}_i \in \mathcal{A}^{N_t}} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{H}_i \mathbf{x}_i\|_2^2, \tag{4.5}$$

$$= \arg \min_{x_{(i,j)} \in \mathcal{A}} \|\mathbf{y} - \sum_{i=1}^{K} \sum_{j=1}^{N_t} \mathbf{h}_{(i,j)} x_{(i,j)}\|_2^2 \tag{4.6}$$

However, it can be noted that the number of possible combinations is exponential in the number of users and the number of antennas at each user which makes the realization of ML detection computationally prohibitive.

## 4.3 Types of Index Modulation

In this section, we discuss the key concept behind SM, GSM, SSK, and MBM with the required mathematical formulation and block diagrams. We also provide the ML detection rule for each of the aforementioned IM schemes for optimal detection of the transmitted information. First, we discuss the SM scheme followed by the discussion on its generalized version, i.e., GSM scheme. Next, we discuss the SSK scheme which is a low-complexity modification of SM. Finally, we provide a brief introduction to the recently proposed MBM scheme.

### *4.3.1 Spatial Modulation*

SM (Mesleh 2008), where indexing is performed across the antennas (i.e., spatial dimension), is the first IM scheme proposed in the literature. In SM, only one antenna out of the available antennas is active at a time which generates different single active antenna patterns termed as *antenna activation patterns* (Di Renzo et al. 2011). The information to be encoded is mapped on the antenna activity pattern by selecting one of the possible patterns based on the incoming information bits. As depicted in Fig. 4.4, the index mapper is connected to the antenna mapper which is used to activate the switch between the RF chain and the selected antenna for transmitting the processed information obtained from the symbol mapper for the *i*th user.

For simplicity, let us consider a transmitter with $N_t$ antennas. Then, the total number of antenna activation patterns by selecting one antenna out of $N_t$ available antennas equals to $\binom{N_t}{1}$. Therefore, the total number of bits which can be used for selecting one of these patterns is $\lfloor \log_2 \binom{N_t}{1} \rfloor$, where $\lfloor \cdot \rfloor$ is the flooring operation. After selecting the active antenna, a symbol is selected from the modulation alphabet by using $\log_2 |\mathcal{A}|$ bits for transmission through the active antenna. The SE of SM in terms

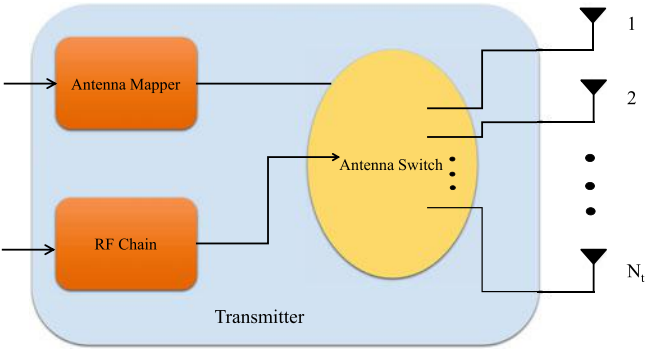



**Fig. 4.4** Transmitter block diagram for SM

of bits per channel use (bpcu), i.e., the total number of bits that can be transmitted by a single user in the $K$ user system is given by

$$\eta_{\mathrm{SM}} = \left( \underbrace{\lfloor \log_2 \binom{N_{\mathrm{t}}}{1} \rfloor}_{\text{Index mapping}} + \underbrace{\log_2 |\mathcal{A}|}_{\text{Symbol mapping}} \right) \quad \text{bpcu.} \tag{4.7}$$

These $\eta_{\mathrm{SM}}$ bits can be conveyed through the symbol vector transmitted from the user, say $i$th user, which can be written as

$$\mathbf{x}_i = \left[ 0, 0, \cdots, \underbrace{x_{(i,k)}}_{\text{active antenna}}, \ldots, 0, 0 \right],$$

where $k$th antenna is active which transmits a symbol $x_{(i,k)} \in \mathcal{A}$. Similarly, each user maps the incoming information bits into a transmit vector for conveying the information to the BS. The symbol vector received at the BS can be written as

$$\mathbf{y} = \mathbf{H}_1 \mathbf{x}_1 + \mathbf{H}_2 \mathbf{x}_2 + \cdots + \mathbf{H}_K \mathbf{x}_K + \mathbf{n}$$

which is rewritten as

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{h}_{(i,k)} \mathbf{x}_{(i,k)} + \mathbf{n}, \quad \text{for } k = 1, 2, \ldots, N_{\mathrm{t}}, \tag{4.8}$$

where $k$ differs for each $i$ depending on the incoming information. For example, let us consider a system with single user having $N_{\mathrm{t}} = 4$ and using BPSK modulation, the set of all the possible transmit vector is given by

$$\mathbb{S}_{\mathrm{SM}} = \left\{ \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} +1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{k=1}, \underbrace{\begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ +1 \\ 0 \\ 0 \end{bmatrix}}_{k=2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ +1 \\ 0 \end{bmatrix}}_{k=3}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ +1 \end{bmatrix}}_{k=4} \right\}$$

Therefore, by using $\mathbb{S}_{\mathrm{SM}}$ we can obtain $\eta_{\mathrm{SM}} = 3$ bpcu where 1 bit can be transmitted by using a symbol from BPSK and other 2 bits by using one activity pattern, i.e., $\lfloor \log_2 \binom{4}{1} \rfloor$.

#### 4.3.1.1 ML Detection in SM

The ML solution in SM massive-MIMO can be obtained similar to that of the massive-MIMO system (discussed in Sect. 4.2.1). In SM, each user transmits a vector selected from $\mathbb{S}_{\text{SM}}$, the set of all the possible SM transmit vectors. Mathematically, the ML solution at the receiver to achieve the optimal detection can be written as

$$\widehat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x} \in \mathbb{S}_{\text{SM}}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \tag{4.9}$$

This can further be simplified as

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \min_{x_{(i,k)} \in \mathcal{A}, \mathbf{h}_{(i,k)} \in \mathbf{H}_i, \forall i} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{h}_{(i,k)} \mathbf{x}_{(i,k)}\|_2^2. \tag{4.10}$$

It can easily be observed from here that a joint search is performed over $x_{(i,k)} \in \mathcal{A}$ and $\mathbf{h}_{(i,k)} \in \mathbf{H}_i$.

### *4.3.2 Generalized Spatial Modulation*

SM, which is simplest form of IM schemes, is limited in SE due to only one active antenna. Therefore, GSM, which is an spectrally efficient advancement of SM, proposes to activate multiple antennas for transmission of multiple data streams simultaneously (Younis et al. 2010). However, not all the antennas are active in GSM, which makes it different from spatial multiplexing-based MIMO transmission. The higher SE in GSM is achieved at the cost of increased architecture complexity, which is due to the requirement of dedicated RF chains for each of the active antenna as depicted in Fig. 4.5. In GSM, say $N_a$ number of antennas are active out of all the available $N_{\text{t}}$ antennas, and therefore, the total number of antenna activation patterns in GSM is $\binom{N_{\text{t}}}{N_{\text{a}}}$. The total number of bits which can be used for selecting one of these activation patterns is $\lfloor \log_2 \binom{N_{\text{t}}}{N_{\text{a}}} \rfloor$. Further, each of the $N_{\text{a}}$ active antennas transmits a symbol selected from the modulation alphabet $\mathcal{A}$, which results in transmission of $N_{\text{a}} \log_2 |\mathcal{A}|$ information bits. Hence, the total number of information bits which can be conveyed form a single GSM user is

$$\eta_{\text{GSM}} = \left( \underbrace{\lfloor \log_2 \binom{N_{\text{t}}}{N_{\text{a}}} \rfloor}_{\text{Antenna Index Mapping}} + \underbrace{N_{\text{a}} \log_2 |\mathcal{A}|}_{\text{Symbol Mapping}} \right) \text{ bpcu.} \tag{4.11}$$

which is significantly higher than that of the SM scheme. The transmit vector in GSM consists of multiple non-zero entries and can be written as

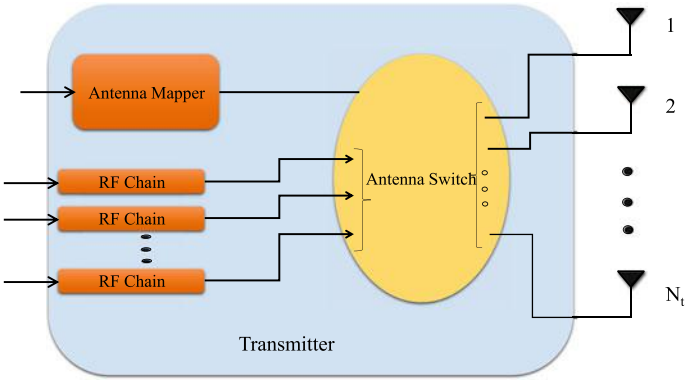



**Fig. 4.5** Transmitter block diagram of GSM

$$\mathbf{x}_i = \Big[\, 0, \cdots, \underbrace{x_{i_1}}_{\text{active antenna 1}}, \ldots 0, \cdots, \underbrace{x_{i_2}}_{\text{active antenna 2}}, \ldots, 0, \ldots, \underbrace{x_{i_{N_\text{a}}}}_{\text{active antenna } N_\text{a}}, \ldots, 0 \,\Big],$$

where each of the $x_{i_1}, x_{i_2}, \ldots, x_{i_{N_\text{a}}}$ is selected from the modulation alphabet $\mathcal{A}$ and $i_1, i_2, \ldots, i_{N_\text{a}}$ are the indices of the active antennas at the $i$th user. The received symbol vector at the BS can now be written as

$$\mathbf{y} = \sum_{i=1}^{K} \sum_{k=1}^{N_\text{a}} \mathbf{h}_{i_k} \mathbf{x}_{i_k} + \mathbf{n}. \tag{4.12}$$

For example, consider a multiple antenna system with $N_\text{t} = 4$, $N_\text{a} = 2$ and BPSK modulation, the set of all the possible transmit vectors can be given by

$$\mathbb{S}_\text{GSM} = \left\{ \underbrace{\begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} +1 \\ +1 \\ 0 \\ 0 \end{bmatrix}}_{k=1,2}, \ldots, \underbrace{\begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \end{bmatrix}}_{k=1,3}, \ldots, \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix}}_{k=1,4}, \ldots \underbrace{\begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix}}_{k=2,3}, \ldots, \underbrace{\begin{bmatrix} 0 \\ -1 \\ 0 \\ -1 \end{bmatrix}}_{k=2,4}, \ldots, \underbrace{\begin{bmatrix} 0 \\ 0 \\ +1 \\ +1 \end{bmatrix}}_{k=3,4} \right\}$$

The total number of vectors in the set $\mathbb{S}_\text{GSM}$ is $\binom{N_\text{t}}{N_\text{a}} \times |\mathcal{A}|^{N_\text{a}} = 6 \times 2^2 = 24$ which can convey $\lfloor \log_2(24) \rfloor = 4$ bits of information. It is worth noting that not all the patterns are used here for transmission of information, and therefore, only 16 out of 24 patterns are used in this example. Also, spatial multiplexing and SM are the special cases of GSM with all active antennas and one active antenna, respectively. In spatial multiplexing, $N_\text{a} = N_\text{t}$ whereas in SM $N_\text{a} = 1$.

#### 4.3.2.1 ML Detection in GSM

Optimal detection in GSM can be performed by using the ML search over the set of all the possible combinations of the transmit vector. The ML solution for GSM scheme can be obtained as

$$\widehat{\mathbf{x}}_{\mathrm{ML}} = \arg \min_{x_{i_k} \in \mathcal{A}, \mathbf{h}_{i_k} \in \mathbf{H}_i, \forall i} \|\mathbf{y} - \sum_{i=1}^{K} \sum_{k=1}^{N_{\mathrm{a}}} \mathbf{h}_{i_k} \mathbf{x}_{i_k}\|_2^2. \tag{4.13}$$

In obtaining the ML solution, a joint search is performed over $x_{i_k} \in \mathcal{A}$ and $\mathbf{h}_{i_k} \in \mathbf{H}_i$ for all $i = 1, 2, \ldots, K$ and $k = 1, 2, \ldots, N_{\mathrm{a}}$. It can easily be observed that obtaining the ML solution in GSM is computationally expensive as compared to SM scheme.

### *4.3.3 Space Shift Keying and Generalized Space Shift Keying*

SSK and GSSK are simplified forms of SM and GSM, respectively. In SSK, only one antenna is active which does not transmit the symbol from a conventional modulation alphabet (Jeganathan et al. 2009). The active antenna in SSK transmits only one type of symbol, i.e., 1 whereas all the other antennas remain off. If the number of antennas is $N_{\mathrm{t}}$, then the total number of possible active antenna combinations are $\binom{N_{\mathrm{t}}}{1}$ which can convey $\lfloor \log_2 \binom{N_{\mathrm{t}}}{1} \rfloor$ bits at a time. On the other hand, in GSSK multiple antennas are active similar to GSM but these antennas only transmit 1 (Jeganathan et al. 2008). In GSSK, if $N_{\mathrm{a}}$ antennas are active, then the total number of active antenna combinations is $\binom{N_{\mathrm{t}}}{N_{\mathrm{a}}}$ and the total bits which can be conveyed by selection of an activity pattern is $\lfloor \log_2 \binom{N_{\mathrm{t}}}{N_{\mathrm{a}}} \rfloor$ bits at a time. The transmit symbol vector in SSK can be written as

$$\mathbf{x}_i = \left[ 0, 0, \cdots, \underbrace{1}_{\text{active antenna}}, \ldots, 0, 0 \right],$$

and the received symbol vector in SSK is given by

$$\mathbf{y}_{\mathrm{SSK}} = \sum_{i=1}^{K} \mathbf{h}_{i_k} + \mathbf{n}, \quad \text{for } k = 1, 2, \ldots, N_{\mathrm{t}}. \tag{4.14}$$

Similarly, the transmit vector in GSSK can be written as

$$\mathbf{x}_i = \left[ 0, \cdots, \underbrace{1}_{\text{active antenna 1}}, \ldots 0, \cdots, \underbrace{1}_{\text{active antenna 2}}, \ldots, 0, \cdots, \underbrace{1}_{\text{active antenna } N_{\mathrm{a}}}, \ldots, 0 \right],$$

and the received symbol vector is

$$\mathbf{y}_{\text{GSSK}} = \sum_{i=1}^{K} \sum_{k=1}^{N_{\text{a}}} \mathbf{h}_{i_k} + \mathbf{n}. \tag{4.15}$$

For example, if $N_{\text{t}} = 4$, then the set of all the possible transmit vector in SSK is

$$\mathbb{S}_{\text{SSK}} = \left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{k=1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{k=2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{k=3}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{k=4} \right\}$$

and the set of all the possible transmit vectors can be written as

$$\mathbb{S}_{\text{GSSK}} = \left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{k=1,2}, \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{k=1,3}, \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{k=1,4}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}}_{k=2,3}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{k=2,4}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}_{k=3,4} \right\}$$

#### 4.3.3.1 ML Detection in SSK and GSSK

It can be observed that the problem of detection in SSK and GSSK is simplified in the sense that there is no need to perform a search over the set $\mathcal{A}$. Therefore, the ML detection problem in SSK can be formulated as

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{h}_{i_k} \in \mathbf{H}_i, \forall i} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{h}_{i_k}\|_2^2. \tag{4.16}$$

and that of in GSSK, ML solution can be obtained as

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{h}_{i_k} \in \mathbf{H}_i, \forall i} \|\mathbf{y} - \sum_{i=1}^{K} \sum_{k=1}^{N_{\text{a}}} \mathbf{h}_{i_k}\|_2^2. \tag{4.17}$$

The ML search now reduces to search only over the set $\mathbf{H}$ in contrast to the SM and GSM where the search also involves set $\mathcal{A}$, which ultimately increases the computational complexity of detector at the receiver.

### 4.3.4 Media-Based Modulation

MBM is a recently proposed IM scheme where indexing is performed across the parasitic elements called as RF mirrors (Khandani 2013). These elements are placed near the transmit antennas. The ON/OFF switching of these RF mirrors generates different channel fade realizations at a far field in a rich scattering environment. The ON/OFF activity of the mirrors is chosen based on the incoming information bits. Once a switching pattern is selected for a particular antenna, the antenna then transmits a symbol selected from the modulation alphabet as depicted in Fig. 4.6 (Khandani 2014). Note that the key difference between other IM schemes discussed so far and MBM is that, in MBM all the antennas transmit information. If there are $m_{\text{rf}}$ RF mirrors are placed near an antenna, then it generates $2^{m_{\text{rf}}}$ ON/OFF switching patterns (Naresh and Chockalingam 2017). Therefore, to select a particular pattern requires $\lfloor \log_2 2^{m_{\text{rf}}} \rfloor = m_{\text{rf}}$ information bits which is linear with respect to the number of RF mirrors. This is in contrast to the logarithmic increase in information conveying capability of SM and GSM with respect to the total number of antennas. Other than $m_{\text{rf}}$ bits, each antenna also transmits a symbol which convey additional $\log_2 |\mathcal{A}|$ information bits. Therefore, the total SE per antenna in terms of bpcu in MBM scheme is

$$\eta_{\text{MBM}} = \left( \underbrace{m_{\text{rf}}}_{\text{Index Mapping}} + \underbrace{\log_2 |\mathcal{A}|}_{\text{Symbol Mapping}} \right) \quad \text{bpcu.} \tag{4.18}$$

Let us consider a single antenna user with $m_{\text{rf}} = 2$ as shown in Fig. 4.7. The set of possible ON/OFF activity patterns is $n_{\text{a}} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ where each of the activity pattern results in an independent channel realization $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ and $\mathbf{h}_4$, respectively. Based on the incoming information bits, one of these activity pattern is selected to transmit information bearing symbol selected from the modulation alphabet. Let us consider the composite channel matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4]$, then the composite transmit vector for a single antenna MBM transmitter is $\mathbb{S}_{\text{MBM}} = \{[\alpha_i, 0, 0, 0], [0, \alpha_i, 0, 0], [0, 0, \alpha_i, 0], [0, 0, 0, \alpha_i]\}$. It is worth noting that the single antenna user transmits $\alpha_i \in \mathcal{A}$ only from the antenna, but for mathematical simplicity we consider the concept of composite channel and composite transmit vector.

For generalization of the transmit vector to the case of multiple users, let us consider single antenna users for simplicity. Each user consists of $m_{\text{rf}}$ RF mirrors, $M = 2^{m_{\text{rf}}}$ mirror activation patterns and use $\mathcal{A}$ modulation alphabet for transmission of a symbol. Let $\mathbf{h}_k^j$ denote the $N_r \times 1$ channel state vector corresponding to the $j$th MAP selected by the $k$th user for transmitting the $\alpha_k \in \mathbb{A}$, where $\mathbf{h}_k^j = [h_{1,k}^j, h_{2,k}^j, \ldots, h_{N_r,k}^j]^T$. Each $h_{i,k}^j$ is assumed to be independent and identically distribute (i.i.d.) complex Gaussian random variable with zero mean and unit variance, i.e., $\sim \mathcal{CN}(0, 1)$. Let us define the matrix $\mathbf{H}_k = [\mathbf{h}_k^1, \mathbf{h}_k^2, \ldots, \mathbf{h}_k^M]$ and the MBM signal set for a single user as

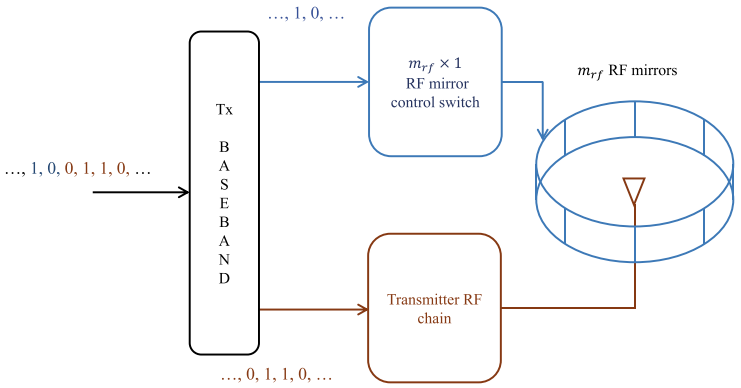



**Fig. 4.6** Transmitter block diagram of MBM

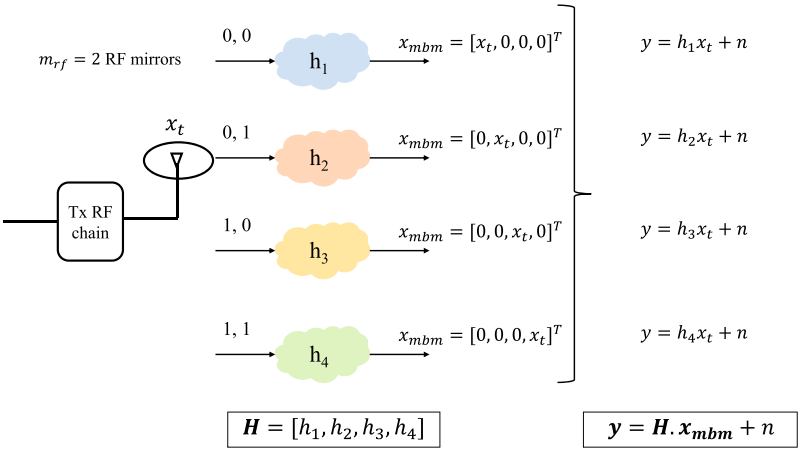



**Fig. 4.7** Single antenna MBM system

$$\begin{aligned} &\mathbb{S}_{\text{MBM}} = \{\mathbf{s}_{j,i} : j = 1, 2, \ldots, M, i = 1, 2, \ldots, |\mathbb{A}|\}, \\ &\text{s.t. } \mathbf{s}_{j,i} = [0, \ldots, 0, \underbrace{\alpha_i}_{j\text{th coordinate}}, 0, \ldots, 0]^T, \alpha_i \in \mathbb{A}, \end{aligned} \tag{4.19}$$

which means that $\mathbf{s}_{j,i}$ is an $M \times 1$ vector with only one nonzero entry $\alpha_i \in \mathbb{A}$ corresponding to the $j$th channel fade realization. For the $k$th user, let us denote $\mathbf{x}_k \in \mathbb{S}_{\text{MBM}}$, and therefore, the received vector at the BS can be written as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{n}, \tag{4.20}$$

where $\mathbf{n}$ is an $N_r \times 1$ additive white Gaussian noise (AWGN) vector with $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_r})$. Further, we can rewrite the received vector as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.21}$$

where $\mathbf{H} = [\mathbf{H}_1 \quad \mathbf{H}_2 \quad \dots \quad \mathbf{H}_K]$ is an $N_r \times KM$ multi-user MBM-MIMO channel matrix and $\mathbf{x} = [\mathbf{x}_1^T \quad \mathbf{x}_2^T \quad \dots \quad \mathbf{x}_K^T]^T$ is $KM \times 1$ multi-user MBM-MIMO transmit symbol vector.

At receiver, the objective is to detect the ML solution given the received vector $\mathbf{y}$ and the channel state information $\mathbf{H}$ as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{S}_{\text{MBM}}^K} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \tag{4.22}$$

However, finding ML solution requires an exhaustive search over the set of all the possible vectors, i.e., $\mathbb{S}_{\text{MBM}}^K$ which is computationally impractical in massive-MIMO systems.

## 4.4 Multi-dimensional Index Modulation

In this section, we discuss an improved IM scheme termed as MIM wherein multiple IM schemes are integrated to obtain higher SE. First, we discuss the space-time IM scheme which is the simplest form of MIM. Next, we extend this concept to the spatial MBM which achieves significantly high SE over other IM and MIM schemes.

### *4.4.1 Space-Time IM*

In space-time IM, multiple resources are utilized for indexing such as time-slots and antennas at the transmitter apart from transmitting the symbol selected from the modulation alphabet (Shamasundar 2017). Let us consider $T$ time-slots and $N_\text{t}$ antennas are available for indexing. The total number time-slot activity patterns are $\binom{T}{1}$, and for each time-slot activity pattern, there are $\binom{N_\text{t}}{1}$ antenna activity patterns possible. Therefore, the total number of bits that can be conveyed by indexing across time-slots is $\lfloor \log_2 \binom{T}{1} \rfloor$ and that of across antenna is $\lfloor \log_2 \binom{N_\text{t}}{1} \rfloor$. The incoming information bits are divided into three part: One part is used for selecting the time-slot, second part is used for selecting an antenna, and the third part is used to select a symbol from $\mathcal{A}$ for transmission. The SE in space-time IM in terms of bpcu is

$$\eta_{\text{ST-IM}} = \lfloor \log_2 \binom{T}{1} \rfloor + \lfloor \log_2 \binom{N_\text{t}}{1} \rfloor + \log_2 |\mathcal{A}|. \tag{4.23}$$

For example, if $T = 2$, $N_t = 4$ and BPSK modulation is considered, then the set of possible transmit vector is

$$\mathbb{S}_{\text{ST-IM}} = \left\{ \left\{ \underbrace{\begin{bmatrix} -1 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} +1 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} 0 \\ +1 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=2} \right\} \right\}$$
$$\left\{ \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} -1 \\ 0 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} +1 \\ 0 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{T=2} \right\}, \left\{ \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{T=1}, \underbrace{\begin{bmatrix} 0 \\ +1 \end{bmatrix}}_{T=2} \right\} \right\} \tag{4.24}$$

The transmit symbol vector of $i$th user for ST-IM can be written as

$$\mathbf{x}_i^{(t)} = [0, \ldots, 0, x_i(k, \tau), 0, \ldots, 0], \tag{4.25}$$

where $k$ is the active antenna index, $\tau \in \{1, 2, \ldots, T\}$, $x_i(k, \tau) \in \mathcal{A}$, i.e., $x_i(k, t) = 0$ for all $t \neq \tau$. With the assumption that the channel state information is block faded for $T$ time-slots, then the received vector in ST-IM can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{H}_i \mathbf{x}_i^{(t)} + \mathbf{n} \\ &= \sum_{i=1}^{K} \mathbf{h}_{i,k} x_i(k, \tau) + \mathbf{n} \end{aligned} \tag{4.26}$$

Now, let us consider the composite transmit vector $\mathbf{x} = [\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \ldots, \mathbf{x}_K^{(t)}]$. The ML detector in ST-IM for obtaining the optimal solution can be formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}_i^{(t)} \in \mathbb{S}_{\text{ST-IM}}^K} \|\mathbf{y} - \mathbf{H}_i \mathbf{x}_i^{(t)}\|_2^2, \tag{4.27}$$

$$= \arg \min_{x_i(k,\tau) \in \mathcal{A}} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{h}_{i,k} x_i(k, \tau)\|_2^2. \tag{4.28}$$

### *4.4.2 Spatial Modulated Media-Based Modulation*

SM-MBM is a new MIM scheme where the concept of SM and MBM is integrated to obtain a higher SE over both SM and MBM (Shamasundar 2017). In SM-MBM, we consider multiple antenna transmitter with each antenna having multiple RF mirrors deployed nearby each antenna as depicted in Fig. 4.8.

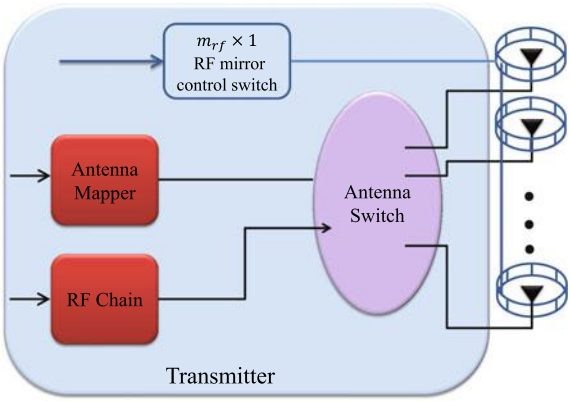



**Fig. 4.8** SM-MBM transmitter block diagram

In SM-MBM, the information is conveyed by using antenna activation pattern, mirror activation pattern, and the symbol transmitted selected from the modulation alphabet. The total number of information bits that can be conveyed by using antenna activation pattern and mirror activation pattern is $\lfloor \log_2 \binom{N_\mathrm{t}}{1} \rfloor + m_\mathrm{rf}$. Additionally, $\log_2 |\mathcal{A}|$ bits are also conveyed using a symbol selected from the modulation alphabet $\mathcal{A}$. This results in the total SE of SM-MBM scheme as

$$\eta_{\text{SM-MBM}} = \left( \underbrace{\lfloor \log_2 \binom{N_\mathrm{t}}{1} \rfloor}_{\text{Antenna index mapping}} + \underbrace{m_\mathrm{rf}}_{\text{RF mirror index mapping}} + \underbrace{\log_2 |\mathcal{A}|}_{\text{Symbol mapping}} \right) \text{ bpcu.} \tag{4.29}$$

The transmit symbol vector $\mathbf{x}_i$ from the $i$th user can be written as $\left[0, \cdots, 0, x_{(i,k,p)}, 0, \ldots, 0\right]$, where $x_{i,k,p} \in \mathcal{A}$ is the information symbol from the $k$th antenna and $p$th mirror activation pattern. It is worth noting that $k \in \{1, 2, \ldots, N_\mathrm{t}\}$ and $p \in \{1, 2, \ldots, 2^{m_\mathrm{rf}}\}$, and therefore, the size of the symbol vector $\mathbf{x}_i$ is $2^{m_\mathrm{rf}} \times N_\mathrm{t}$ with only one nonzero entry. Similarly, the channel matrix $\mathbf{H}_i$ consists of $2^{m_\mathrm{rf}} \times N_\mathrm{t}$ columns with each column of size $N_r \times 1$. The received vector in SM-MBM system can be written as

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^{K} \mathbf{H}_i \mathbf{x}_i + \mathbf{n}, \\ &= \sum_{i=1}^{K} \mathbf{h}_{(i,k,p)} x_{(i,k,p)} + \mathbf{n}. \end{aligned} \tag{4.30}$$

**Table 4.1** IM and MIM schemes with their SE

| IM and MIM schemes | Indexing | SE (bpcu) |
|---|---|---|
| SM | Antenna | $\eta_{\text{SM}} = \left( \underbrace{\lfloor \log_2 \binom{N_\text{t}}{1} \rfloor}_{\text{Antenna index mapping}} + \underbrace{\log_2 \lvert\mathcal{A}\rvert}_{\text{Symbol mapping}} \right)$ |
| GSM | Antenna | $\eta_{\text{GSM}} = \left( \underbrace{\lfloor \log_2 \binom{N_\text{t}}{N_\text{a}} \rfloor}_{\text{Antenna index mapping}} + \underbrace{N_\text{a} \log_2 \lvert\mathcal{A}\rvert}_{\text{Symbol mapping}} \right)$ |
| SSK | Antenna | $\eta_{\text{SSK}} = \left( \underbrace{\lfloor \log_2 \binom{N_\text{t}}{1} \rfloor}_{\text{Antenna index mapping}} \right)$ |
| GSSK | Antenna | $\eta_{\text{SSK}} = \left( \underbrace{\lfloor \log_2 \binom{N_\text{t}}{N_\text{a}} \rfloor}_{\text{Antenna index mapping}} \right)$ |
| MBM | RF mirrors | $\eta_{\text{MBM}} = \left( \underbrace{m_\text{rf}}_{\text{RF mirror index mapping}} + \underbrace{\log_2 \lvert\mathcal{A}\rvert}_{\text{Symbol mapping}} \right)$ bpcu |
| SM-MBM | Antenna and RF mirrors | $\eta_{\text{SM-MBM}} = \left( \underbrace{\lfloor \log_2 \binom{N_\text{t}}{1} \rfloor}_{\text{Antenna index mapping}} + \underbrace{m_\text{rf}}_{\text{RF mirror index mapping}} + \underbrace{\log_2 \lvert\mathcal{A}\rvert}_{\text{Symbol mapping}} \right)$ bpcu |

Next, to detect the transmitted information symbol optimally, we introduce the ML detection for SM-MBM which can be written as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}_i \in \mathbb{S}^K_{\text{SM-MBM}}} \|\mathbf{y} - \mathbf{H}_i \mathbf{x}_i\|_2^2, \tag{4.31}$$

$$= \arg \min_{x_{(i,k,p)} \in \mathcal{A}} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{h}_{(i,k,p)} x_{(i,k,p)}\|_2^2, \tag{4.32}$$

where $\mathbb{S}^K_{\text{SM-MBM}}$ is the set of all the possible composite transmit vectors from the $K$ users jointly.

This completes our discussion on different IM and MIM schemes in detail. All the schemes are listed in Table 4.1 with their SE in terms of bpcu.

## 4.5 Conclusion

In this chapter, we introduced different IM schemes as a low-complexity solution for B5G wireless systems. In particular, first, we discussed SM, GSM, SSK, and MBM schemes with their mathematical formulation in detail. We also discussed the ML solution for each of these IM techniques. Through discussion, it is revealed that SSK is the simplest IM scheme and require low-complexity receiver design for detection of information symbols. SM and GSM involve the transmission of additional information through a symbol selected from the modulation alphabet which increases its SE at the cost of increased detection complexity and requirement of dedicated RF chain(s). On the other hand, MBM, which is a new IM scheme, has improved SE over SSK, SM, and GSM due to linear increase in SE with respect to the number in RF mirrors. Next, we discussed multi-dimensional IM schemes which involve combination of two or more IM schemes to convey the information from transmitter to the receiver thereby enhancing the SE further. IM and MIM are the emerging scheme with significant advantages in terms of inter-channel interference, reliability, and saving in resources. However, there is a lot of research scope to develop low-complexity receiver architectures for detection and demapping of the transmitted information in IM and MIM. As a future work, application of IM and MIM in the downlink of NOMA systems is an interesting work which could replace the complex successive interference cancellation operation at the user end without affecting the SE.

**Acknowledgements** The authors would like to thank Start-up research grant, Science and Engineering Research Board (SERB) (File no. SRG/2019/000654) of Department of Science and Technology (DST), Government of India for supporting the research work. Authors would also like to thank Visvesvaraya PhD scheme of Ministry of Electronics and Information Technology, Government of India for their support.

## References

Andrews JG et al (2014) What will 5G be? IEEE J Sel Areas Commun 32:1065–1082
Basar E et al (2013) Orthogonal frequency division multiplexing with index modulation. IEEE Trans Sig Process 61(22):5536–5549
Basar E (2015) Multiple-input multiple-output OFDM with index modulation. IEEE Sig Process Lett 22(12):2259–2263
Basar E (2016) Index modulation techniques for 5G wireless networks. IEEE Commun Mag 54(7):168–175
Boccardi F et al (2014) Five disruptive technology directions for 5G. IEEE Commun Mag 52(2):74–80
Dai L et al (2018) A survey of non-orthogonal multiple access for 5G. IEEE Commun Surv Tutor 20(3):2294–2323
Di Renzo M, Haas H, Grant PM (2011) Spatial modulation for multiple antenna wireless systems: a survey. IEEE Commun Mag 49(12):182–191
Ding Z et al (2017) A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends. IEEE J Sel Areas Commun 35(10):2181–2195
Gao Z, Dai L, Qi C, Yuen C, Wang Z (2017) Near-optimal signal detector based on structured compressive sensing for massive SM-MIMO. IEEE Trans Veh Technol 66(2):1860–1865
Jeganathan J, Ghrayeb A, Szczecinski L, Ceron A (2009) Space shift keying modulation for MIMO channels. IEEE Trans Wirel Commun 8(7):3692–3703
Jeganathan J, Ghrayeb A, Szczecinski L (2008) Generalized space shift keying modulation for MIMO channels. In: Proceedings of the IEEE 19th international symposium on personal, indoor and mobile radio communications, PIMRC, Cannes, France, pp 1–5
Khandani AK (2013) Media-based modulation: a new approach to wireless transmission. In: Proceedings of IEEE ISIT', pp 3050–3054
Khandani AK (2014) Media-based modulation: converting static Rayleigh fading to AWGN. In: Proceedings of IEEE ISIT' 2014, pp 1549–1553
Lu L, Li GY, Swindlehurst AL, Ashikhmin A, Zhang R (2014) An overview of massive MIMO: benefits and challenges. IEEE J Sel Top Sig Process 8(5):742–758
Mesleh RY et al (2008) Spatial modulation. IEEE Trans Veh Technol 57(4):2228–2241
Narasimhan TL, Raviteja P, Chockalingam A (2015) Generalized spatial modulation in large-scale multiuser MIMO systems. IEEE Trans Wirel Commun 14(7):3764–3779
Naresh Y, Chockalingam A (2017) On media-based modulation using RF mirrors. IEEE Trans Veh Technol 66(6):4967–4983
Ozturk E et al (2016) Generalized frequency division multiplexing with index modulation. In: Proceedings of IEEE Globecom workshops (GC Wkshps), Washington, DC, USA, pp 1–6
Rusek F et al (2013) Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process Mag 30(1):40–60
Shamasundar B et al (2017) Multidimensional index modulation in wireless communications. IEEE Access 6:589–604
Wen M, Cheng X, Yang L (2017) Index modulation for 5G wireless communications (wireless networks series). Springer, Berlin, Germany
Yang S, Hanzo L (2015) Fifty years of MIMO detection: the road to large-scale MIMOs. IEEE Commun Surveys Tuts 17(4):1941–1988
Younis A, Seramovski N, Mesleh R, Haas H (2010) Generalised spatial modulation. In: Proceedings conference record of the forty fourth Asilomar conference on signals, systems and computers, Pacific Grove, CA, USA, pp 1498–1502

# Chapter 5
# Sparse Code and Hybrid Multiple Access Techniques


**Sanjeev Sharma and Kuntal Deka**

**Abstract** Next-generation wireless networks require higher spectral efficiency and lower latency to meet the demands of various upcoming applications. Recently, non-orthogonal multiple access (NOMA) schemes are introduced in the literature for 5G and beyond. Various forms of NOMA are considered like power domain, code domain, pattern division multiple access, etc. to enhance the spectral efficiency of wireless networks. In this chapter, we introduce the code domain-based sparse code multiple access (SCMA) NOMA scheme to enhance the spectral efficiency of a wireless network. The design and detection of an SCMA system are analyzed in this chapter. Also, the method for codebooks design and its impact on system performance are highlighted. A hybrid multiple access scheme is also introduced using both code-domain and power-domain NOMA. Furthermore, simulation results are included to show the impact of various SCMA system parameters.

**Keywords** NOMA · SCMA · HMA · Message passing algorithm

## 5.1 Introduction of NOMA

In this chapter, we focus on the sparse code multiple access (SCMA) and hybrid multiple access (HMA) schemes for next-generation wireless systems. From 1G to 4G wireless systems are based on the orthogonal multiple access (OMA) techniques such as time division multiple access (TDMA) and frequency division multiple access (FDMA), where resources are allocated to each user exclusively. OMA-based system has lower spectral efficiency when some bandwidth resources are allocated to users

S. Sharma (✉)
Indian Institute of Technology (BHU) Varanasi, Varanasi, Uttar Pradesh 221005, India
e-mail: sanjeev.ece@iitbhu.ac.in

K. Deka
Indian Institute of Technology Goa, Goa College of Engineering Campus, Farmagudi, Ponda, Goa 403401, India
e-mail: kuntal@iitgoa.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_5

with poor channel state information (CSI). Further, OMA techniques may not be suitable for a wireless network which requires very high spectral efficiency, very low latency, and massive device connectivity. Therefore, OMA-based system may not be able to support the connectivity of billions devices in wireless network for various futuristic applications. Hence, researchers have started to focus on NOMA-based systems to achieve higher spectral efficiency and lower latency for next-generation wireless networks. In NOMA, one or more resources are allocated to more than one user simultaneously to enhance the spectral efficiency of network. Further, power domain (PD) and code domain NOMA-based systems are mainly explored in the literature (Dai et al. 2017; Sharma et al. 2018; Yang et al. 2019). Recently, NOMA-based systems are also analyzed by considering the multiple-input multiple-output (MIMO) techniques (Pan et al. 2018; Yang et al. 2018; Yuan et al. 2018).

SCMA is a code domain (CD) NOMA technique, in which a user occupies more than one orthogonal resource[1] for communication. In SCMA, multidimensional codebooks are used due to which shaping gain can be enhanced. Further, SCMA can be thought of as an extended version of low density spreading (LDS) multiple access technique. Recently, SCMA has brought a significant interest from the researchers and the scientists for 5G and beyond wireless networks.

## 5.2 SCMA System Model

Suppose the number of orthogonal resource elements and users in a system are $K$ and $J$, where $J > K$. Therefore, $J$ users communicate using $K$ resources and this system is often referred to as $J \times K$ SCMA system. In SCMA, each user's information is directly mapped to a multidimensional complex vector for transmission. Let each user has a codebook $\mathcal{X}_j, j = 1, 2, \ldots, J$, which contains $M$ complex columns of dimension $K$. Therefore, each user codebook $\mathcal{X}_j$ has a size is $K \times M$ with complex elements and can be written as $\mathcal{X}_j = \left[ \mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jM} \right] \in \mathbb{C}^{K \times M}$. For example, $M = 4$, each data symbol has two bits, i.e., $\log_2(M)$ bits. Each column of codebook $\mathcal{X}_j$ corresponds to a data symbol. Hence, first, second, third, and fourth columns are selected corresponding data symbols [00], [01], [10], and [11], respectively, for transmission. The encoding operation of an SCMA system is illustrated in Fig. 5.1.

Consider the example of $6 \times 4$ SCMA system. The 6 codebooks and their superposition are depicted in Fig. 5.2. Each codebook has four columns corresponding to the four data symbols. Further, only two same rows are non-zeros in each codebook. Therefore, the codebooks are sparse.[2] The sparsity of the codebooks can help to achieve low complexity detection of users symbols. Each user selects a column from their codebooks, and all the codewords are summed, as shown in Fig. 5.2. The values of the non-zero elements and their locations in the codebooks can be selected in such

[1] Orthogonal time slots or frequency bands or codes are referred to as orthogonal resource in the system.

[2] Only a few elements in a signal are non-zeros as compared to the total number of elements.

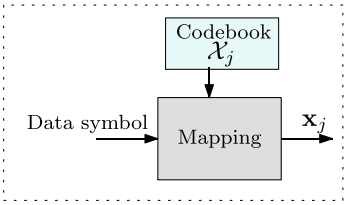



**Fig. 5.1** Data symbol to codeword mapping

a way that the system performance is optimized. Next section, we briefly summarize the downlink and the uplink SCMA system model.

### 5.2.1 Downlink SCMA System

In downlink SCMA, all users' information is broadcast from the base station to and each user node receives the sum of all users' codewords. The received signal $\mathbf{y}_i = \left[y_{i1}, \ldots, y_{iK}\right]$ at the $i$th user can be expressed as (Sharmaet al. 2019a)

$$\mathbf{y}_i = \text{diag}\,(\mathbf{h}_i) \sum_{j=1}^{J} \sqrt{P_j}\mathbf{x}_j + \mathbf{n}_i, \tag{5.1}$$

where $\mathbf{h}_i = [h_{i1}, h_{i2}, \ldots, h_{iK}]$ is the channel impulse response vector at the $i$th user and $P_j$ is the power assigned to the $j$th user. $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jK}] \in \mathbb{C}^K$ is the $j$th SCMA codeword and $\mathbf{n}_i$ is the additive white Gaussian noise (AWGN) at the $i$th user and is Gaussian distributed, i.e., $\mathbf{n}_i \sim \mathcal{CN}(0, N_0\mathbf{I}_K)$. A downlink $J \times K$ SCMA system is shown in Fig. 5.3. Message passing algorithm (MPA) is employed by each user to decode the data symbol.

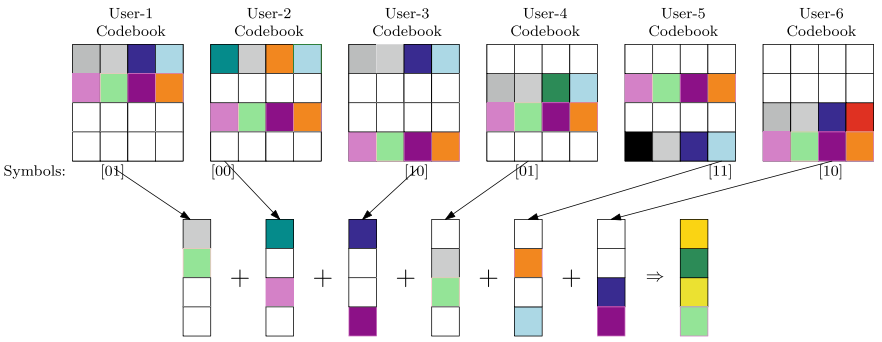



**Fig. 5.2** SCMA codebook model for $J = 6$ and $K = 4$

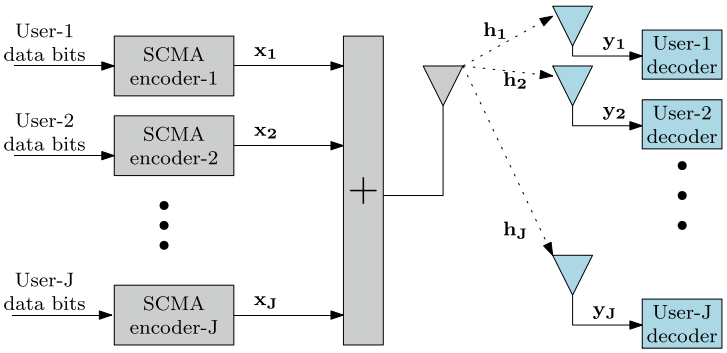



**Fig. 5.3** Downlink SCMA system

#### 5.2.1.1 Uplink SCMA System

In uplink SCMA, each user access the respective channel to transmit information to the base station, as shown in Fig. 5.4. Consider a system involving $J$ users sharing $K$ orthogonal resource elements. Each user has a codebook $\mathcal{X}_j$ which contains $M$, $K$-dimensional constellations: $\mathcal{X}_j = \left\{\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jM}\right\}$. In Rayleigh fading, the received signal $\mathbf{y}$ is expressed as

$$\mathbf{y} = \sum_{j=1}^{J} \sqrt{P_j}\,\mathrm{diag}(\mathbf{h}_j)\mathbf{x}_j + \mathbf{n}, \tag{5.2}$$

where $\mathbf{h}_j = [h_{1j}, h_{2j}, \ldots, h_{Kj}] \in \mathbb{C}^K$ is the channel impulse response vector between the receiver and $j$th user. The envelope of each $h_{jk}, k = 1, 2, \ldots, K$ is Rayleigh distributed. $P_j$ is the power of the $j$th user. In uplink SCMA system users' data symbol can be decoded jointly using the MPA at the base station.

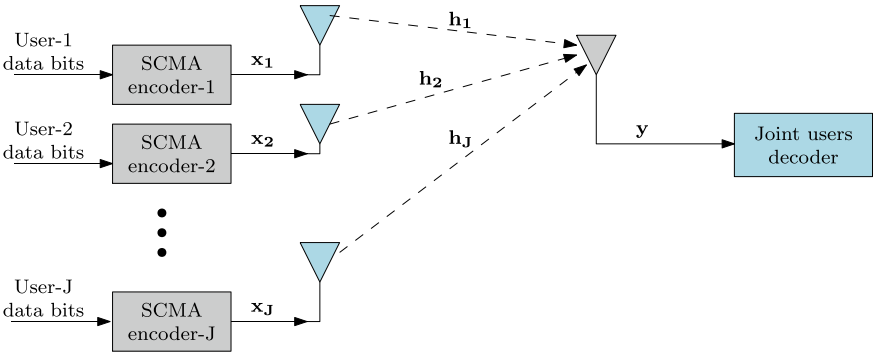



**Fig. 5.4** Uplink SCMA system

### 5.2.2 Design and Optimization of SCMA Codebooks

The performance of an SCMA system depends on the codebooks and their assignment to the users. The methods to design optimum codebooks in guaranteed fashion do not exist in literature. Usually the codebooks are designed using sub-optimal methods. For example in Zhang et al. (2016), codebooks are designed by optimizing the single dimensional complex codewords instead using the multidimensional codewords. Similarly, in Yu et al. (2018), quadrature amplitude modulation and phase rotation-based codebooks are designed using sub-optimal method.

The codebook design problem can be formulated by maximizing the sum rate of the users as (Zhang et al. 2016)

$$R_{\mathrm{s}} \leq I(\mathbf{y}; \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_J), \tag{5.3}$$

where $\mathbf{y}$ is the $K$ dimensional received signal vector and $I(\cdot)$ is the mutual information between the received signal $\mathbf{y}$ and multi-user codewords $\{\mathcal{X}_j\}_{j=1}J$. Therefore, the codebooks are optimized as

$$\{\mathcal{X}_1^{\mathrm{opt}}, \mathcal{X}_2^{\mathrm{opt}}, \ldots, \mathcal{X}_J^{\mathrm{opt}}\} = \arg\max_{\mathcal{X}_j} I(\mathbf{y}; \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_J). \tag{5.4}$$

Hence, the set of codewords $\{\mathcal{X}_1^{\mathrm{opt}}, \mathcal{X}_2^{\mathrm{opt}}, \ldots, \mathcal{X}_J^{\mathrm{opt}}\}$ represents the optimal value of codebooks for which the sum rate of users $R_{\mathrm{s}}$ has the maximum value.

***Example*** In this example, a codebook design method is illustrated for each user by considering $6 \times 4$ SCMA system. Codebooks are designed by maximizing the mutual information between the received signal $\mathbf{y}$ and interfering users' intermingled data, and the shaping gain of the constellation points (Sharma et al. 2018). Let the codewords of $d_f$ users are summed at each resource node for transmission. From the summed codeword, each user information has to be recovered. The constellation points must be designed in such a way that the sums are distinct.

Let $Y$ be the received signal over one resource element and $S$ be the sum of the codewords of $d_f$ users. The alphabet $\mathcal{S}$ for $S$ contains $M^{d_f}$ distinct sum values and is denoted by

$$\mathcal{S} = \left\{s_1, s_2, \ldots, s_{M^{d_f}}\right\}$$

where $M$ is the modulation order and $\{s_i\}_{i=1}^{M^{d_f}}$ s is the set of distinct sum values. Mutual information between $Y$ and $S$ is expressed (Sharma et al. 2018) as

$$\begin{aligned} I_m(Y; S) = \log|\mathcal{S}| - \frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\frac{1}{\pi N_0}\int_{y\in\mathbb{C}} \exp\left(-\frac{\|y - s_j\|^2}{N_0}\right) \\ \times \log\left[\sum_{i=1}^{|\mathcal{S}|}\exp\left(\frac{\|y - s_j\|^2 - \|y - s_i\|^2}{N_0}\right)\right] \mathrm{d}y. \end{aligned} \tag{5.5}$$

However, optimization of mutual information $I_m(Y;S)$ is difficult and one can choose the lower bound of $I_m(Y;S)$ as (Sharma et al. 2018)

$$I_m^L = \log|\mathcal{S}| - \log\left[1 + \frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\sum_{\substack{i=1\\ i\neq j}}^{|\mathcal{S}|}\exp\left(-\frac{1}{4N_0}||s_j - s_i||^2\right)\right]. \tag{5.6}$$

Further, the shaping gain of the constellation points is optimized to get an improved performance of the codebooks. It is widely accepted that a high value of shaping gain is obtained by adopting a nearly circular constellation boundary. Moreover, the irregularity in the constellation points tends to increase the diversity, which results in a higher value of shaping gain (Sharma et al. 2018). The shaping gain of a region $\mathcal{R}$ is expressed as (Forney and Wei 1989)

$$\gamma_s(\mathcal{R}) = \frac{[V(\mathcal{R})]^{\frac{2}{n}}}{6\mathcal{E}_{\text{av}}} \tag{5.7}$$

where $V(\mathcal{R})$ is the volume of the region $\mathcal{R}$, $n$ is the dimension, and $\mathcal{E}_{\text{av}}$ is the average energy of the constellation points.

The minimum Euclidean distance $d_{\min}$ of the constellation points plays an important role in the design of codebooks apart from the shaping gain and mutual information. In this example, we considered Euclidean distance as unity. Therefore, a constraint on the constellation points is imposed as

$$\mathcal{E}_{\text{av}} = \frac{1}{M}\sum_{m=1}^{M}\|\boldsymbol{x}_m\|^2 = 1.$$

Therefore, one can jointly optimize mutual information, shaping gain, and the minimum Euclidean distance to design codebooks. However, this joint optimization is difficult and a sub-optimal approach can be employed by optimizing each one of the objectives separately.

Let $M = 4$, $N = 2$, and $d_f = 3$. One can select 4-points using pulse amplitude modulation (PAM) constellation points which are equally spaced in line or space. These constellation points correspond to the first user and are denoted as $\mathcal{U}_1$. $\mathcal{U}_1$ points are expressed as $\mathcal{U}_1 = \{-1, -0.333, 0.333, 1\}$ and these points are scaled to get the unit average energy. Points of other two interfering users $\mathcal{U}_2$ and $\mathcal{U}_3$ are generated by judiciously rotating the constellation points $\mathcal{U}_1$. The angles of $\mathcal{U}_2$ and $\mathcal{U}_3$ are generated by maximizing the mutual information and the shaping gain:

$$\{\theta_2^\star, \theta_3^\star\} = \arg\max_{\{\theta_2,\theta_3\}\in[0\ 2\pi)} I_m^L \tag{5.8}$$

where $I_m^L$ is given in (5.6). $\{\theta_2^\star, \theta_3^\star\}$ are generated using (5.8) and then the shaping gain is optimized by considering the irregularity in the constellations $\mathcal{U}_2$ and $\mathcal{U}_3$. Additionally, during optimization, the Euclidean distance among the constellation points is kept as fixed. The complete codebook from constellation points, $\mathcal{U}_1, \mathcal{U}_2$, and $\mathcal{U}_3$ are distributed over the resources as

$$\mathcal{I} = \begin{bmatrix} \mathcal{U}_1 & 0 & \mathcal{U}_2 & 0 & \mathcal{U}_3 & 0 \\ 0 & \mathcal{U}_2 & \mathcal{U}_3 & 0 & 0 & \mathcal{U}_1 \\ \mathcal{U}_2 & 0 & 0 & \mathcal{U}_1 & 0 & \mathcal{U}_3 \\ 0 & \mathcal{U}_1 & 0 & \mathcal{U}_3 & \mathcal{U}_2 & 0 \end{bmatrix},$$

where $\mathcal{I}$ indicator matrix. The values of optimum angles $\theta_2^\star = 60°, \theta_3^\star = 120°$ are obtained at 10 dB. The generated constellation points are given as

$$\begin{aligned} \mathcal{U}_1 &= [-1, -0.333, 0.333, 1] \\ \mathcal{U}_2 &= [-0.1109 - 0.3i, 0.6 + 1i, -0.6 - 1i, 0.1109 + 0.3i] \\ \mathcal{U}_3 &= [0.3 - 0.3i, -0.6 + 1i, 0.6 - 1i, -0.3 + .3i]. \end{aligned} \tag{5.9}$$

The codebooks of all users in the $6 \times 4$ SCMA system are expressed in Table 5.1.

### 5.2.3 Symbol Detection in SCMA

In this section, the detection of the users (known as multi-user detection (MUD)) is described. Usually message passing algorithm (MPA) is used for MUD. The MPA can be best explained graphically in terms of the factor graph. Therefore, first we explain the factor graph.

#### 5.2.3.1 Factor Graph Representation

The codewords in SCMA are sparse, and only a few components of a codeword are non-zero. This feature can be specified in terms of a $K \times J$ matrix $F$ called the factor graph matrix. For example, consider $F$ shown in (5.10) for an SCMA system with $J = 6, K = 4$, and $\lambda = 150\%$.

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{5.10}$$

The 1s present in the $j$th column specify the locations of non-zero components of the codewords for the $j$th user. Since, the codewords are sparse, $F$ is also sparse. The matrix $F$ can also be graphically represented by a factor graph as shown in

**Table 5.1** Designed codebooks of all 6 users

$$\mathcal{X}_1 = \left\{ \begin{bmatrix} -1 \\ 0 \\ -0.1109 - 0.3i \\ 0 \end{bmatrix} \begin{bmatrix} -0.333 \\ 0 \\ 0.6 + 1i \\ 0 \end{bmatrix} \begin{bmatrix} 0.333 \\ 0 \\ -0.6 - 1i \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0.1109 + 0.3i \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{X}_2 = \left\{ \begin{bmatrix} 0 \\ -0.1109 - 0.3i \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0.6 + 1i \\ 0 \\ -0.333 \end{bmatrix} \begin{bmatrix} 0 \\ -0.6 - 1i \\ 0 \\ 0.333 \end{bmatrix} \begin{bmatrix} 0 \\ 0.1109 + 0.3i \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$\mathcal{X}_3 = \left\{ \begin{bmatrix} -0.6 - 1i \\ 0.3 - 0.3i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -0.1109 - 0.3i \\ -0.6 + 1i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.1109 + 0.3i \\ 0.6 - 1i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.6 + 1i \\ -0.3 + .3i \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{X}_4 = \left\{ \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0.3 - 0.3i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.333 \\ -0.6 + 1i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0.333 \\ 0.6 - 1i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ -0.3 + .3i \end{bmatrix} \right\}$$

$$\mathcal{X}_5 = \left\{ \begin{bmatrix} 0.3 - 0.3i \\ 0 \\ 0 \\ -0.6 - 1i \end{bmatrix} \begin{bmatrix} -0.6 + 1i \\ 0 \\ 0 \\ -0.1109 - 0.3i \end{bmatrix} \begin{bmatrix} 0.6 - 1i \\ 0 \\ 0 \\ 0.1109 + 0.3i \end{bmatrix} \begin{bmatrix} -0.3 + .3i \\ 0 \\ 0 \\ 0.6 + 1i \end{bmatrix} \right\}$$

$$\mathcal{X}_6 = \left\{ \begin{bmatrix} 0 \\ -1 \\ 0.3 - 0.3i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.333 \\ -0.6 + 1i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0.333 \\ 0.6 - 1i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -0.3 + .3i \\ 0 \end{bmatrix} \right\}$$

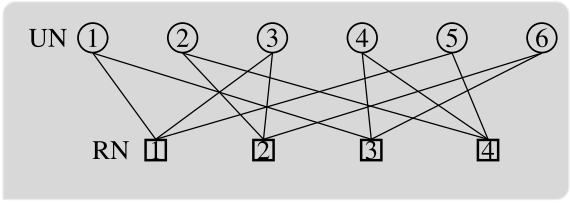



**Fig. 5.5** Factor graph of six users ($J = 6$) and four resource nodes ($K = 4$) where three users connect with one resource node

Fig. 5.5. Corresponding to each column and each row of $F$, there is a user node and a resource node, respectively. And, against a '1' in $F$, there is an edge between the corresponding user node and resource node.

#### 5.2.3.2 Detection Using MPA

The MUD for SCMA is a crucial task. The factor graph of the SCMA system is sparse. Therefore, MUD can be carried out by applying MPA over the sparse graph.

The MPA and its variants have been extensively used in many applications like decoding of low-density parity-check (LDPC), detection in multiple-input multiple-output (MIMO) antenna systems, etc.

In the following, we describe the version of MPA which is used in MUD for SCMA system. The underlying principle of the MPA is the maximum *a posteriori* rule. It is numerically exhaustive to perform the MAP detection for all the users in a block at one shot. Instead, user-wise MAP detection is carried out. The user-specific rule is given by

$$\hat{\mathbf{x}}_j = \arg \max_{\mathbf{x}_{jm} \in \mathcal{X}_j} V_j \left( \mathbf{x}_{jm} \right) \tag{5.11}$$

where $\mathcal{X}_j$ is the codebook dedicated for the $j$th user, $\mathbf{x}_{jm}$ is the $m$th codeword of $\mathcal{X}_j$, and $V_j \left( \mathbf{x}_{jm} \right)$ is the a posteriori probability of the $j$th user's codeword being $\mathbf{x}_{jm}$ given the received signal $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T$ with $m = 1, 2, \ldots, M$.

The detection in (5.11) is carried out iteratively by applying MPA. The detection process can be conveniently visualized graphically in terms of the factor graph as shown in Fig. 5.6. At the outset, the signals from the channel enter the resource nodes as shown in Fig. 5.6. Note that unlike the LDPC codes where the user nodes (variable nodes) receive the channel values, here, the resource nodes (check nodes) receive the same. Then the user nodes and the resource nodes exchange messages iteratively to extract the data of different users. The steps for MPA-based MUD are presented (Sharma et al. 2018) below.

Let $U_{k \to j}^{(l)}$ and $V_{j \to k}^{(l)}$ be the messages sent by the $k$th resource node to the $j$th user node and by the $j$th user node to the $k$th resource node, respectively, during the $l$th iteration. Here, the messages are vectors of length $M$. The $m$th components $U_{k \to j}^{(l)} \left( \mathbf{x}_{jm} \right)$ and $V_{j \to k}^{(l)} \left( \mathbf{x}_{jm} \right)$ correspond to the $m$th codeword $\mathbf{x}_{jm}$ present in the codebook $\mathcal{X}_j$. The steps of the MPA are outlined below.

1. ***Initialization***: At the beginning, the messages are equally likely. Therefore, set $V_{j \to k}^{(0)} \left( \mathbf{x}_{jm} \right) = 1/M$. Note that in the case of the LDPC codes, the initialization is done according to the received channel values. Here, since the user nodes do not receive the signals from the channel, the messages (probability values) from the user nodes are set uniformly over the all possible symbols.

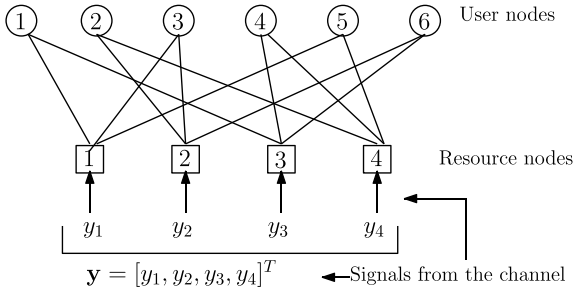



**Fig. 5.6** MPA-based detection over factor graph

2. ***Resource node update***: The resource nodes receive the output signals $\mathbf{y}$ from the channel. The messages from the resource nodes are updated as follows:

$$U_{k\to j}^{(l)}(\mathbf{x}_{jm}) = \sum_{\mathbf{c}\in C_{kj}} \frac{1}{\pi N_0} \exp\Bigg[ -\frac{1}{N_0}|y_k - \sqrt{P_j}h_{kj}x_{jmk} - \sum_{j'\in\mathcal{M}_k^j} \sqrt{P_{j'}}h_{ki}c_{j'k}|^2\Bigg] \prod_{j'\in\mathcal{M}_k^j} V_{j'\to k}^{(l-1)}\left(\mathbf{c}_{j'}\right) \tag{5.12}$$

where,

- $\mathcal{M}_k$ is the set of the user nodes connected to the $k$th resource node and $\mathcal{M}_k^j = \mathcal{M}_k \setminus \{j\}$.
- Suppose $\mathcal{M}_k^j = \left\{j_1, j_2, \ldots, j_{d_f-1}\right\}$. Then, $C_{kj}$ is the Cartesian product of the codebooks as defined below:

$$C_{kj} = \mathcal{X}_{j_1} \times \mathcal{X}_{j_2} \times \cdots \times \mathcal{X}_{j_{d_f-1}}.$$

- $\mathbf{c} = \left(\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \ldots, \mathbf{c}_{j_{d_f-1}}\right)$ is a member of $C_{kj}$ and $\mathbf{c}_{j'} = \left(c_{j'1}, c_{j'2}, \ldots, c_{j'K}\right)$.
- $\mathbf{x}_{jm} = \left(x_{jm1}, x_{jm2}, \ldots, x_{jmK}\right)$.

The step in (5.12) is known as the sum-product rule. The AWGN channel model is assumed.

3. ***User node update***: The messages from the user nodes are updated as follows:

$$V_{j\to k}^{(l)}\left(\mathbf{x}_{jm}\right) = \prod_{k'\in\mathcal{N}_j^k} U_{k'\to j}^{(l)}(\mathbf{x}_{jm}) \tag{5.13}$$

where $\mathcal{N}_j$ is the set of the resource nodes connected to the $j$th user node and $\mathcal{N}_j^k = \mathcal{N}_j \setminus \{k\}$.

4. ***Stopping rule***: If the messages get converged or the maximum number of iterations $\kappa$ is exhausted, then stop and proceed for decision making in Step (5). Otherwise, set $l = l + 1$ and go to Step (2).

5. ***Decision***: Compute

$$V_j\left(\mathbf{x}_{jm}\right) = \prod_{k\in\mathcal{N}_j} U_{k\to j}^{(l)}(\mathbf{x}_{jm}). \tag{5.14}$$

The estimates of the transmitted codewords are found as follows:

$$\hat{\mathbf{x}}_j = \arg\max_{\mathbf{x}_{jm}\in\mathcal{X}_j} V_j\left(\mathbf{x}_{jm}\right).$$

The probability-domain MPA is usually numerically unstable. Therefore, the messages must be normalized in each iteration.

### 5.2.4 Example

Now we give a complete example of the SCMA system. We explain the procedure for SCMA encoding and detection. Suppose, the codebooks shown in Table 5.2 are considered.

Consider the following data symbols for the 6 users: (2, 2, 1, 1, 3, 4). Then, the set of the codewords chosen for transmission can be presented as shown below:

$$\begin{pmatrix} -0.4022 & 0 & -0.7247-1.2078i & 0 & 0.7247-1.2078i & 0 \\ 0 & 0.7247+1.2078i & 0.3623-0.3623i & 0 & 0 & 1.2078 \\ 0.7247+1.2078i & 0 & 0 & -1.2078 & 0 & -0.3623+0.3623i \\ 0 & -0.4022 & 0 & 0.3623-0.3623i & 0.1339+0.3623i & 0 \end{pmatrix}$$

In the above, the $i$th column of the matrix represents the codeword transmitted by the $i$th user. Note that here the number of resources is $K = 4$ and therefore the length of each codeword is also 4. The sum of the superimposed codewords is given by:

**Table 5.2** Codebooks of 6 users

$$\mathcal{X}_1 = \left\{ \begin{bmatrix} -1.2078 \\ 0 \\ -0.1339-0.3623i \\ 0 \end{bmatrix} \begin{bmatrix} -0.4022 \\ 0 \\ 0.7247+1.2078i \\ 0 \end{bmatrix} \begin{bmatrix} 0.4022 \\ 0 \\ -0.7247-1.2078i \\ 0 \end{bmatrix} \begin{bmatrix} 1.2078 \\ 0 \\ 0.1339+0.3623i \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{X}_2 = \left\{ \begin{bmatrix} 0 \\ -0.1339-0.3623i \\ 0 \\ -1.2078 \end{bmatrix} \begin{bmatrix} 0 \\ 0.7247+1.2078i \\ 0 \\ -0.4022 \end{bmatrix} \begin{bmatrix} 0 \\ -0.7247-1.2078i \\ 0 \\ 0.4022 \end{bmatrix} \begin{bmatrix} 0 \\ 0.1339+0.3623i \\ 0 \\ 1.2078 \end{bmatrix} \right\}$$

$$\mathcal{X}_3 = \left\{ \begin{bmatrix} -0.7247-1.2078i \\ 0.3623-0.3623i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -0.1339-0.3623i \\ -0.7247+1.2078i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.1339+0.3623i \\ 0.7247-1.2078i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.7247+1.2078i \\ -0.3623+0.3623i \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{X}_4 = \left\{ \begin{bmatrix} 0 \\ 0 \\ -1.2078 \\ 0.3623-0.3623i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.4022 \\ -0.7247+1.2078i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0.4022 \\ 0.7247-1.2078i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1.2078 \\ -0.3623+0.3623i \end{bmatrix} \right\}$$

$$\mathcal{X}_5 = \left\{ \begin{bmatrix} 0.3623-0.3623i \\ 0 \\ 0 \\ -0.7247-1.2078i \end{bmatrix} \begin{bmatrix} -0.7247+1.2078i \\ 0 \\ 0 \\ -0.1339-0.3623i \end{bmatrix} \begin{bmatrix} 0.7247-1.2078i \\ 0 \\ 0 \\ 0.1339+0.3623i \end{bmatrix} \begin{bmatrix} -0.3623+0.3623i \\ 0 \\ 0 \\ 0.7247+1.2078i \end{bmatrix} \right\}$$

$$\mathcal{X}_6 = \left\{ \begin{bmatrix} 0 \\ -1.2078 \\ 0.3623-0.3623i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.4022 \\ -0.7247+1.2078i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0.4022 \\ 0.7247-1.2078i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1.2078 \\ -0.3623+0.3623i \\ 0 \end{bmatrix} \right\}$$

$$\mathbf{r} = \begin{pmatrix} -0.4022 - 2.4156i \\ 2.2948 + 0.8454i \\ -0.8454 + 1.5701i \\ 0.0941 + 0.0000i \end{pmatrix}.$$

We consider a simple complex AWGN channel model. Suppose, the AWGN noise samples at $\frac{E_b}{N_0} = 3\,\text{dB}$ are given by:

$$\mathbf{n} = \begin{pmatrix} 0.0018 - 1.2008i \\ 1.2143 + 0.4227i \\ 0.3323 + 0.4232i \\ 0.1488 - 0.6993i \end{pmatrix}$$

Then the 4-dimensional received signal $\mathbf{y} = \mathbf{r} + \mathbf{n}$ becomes

$$\mathbf{y} = \begin{pmatrix} -0.4004 - 3.6164i \\ 3.5091 + 1.2681i \\ -0.5132 + 1.9933i \\ 0.2428 - 0.6993i \end{pmatrix}$$

Given this value of $\mathbf{y}$, the individual user's data symbol is to be estimated now. For that we consider the probability-domain MPA described in Sect. 5.2.3.2. In the MPA-based SCMA detection process, the messages are vectors of length $M$ (in this case $M = 4$). There is no exchange of messages between a user node and a resource node if they are not connected through an edge. The vector messages for a particular direction in the entire factor graph can be stored in a matrix of size $KM \times J$. First, the messages from the user nodes are initialized equally likely as described in Sect. 5.2.3.2. These initialized messages are stored in the matrix $\mathbf{V}$ as shown below.

$$\mathbf{V} = \left(\begin{array}{cccccc} 0.25 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.25 & 0 & 0.25 & 0 \\ \hline 0 & 0.25 & 0.25 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0.25 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0.25 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0.25 & 0 & 0 & 0.25 \\ \hline 0.25 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0.25 \\ \hline 0 & 0.25 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0.25 & 0.25 & 0 \end{array}\right) \tag{5.15}$$

Observe from (5.15) that $\mathbf{V}$ contains $K = 4$ blocks of rows of size $M = 4$. The number of columns is $J = 6$. After the initialization step, the messages from the resource nodes are updated according to (5.12). These messages are stored in a matrix $\mathbf{U}$ as shown below:

$$\mathbf{U} = \left(\begin{array}{cccccc}
0.1764 & 0 & 0.9962 & 0 & 0.0042 & 0 \\
0.6415 & 0 & 0.0038 & 0 & 0.0000 & 0 \\
0.1784 & 0 & 0.0000 & 0 & 0.9958 & 0 \\
0.0038 & 0 & 0.0000 & 0 & 0.0000 & 0 \\
\hline
0 & 0.0000 & 0.7860 & 0 & 0 & 0.0000 \\
0 & 0.9986 & 0.0000 & 0 & 0 & 0.0000 \\
0 & 0.0000 & 0.2042 & 0 & 0 & 0.0079 \\
0 & 0.0014 & 0.0098 & 0 & 0 & 0.9921 \\
\hline
0.0328 & 0 & 0 & 0.2660 & 0 & 0.0168 \\
0.6324 & 0 & 0 & 0.4763 & 0 & 0.6641 \\
0.0002 & 0 & 0 & 0.2362 & 0 & 0.0000 \\
0.3347 & 0 & 0 & 0.0215 & 0 & 0.3191 \\
\hline
0 & 0.1512 & 0 & 0.3688 & 0.2511 & 0 \\
0 & 0.3453 & 0 & 0.0387 & 0.3669 & 0 \\
0 & 0.2673 & 0 & 0.3406 & 0.3087 & 0 \\
0 & 0.2362 & 0 & 0.2519 & 0.0733 & 0
\end{array}\right) \tag{5.16}$$

In order to demonstrate the resource node update, we consider the calculation of $U_{1\to 1}(1)$ during the first iteration. Note that the 1st resource node is connected to user nodes 1, 3, and 5. Since we are calculating the 1st component of $U_{1\to 1}$, the data symbol $v_1$ for the 1st user node is fixed at 1. There are $M^{d_f-1} = 4^2 = 16$ different combinations for $v_3$ and $v_5$. Table 5.3 shows the detailed calculation for each of the combinations.

As an example, consider the computation of the combination corresponding to $v_1 = 1, v_3 = 2, v_5 = 3$. Here, $x_{111} = -1.2078, c_{32} = -0.1339 - 0.3623i$, and $c_{53} = 0.7247 - 1.2078i$. At $\frac{E_b}{N_0} = 3\text{dB}$, we have $N_0 = 0.5012$. With these values, the term is computed as:

$$\begin{aligned}
\text{term} &= \frac{1}{\pi N_0} \exp\left[-\frac{1}{N_0}\,|y_1 - x_{111} - c_{32} - c_{53}|^2\right] \times V_{3\to 1}(2) \times V_{5\to 1}(3) \\
&= \frac{1}{0.5\pi} \exp\left[-\frac{1}{0.5}\,|-0.4 - 3.6i + 1.2 + 0.1 + 0.36i - 0.7 + 1.2i|^2\right] (0.25)^2 \\
&= 8.5\text{e}{-6}
\end{aligned}$$

Finally, the terms corresponding to all the combinations are summed up to produce $U_{1\to 1}(1) = 6.2e-4$. The values of $U_{1\to 1}(2)$, $U_{1\to 1}(3)$, and $U_{1\to 1}(4)$ are $0.0022$, $6.2471e-4$ and $1.33e-5$, respectively. These components are normalized so that we have $\mathbf{U}_{1\to 1} = [0.1764, 0.6415, 0.1784, 0.0038]^T$.

**Table 5.3** Computation of $U_{1\to1}(1)$: A total of 16 terms/combinations in the sum

| $v_1$ | $v_3$ | $v_5$ | Term |
|---|---|---|---|
| 1 | 1 | 1 | 6.1e−7 |
| 1 | 1 | 2 | 7.1e−18 |
| 1 | 1 | 3 | 6.1e−4 |
| 1 | 1 | 4 | 6.9e−12 |
| 1 | 2 | 1 | 1.2e−9 |
| 1 | 2 | 2 | 8.8e−22 |
| **1** | **2** | **3** | **8.5e−6** |
| 1 | 2 | 4 | 6.2e−15 |
| 1 | 3 | 1 | 1.5e−13 |
| 1 | 3 | 2 | 3.9e−27 |
| 1 | 3 | 3 | 8.8e−9 |
| 1 | 3 | 4 | 2.2e−19 |
| 1 | 4 | 1 | 1.9e−19 |
| 1 | 4 | 2 | 3.2e−34 |
| 1 | 4 | 3 | 8.1e−14 |
| 1 | 4 | 4 | 1.3e−25 |
| $U_{1\to1}(1)$ (sum) | | | 6.2e−4 |

After the computation of all the messages from the resource nodes, the messages from the user nodes are updated according to (5.13). These messages are shown below:

$$\mathbf{V} = \begin{pmatrix} 0.0328 & 0 & 0.7860 & 0 & 0.2511 & 0 \\ 0.6324 & 0 & 0.0000 & 0 & 0.3669 & 0 \\ 0.0002 & 0 & 0.2042 & 0 & 0.3087 & 0 \\ 0.3347 & 0 & 0.0098 & 0 & 0.0733 & 0 \\ 0 & 0.1512 & 0.9962 & 0 & 0 & 0.0168 \\ 0 & 0.3453 & 0.0038 & 0 & 0 & 0.6641 \\ 0 & 0.2673 & 0 & 0 & 0 & 0 \\ 0 & 0.2362 & 0.0000 & 0 & 0 & 0.3191 \\ 0.1764 & 0 & 0 & 0.3688 & 0 & 0 \\ 0.6415 & 0 & 0 & 0.0387 & 0 & 0 \\ 0.1784 & 0 & 0 & 0.3406 & 0 & 0.0079 \\ 0.0038 & 0 & 0 & 0.2519 & 0 & 0.9921 \\ 0 & 0 & 0 & 0.2660 & 0.0042 & 0 \\ 0 & 0.9986 & 0 & 0.4763 & 0 & 0 \\ 0 & 0 & 0 & 0.2362 & 0.9958 & 0 \\ 0 & 0.0014 & 0 & 0.0215 & 0 & 0 \end{pmatrix}$$

Then the *a posteriori* probability vectors are updated as per (5.14). These values are shown below:

$$\mathbf{V}_{\text{posterior}} = \begin{pmatrix} 0.0058 & 0.0000 & 0.7830 & 0.0981 & 0.0010 & 0.0000 \\ 0.4056 & 0.3448 & 0.0000 & 0.0184 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0804 & 0.3074 & 0.0000 \\ 0.0013 & 0.0003 & 0.0000 & 0.0054 & 0.0000 & 0.3166 \end{pmatrix}$$

Then by identifying the indices of the maximum component in the columns, the user's symbols are estimated as 2, 2, 1, 1, 3, 4. Note that the normalization must be applied only to $\mathbf{U}$ and $\mathbf{V}$ as these will be used further in the next round of iterations. Usually there is no concrete stopping rule as in the case of LDPC codes where a zero syndrome vector terminates the iterations. Instead, here, the algorithm is stopped when there is no significant changes in $\mathbf{V}_{\text{posterior}}$ or the maximum number of iterations are exhausted.

### 5.2.5 *Numerical Results and Discussions*

In this section, the numerical results are shown for an SCMA system.

In Fig. 5.7, average SER performance of SCMA system is shown. We consider SCMA system performance using various codebooks for comparison in Fig. 5.7. Further, in Fig. 5.7 the SCMA system's SER performance for nine users ($J = 9$) and six resource elements ($K = 6$) over AWGN channel is considered. Therefore, SCMA system's performance depends on the design of codebooks of each user as observed in Fig. 5.7. The gain using the codebook in Sharma et al. (2018) is around 1 dB as observed in Fig. 5.7.

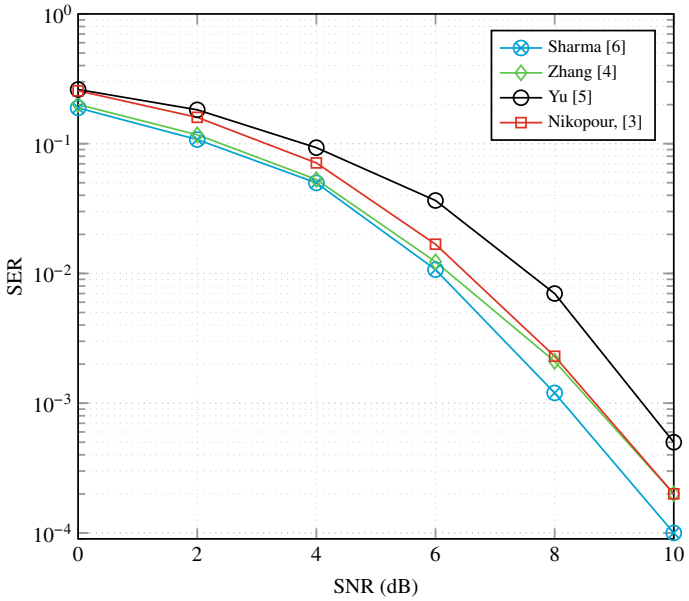



**Fig. 5.7** BER performance of the SCMA system using various codebooks for $J = 9$ and $K = 6$ (overloading factor $\lambda = 150\%$ ) in AWGN channel

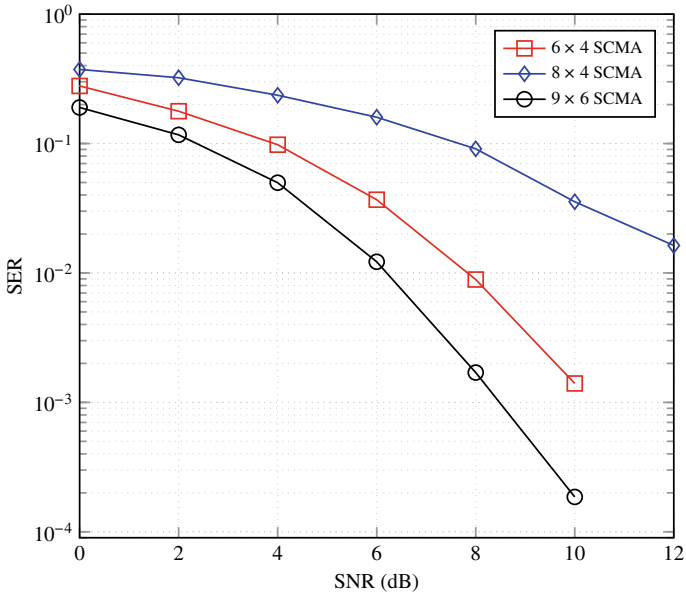



**Fig. 5.8** BER performance of the SCMA system using various $J$ and $K$ in AWGN channel

Further, $6 \times 4$, $8 \times 4$, and $9 \times 6$ SCMA system's SER performance is shown in Fig. 5.8 over AWGN channel. $8 \times 4$ SCMA system has worst performance due to high overloading factor as compared to $6 \times 4$ and $9 \times 6$ SCMA systems. Further, $9 \times 6$ SCMA system has better performance than the $6 \times 4$ due to less symbol collisions among the users. Furthermore, same results are observed over Rayleigh fading channel in Fig. 5.9.

## 5.3 Hybrid Multiple Access-Based System Design

In this section, we discuss a hybrid multiple access (HMA) approach to further enhance the spectral efficiency of next-generation wireless network. In real scenario, users' distribution around the base station is uniform. Further, based on users' channel gain, users are grouped into near users (NU) and far users (FU) in a cell. FU have higher pathless than NU in system. Furthermore, the difference can also be possible between the number of FU (more) and the NU (less) in a system due to a higher area of peripheral than the near area of a base station. Therefore, a conventional PD-NOMA approach is not optimal for pairing all the FU and NU in a cell. Since, some FU are left after NU and FU pairing, as shown in Fig. 5.11. Further, the SCMA-based method to connect all the users to a base station is not optimal since it requires multiple SCMA group, where each SCMA group uses orthogonal resources. Therefore, a

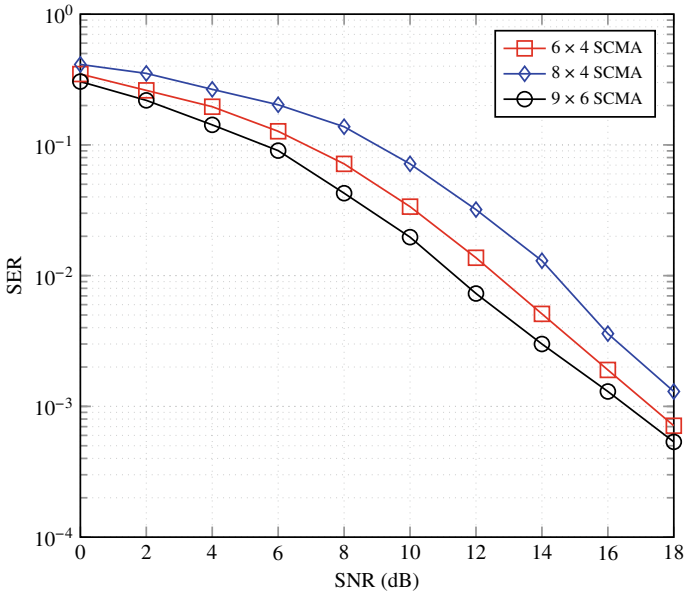



**Fig. 5.9** BER performance of the SCMA system using various $J$ and $K$ in Rayleigh channel

HMA scheme is considered to connect all the users in a system using minimum number of resources.

In HMAS, users are divided into multiple groups based on their channel gain difference. For example, in a downlink scenario, users are partitioned into two groups, say Group-1 and Group-2, as shown in Fig. 5.11. Group-1 and Group-2 are near and far to a base station and is referred to as near group (or strong group) and far group (or weak group), respectively. Let $J_1$ and $J_2$ users in Group-1 and Group-2, respectively.

The strong and the weak users in HMA-based system are denoted by $\{s_1, s_2, \ldots, s_{J_1}\}$ and $\{w_1, w_2, \ldots, w_{J_2}\}$, respectively. Let $\left\{\mathbf{h}_i^s\right\}_{i=1}^{J_1}$ and $\left\{\mathbf{h}_j^w\right\}_{j=1}^{J_2}$ denote the channel impulse response vectors experienced by the strong and the weak users, respectively. Therefore, $J_1 \times K$ and $J_2 \times K$ SCMA systems for *Group 1* based on a codebook $C_1$ and *Group 2*, based on a codebook $C_2$, are considered, respectively. Let $\mathbf{x}_i^s \in C_1$ and $\mathbf{x}_j^w \in C_2$ denote the codewords for the $i$th strong user and the $j$th weak user, respectively. The superimposed codewords of all users at the transmitter are written as:

$$\mathbf{x} = \sum_{i=1}^{J_1} \sqrt{P_i^s}\mathbf{x}_i^s + \sum_{j=1}^{J_2} \sqrt{P_j^w}\mathbf{x}_j^w$$

where $P_i^s$ and $P_j^w$ denote the powers assigned to the $i$th strong and $j$th weak user, respectively. The received signal vectors at the strong and the weak users are given by

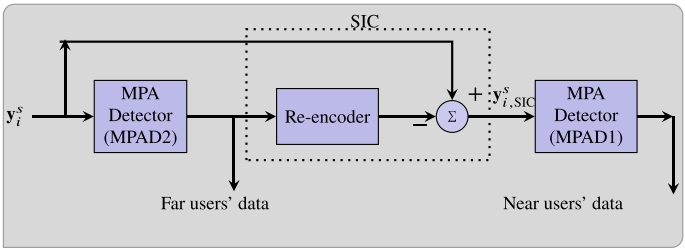



**Fig. 5.10** Detection at the location of the strong/near users (Sharmaet al. 2019a)

$$\begin{aligned}\mathbf{y}_i^s &= \text{diag}\left(\mathbf{h}_i^s\right)\mathbf{x} + \mathbf{n}_i^s, \quad i = 1, 2, \ldots J_1 \quad \text{and} \\ \mathbf{y}_j^w &= \text{diag}\left(\mathbf{h}_j^w\right)\mathbf{x} + \mathbf{n}_j^w, \quad j = 1, 2, \ldots J_2\end{aligned} \tag{5.17}$$

respectively.

MPA detector for the users in *Group 1* operates over the factor graph corresponding to the codebook $C_1$. Such an MPA detector is denoted by MPAD1. Similarly, the MPA detector for *Group 2* users is referred to as MPAD2. First, the detection of the strong users is explained. The detection of the $i$th strong user $s_i$ is illustrated in Fig. 5.10. The SIC of the far users is performed before extracting the data of $s_i$. First, with $\mathbf{y}_i^s$ as input, MPAD2 produces the estimated codewords $\left\{\hat{\mathbf{x}}_j^{w_i}\right\}_{j=1}^{J_2}$. Here, the superscript '$i$' signifies that the detection process is carried out by the $i$th receiver at the location of $s_i$. These estimates are used only for detecting the data of $s_i$. The weak users' signal is recreated and the interference-canceled received signal is obtained as

$$\mathbf{y}_{i,\text{SIC}}^s = \mathbf{y}_i^s - \sum_{j=1}^{J_2} \text{diag}\left(\mathbf{h}_j^w\right)\sqrt{P_j^w}\hat{\mathbf{x}}_j^{w_i}.$$

Using $\mathbf{y}_{i,\text{SIC}}^s$ as the input, the MPA detector MPAD1 estimates the data of $\left\{s_1, \ldots, s_{J_1}\right\}$. From these, the estimate for only $s_i$ is considered while ignoring the rest.

The detection of the weak users is carried out similarly as in the conventional downlink SCMA system. Observe that, in the proposed system, the overloading factor is increased from $\lambda = \frac{J_c}{K}$ to $\lambda_{\text{proposed}} = \frac{J_1+J_2}{K}$ where $J_c$ is the number of users in a conventional SCMA system.

HMAS-based system's average SER performance is shown in Fig. 5.12 with conventional SCMA and PD-NOMA-based systems' performance. HMA and PD-NOMA-based system involve detection based on SIC principle, which suffers from error propagation. Therefore, these systems are outperformed by the purely SCMA-based systems,which involve powerful MPA detection. HMA-based system has better performance than conventional PD-NOMA or SCMA system with a higher overloading factor as observed in Fig. 5.12.

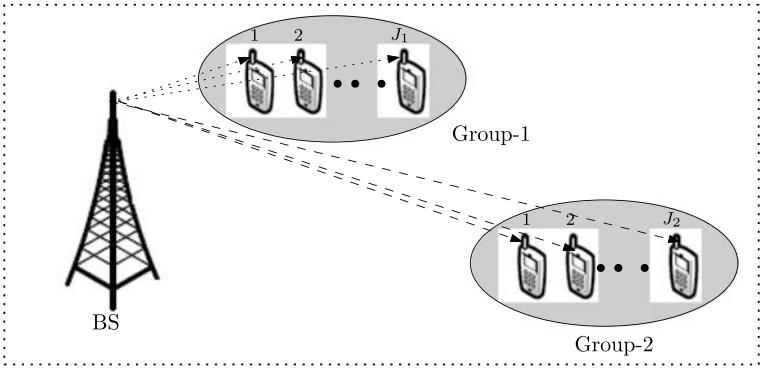



**Fig. 5.11** Hybrid multiple access system

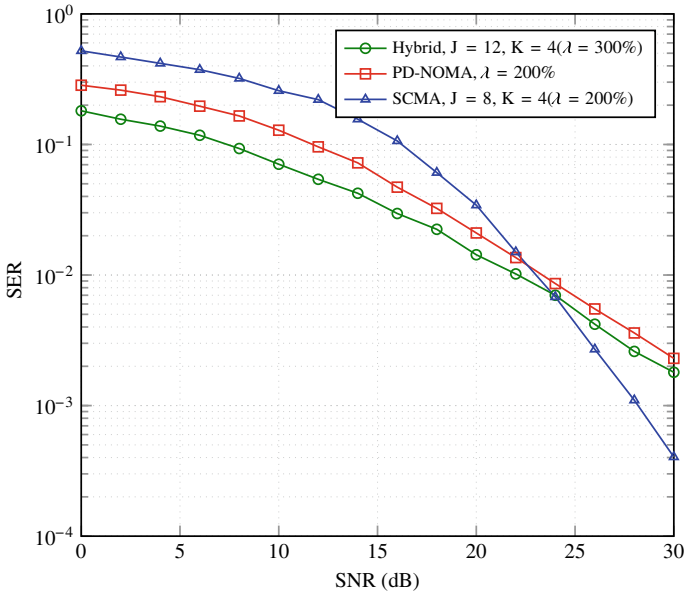



**Fig. 5.12** SER over Rayleigh channel

### *5.3.1 Future Research Direction*

SCMA and HMA schemes can be used to improve the spectral efficiency of 5G and beyond networks. However, some design and implementation issues have to be solved for practical implementation of SCMA and HMA schemes such as follows:

- SCMA or HMA-based system's performance depends on the codebooks assigned to the each user. Therefore, an efficient and universal method of codebooks

design is essential. However, currently, codebooks are designed using sub-optimal approaches and also this design depends on the overloading factor in the system.
- In general, MPA-based detection algorithms are used in SCMA or HMA. However, MPA has higher complexity for higher modulation order size and/or resource degree of each node in factor graph. Therefore, some simple algorithms are essential for practical implementation of SCMA-based system at higher data rate.
- An optimal power allocation scheme is also required to split the total power among the groups in HMA-based system. Since the total sum rate of HMA system can be enhanced using optimal power allocation.
- In practice, the perfect channel information is not available at receiver. Therefore, effect of imperfect channel should be analyzed in SCMA or HMA-based system for detection and estimation.
- Further, massive multi-input and multi-output (MIMO) will be an integral part of 5G and beyond wireless network. Therefore, SCMA and HMA methods should be analyzed with massive MIMO system.

## References

Al-Nahhal I, Dobre OA, Basar E, Ikki S (2019) Low-cost uplink sparse code multiple access for spatial modulation. IEEE Trans Veh Technol 68(9):9313–9317

Dai J, Niu K, Dong C, Lin J (2017) Improved message passing algorithms for sparse code multiple access. IEEE Trans Veh Technol 5(11):9986–9999

Dai X, Zhang Z, Bai B, Chen S, Sun S (2018) Pattern division multiple access: a new multiple access technology for 5G. IEEE Wirel Commun 25(2):54–60

Ding Z, Lei X, Karagiannidis GK, Schober R, Yuan J, Bhargava VK (2017) A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends. IEEE J Sel Areas Commun 35(10):2181–2195

Forney GD, Wei LF (1989) Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations. IEEE J Sel Area Commun 7(6):877–892

Khan WU, Yu Z, Yu S, Sidhu GAS, Liu J (2018) Efficient power allocation in downlink multi-cell multi-user NOMA networks. IET Commun 13(4):396–402

Nikopour H, Baligh H (2013) Sparse code multiple access. In: IEEE 24th international symposium on personal indoor and mobile radio communications (PIMRC), pp 332–336

Pan Z, Luo J, Lei J, Wen L, Tang C (2018) Uplink spatial modulation SCMA system. IEEE Commun Lett 23(1):184–187

Sergienko B, Klimentyev VP (2016) SCMA detection with channel estimation error and resource block diversity. In: IEEE international Siberian conference on control and communications, pp 1–5

Sharma S, Deka K, Bhatia V, Gupta A (2018) SCMA codebook based on optimization of mutual information and shaping gain. In: IEEE Globecom workshops (GC Wkshps), pp 1–6

Sharma S, Deka K, Bhatia V (2019a) Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond. IEEE Commun Lett 23(6):971–974

Sharma S, Deka K, Hong Y (2019b) User activity detection-based large SCMA system for uplink grant-free access. In: IEEE international conference on communications workshops (ICC workshops), pp 1–6

Yan C, Kang G, Zhang N (2017) A dimension distance-based SCMA codebook design. IEEE Access 5:5471–5479

Yan C, Zhang N, Kang G (2018) Downlink multiple input multiple output mixed sparse code multiple access for 5G system. IEEE Access 6:20837–20847

Yang P, Xiao Y, Xiao M, Ma Z (2019) NOMA-aided precoded spatial modulation for downlink MIMO transmissions. IEEE J Sel Top Sig Process 13(3):729–738

Yu L, Lei X, Fan P, Chen D (2015) An optimized design of SCMA codebook based on star-QAM signaling constellations. In: IEEE international conference on wireless communications and signal processing (WCSP), pp 1–5

Yu L, Fan P, Cai D, Ma Z (2018) Design and analysis of SCMA codebook based on star-QAM signaling constellations. IEEE Trans Veh Technol 67(11):10543–10553

Yuan W, Wu N, Guo Q, Li Y, Xing C, Kuang J (2018) Iterative receivers for downlink MIMO-SCMA: message passing and distributed cooperative detection. IEEE Trans Wireless Commun 17(5):3444–3458

Zhang S, Xiao K, Xiao B, Chen Z, Xia B, Chen D, Ma S (2016) A capacity-based codebook design method for sparse code multiple access systems. In: IEEE international conference on wireless communications and signal processing (WCSP), pp 1–5

# Chapter 6
# Implementation of a Non-orthogonal Multiple Access Scheme Under Practical Impairments


**Tu-Trinh Thi Nguyen, Chi-Bao Le, and Dinh-Thuan Do**

**Abstract** This paper presents a practical scenario of non-orthogonal multiple access (NOMA) networks. In particular, a two-way NOMA network with hardware impairments is analyzed. To facilitate the system performance analysis, closed-from expressions are derived for the outage probability and throughput. Furthermore, to provide insights of performance, an asymptotic expression is introduced for the case of high SNR and then limitation of the system under hardware impairments is determined. Since successive interference cancelation (SIC) is adopted at each receiver, SIC affects the detection performance. As such, we consider both imperfect SIC (ipSIC) and perfect SIC cases. In addition, and we further compare system performance of NOMA with traditional orthogonal multiple access (OMA). Finally, to verify the analytical results Monte Carlo simulations presented.

**Keywords** Outage probability · Imperfect/perfect successive interference cancellation · Non-orthogonal multiple access · Impairments

## 6.1 Introduction

To support the high data rate wireless communications, non-orthogonal multiple access (NOMA) has been highly suggested as potential access technology applied in the fifth-generation (5G) communications systems (Ding et al. 2017; Nguyen and Do 2018). The main advantages of NOMA with respective to higher spectral efficiency can be obtained by employing at the transmitters and the receivers with

T.-T. Thi Nguyen · C.-B. Le
Faculty of Electronics Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
e-mail: nguyenthitutrinh@iuh.edu.vn

C.-B. Le
e-mail: lechibao@iuh.edu.vn

D.-T. Do (✉)
Wireless Communications Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
e-mail: dodinhthuan@tdtu.edu.vn



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_6

the superposition coding and successive interference cancelation (SIC), respectively (Do et al. 2019; Do and Le 2018; Song et al. 2017). In particular, SIC is deployed to eliminate interference among multiple users (Zhang et al. 2017a). They confirmed that NOMA can offer a significant development in terms of spectrum effectiveness (Zhang et al. 2017a). The multi-user NOMA system employing SIC receivers is introduced in many research activities to examine the performance of wireless systems. In addition, a randomly located user is considered as recent research in Ding et al. (2016) and Islam and Kwak (2016) to evaluate the outage performance by deploying the order statistic methodology for consideration on the downlink NOMA scenario. Regarding the decoding order, the instantaneous channel state information (CSI) is required at the base station (BS), but it costs extra overhead for CSI. The authors in Shi et al. (2016) studied the optimal power distributed feature and then the optimal decoding order is evaluated with statistical CSI to satisfy the outage of system over independent identical distributed (i.i.d.) channels. To consider uplink scenario with two NOMA users, the closed-form expressions of the outage probability are derived as in Zhang et al. (2016). To analyze the mean coverage probability, the authors in Tabassum et al. (2017) studied multi-cell uplink scenario. However, CSI is often assumed to simpler analysis, i.e., perfect knowledge of CSI is prior known at receivers. Practically, although high accuracy channel estimation is employed, then it is hard to achieve the perfect CSI. Therefore, such a imperfect CSI results in the channel estimation error together with significant overhead consuming in considered systems (Yang et al. 2018). Recently, a downlink NOMA system is studied in terms of the outage probability in case of imperfect CSI as in Yang et al. (2016), where interference is considered as the channel estimation error.

Recently, two-way relaying is considered as a promising technique to allow the base station process multiple signals at both uplink and downlink as in Nguyen and Do (2017). The frequency efficiency is proved that as benefit from two-way relaying (Do and Nguyen 2016; Do 2015; Studer et al. 2010). The fundamental idea of two-way systems is employing the assistance of a relay to exchange information between the two receivers at both directions. As report from Yue et al. (2018), the application of two-way scheme to NOMA is proposed to improve the spectral efficiency in a promising approach systems under full-duplex scenario. A trusted relay is implemented in the presence of an eavesdropper in order to exchange their NOMA signals in two-way methodology (Zheng et al. 2018). The authors presented performance analysis in terms of outage probability and ergodic rate and these expressed is provided and validated completely (Zheng et al. 2018). In other work, the authors in Wang et al. (2019) considered system where the full-duplex two-way NOMA systems is explored by introducing the closed-form expressions of various metrics such as diversity orders, outage probabilities, ergodic rates and system throughput. By assuming of imperfect self-interference cancelation, they derived these performance evaluations including both delay-limited and delay-tolerant transmission modes (Wang et al. 2019). To address further problems, the authors of Zheng et al. (2017b) and Wei et al. (2019) have studied full-duplex two-way NOMA systems. A successive group decoding scheme and a rate splitting strategy were implemented in full-duplex two-way NOMA network (Zheng et al. 2017b). A hybrid design of a two-

way relaying system (namely Hybrid-TWRS) is proposed by combining NOMA and network coding (Wei et al. 2019). Moreover, on the condition of perfect SIC (pSIC) and perfect hardware, the above existing contributions considered about full-duplex two-way NOMA systems. In practical scenarios, when NOMA requires deployment of SIC, several potential implementation issues exist such as complexity scaling and error propagation (Yue et al. 2018). More precisely, errors in decoding are related to these unfavorable factors. To study imperfect SIC in such NOMA systems, model of the residual interference signal (IS) is assumed. It is necessary to examine how imperfect SIC (ipSIC) make influences on two-way NOMA systems. However, the situation of ipSIC of NOMA is not included in both Zheng et al. (2017b) and Wei et al. (2019).

Furthermore, these considered wireless networks owing to the limitation of hardware design, in practice, the transceivers often undergo from several impairments including phase noise, I/Q imbalance and high power amplifier (HPA) non-linearity (Do 2015; Studer et al. 2010). Bjornson et al. (2013) presented the above factors and proposed a general hardware impairments model to exhibit system performance. In practical scenario, it is necessary to study the impact of hardware impairment on the NOMA network. This motivated us to consider impact of hardware impairment in two-way NOMA case.

In order to examine on degraded performance of the considered NOMA under the effects of hardware impairments, we need to evaluate system performance including the outage probability (OP) and throughput. We can summarize main contributions of this work as follows:

- We study two-way NOMA (TWR-NOMA)-based systems over Rayleigh fading channels. We then derive expressions of OP based on achievable SINR in the presence of hardware impairments.
- The closed-form expressions of OP with imperfect SIC are explored to provide a benchmark for comparison with traditional NOMA system with perfect SIC.
- Both OMA and NOMA are examined performance to show performance gap among two considered signals transmitting from the sources.
- To corroborate the derived exact and asymptotic expressions, we provide Monte Carlo simulation results for the proposed system.

The remainder of this book chapter is organized as follows: Section 6.2 presents the system model of the considered TWR-NOMA system and derives the SNR for several scenarios to detect signal. Two main metrics, including the outage and the throughput of such TWR-NOMA system experiencing hardware impairments, are derived in Sect. 6.3. Section 6.4 introduces system performance corresponding to OMA case. To examine and check exactness of each considered scenario along with insights, the detail numerical results and additional discussions are delivered in Sect. 6.5. Finally, closing clarifications are specified in Sect. 6.6.

## 6.2 System Model

This study considers two-way relay NOMA in which one relay helps data communication between pair of NOMA users, as illustration in Fig. 6.1. To facilitate NOMA transmission in two group of users: $G_1 = \{U_1, U_2\}$ and $G_2 = \{U_3, U_4\}$. The communications of the two groups need a relay which employs decode-and-forward (DF) mode, and hence two antennas are facilitated at relay. As shown in Fig. 6.1, the nearby users are denoted by $U_1$ and $U_3$ while $U_2$ and $U_4$ stand for the distant users. All the wireless channels are assumed that experience to be independent quasi-static block Rayleigh fading channels. It is assumed that the channels from user nodes to $R$ and the channels from $R$ to user nodes are reciprocal.Furthermore, $h_1$, $h_2$, $h_3$ and $h_4$ are denoted as the complex channel coefficient of $U_1 \rightarrow BS$, $U_2 \rightarrow BS$, $U_3 \rightarrow BS$ and $U_4 \rightarrow BS$ link, respectively. The channel power gains $|h_1|^2$, $|h_2|^2$, $|h_3|^2$, $|h_4|^2$ are characterized to be exponentially distributed random variables (RVs) with the parameters $\Omega_i$, $i = 1, 2, 3, 4$. The noise terms are denoted as additive white Gaussian noise (AWGN) and its mean power is denoted by $N_0$.

During the first slot, in the uplink side, two users in the left side transmit signals to relay $R$. In the same time, $R$ receives signal from the right group. Therefore, interference signals from the pair of users in the second group occur. In particular, the received signal relay $R$ is given by

$$\begin{aligned} y_{R_{X_1}} &= h_1\sqrt{\gamma_1 P}\,(x_1 + \kappa) + h_2\sqrt{\gamma_2 P}\,(x_2 + \kappa) \\ &\quad + h_3\sqrt{\theta_1\gamma_3 P}\,(x_3 + \kappa) + h_4\sqrt{\theta_2\gamma_4 P}\,(x_4 + \kappa) + \eta_{R_{X_1}}, \end{aligned} \tag{6.1}$$

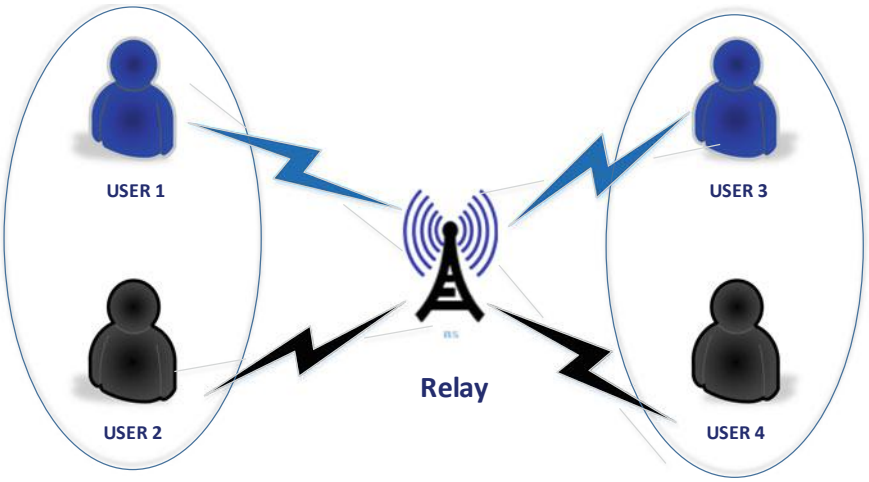



**Fig. 6.1** System model of two-way NOMA systems

where $\theta_1, \theta_2 \in [0, 1]$ represent the impact levels of self-interference at $R$. We denote $P$ as the transmit power at NOMA user. We denote $x_p, x_n$ and $x_q, x_m$ are the signals of $U_1$, $U_2$ and $U_3$, $U_4$, respectively. It is assumed that $\mathrm{E}\left\{x_p^2\right\} = \mathrm{E}\left\{x_n^2\right\} = \mathrm{E}\left\{x_q^2\right\} = \mathrm{E}\left\{x_m^2\right\} = 1$. To serve NOMA users, $\gamma_p, \gamma_n$ and $\gamma_q, \gamma_m$ stand for the corresponding power allocation factors. As required constraint, we have $\gamma_n > \gamma_p$, $\gamma_m > \gamma_q$ and $\gamma_p + \gamma_n = 1$, $\gamma_q + \gamma_m = 1$ they adapt to $\gamma_p + \gamma_n = 1$, $\gamma_q + \gamma_m = 1$. The optimal power allocation factors is beyond the scope of this paper. Here, $\eta_{R_{X_j}}$ stands for the Gaussian noise term at relay. The hardware impairment included aggregate distortion noise of the $A \to B$ link. In addition, $\kappa \triangleq \sqrt{\kappa_A^2 + \kappa_B^2}$ is the aggregate level of residual hardware impairment (RHI) of the $A \to B$ link, where $\kappa_A$ and $\kappa_B$ denote the levels of hardware impairments in node $A$ and node $B$, respectively. As such, the impairment levels of users are assumed the same. Note that these parameters, i.e., $\kappa_A, \kappa_B \geq 0$ are measured as error vector magnitudes (EVMs). Similar interference situation occurs when $R$ receives signal from the right group. In this case, interference signals are departed from the pair of users in the left group. In this case, the received signal at $R$ is given by

$$\begin{aligned} y_{R_{X_2}} &= h_q\sqrt{\gamma_q P}\left(x_q + \kappa\right) + h_m\sqrt{\gamma_m P}\left(x_m + \kappa\right) \\ &\quad + h_p\sqrt{\theta_1\gamma_p P}\left(x_p + \kappa\right) + h_n\sqrt{\theta_1\gamma_n P}\left(x_n + \kappa\right) + \eta_{R_{X_2}}. \end{aligned} \tag{6.2}$$

To simplify in computation, we denote $\rho = P/N_0$ as the transmit signal-to-noise ratio (SNR). The received signal to interference plus noise ratio (SINR) at $R$ to detect $x_p$:

$$\gamma_{x_p}^r = \frac{\gamma_p \rho \left|h_p\right|^2}{\left(\vartheta - 1\right)\gamma_p \rho \left|h_p\right|^2 + \gamma_n \rho \vartheta \left|h_n\right|^2 + \beta_1 + 1}, \tag{6.3}$$

where $(p, q) \in \{(1, 3), (3, 1)\}$, $(n, m) \in \{(2, 4), (4, 2)\}$ and $\beta_1 = \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta\left|h_m\right|^2$. Here we denote $\vartheta = 1 + \kappa^2$. The received signal to interference plus noise ratio (SINR) at $R$ to detect $x_n$

$$\gamma_{x_n}^r = \frac{\gamma_n \rho |h_n|^2}{\upsilon\rho|g|^2 + \left(\vartheta - 1\right)\gamma_n \rho |h_n|^2 + \beta_2 + 1}, \tag{6.4}$$

where $\beta_2 = \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta\left|h_m\right|^2$.

It worth noting that pSIC and ipSIC scenarios can be defined by $\upsilon = 0$: pSIC and $\upsilon = 1$: ipSIC. The SINR at $U_q$ to detect $x_n$ is expressed by

$$\gamma_{q \to x_n}^d = \frac{\lambda_n \rho \left|h_q\right|^2}{\lambda_n \rho \left(\vartheta - 1\right)\left|h_q\right|^2 + \vartheta\lambda_p\rho\left|h_q\right|^2 + \theta_2\vartheta\rho\left|h_q\right|^2 + 1}, \tag{6.5}$$

The SINR at $U_q$ to detect $x_p$ is given by

$$\gamma_{q\to x_p}^{d} = \frac{\lambda_p \rho \left|h_q\right|^2}{\upsilon\rho|g|^2 + \lambda_p \rho\left(\vartheta - 1\right)\left|h_q\right|^2 + \theta_2 \vartheta \rho \left|h_q\right|^2 + 1}. \tag{6.6}$$

We further achieve the SINR at $U_m$ to detect $x_n$ as

$$\gamma_{m\to x_n}^{d} = \frac{\lambda_n \rho |h_m|^2}{\lambda_n \rho\left(\vartheta - 1\right)|h_m|^2 + \vartheta \lambda_p \rho |h_m|^2 + \theta_2 \vartheta \rho |h_m|^2 + 1}, \tag{6.7}$$

where $\delta_1^{\text{th}} = 2^{2r_p} - 1$, $\delta_2^{\text{th}} = 2^{2r_n} - 1$ with $r_p$ is denoted as the target rate at $D_3$ to detect $x_p$, $r_n$ being the target rate at $D_3$ to detect $x_n$.

## 6.3 System Performance Analysis: Outage Performance of NOMA Users' Signals

### *6.3.1 Outage Probability of $x_p$*

To evaluate performance of such TWR-NOMA system, the metric of OP for $x_p$ with ipSIC can be given by

$$\mathrm{P}_{\mathrm{x}_p}^{\mathrm{ipSIC}} = 1 - \Pr\left(\gamma_{x_p}^{r} \geq \delta_1^{\text{th}}\right) \times \Pr\left(\gamma_{q\to x_n}^{d} \geq \delta_2^{\text{th}}, \gamma_{q\to x_p}^{d} \geq \delta_1^{\text{th}}\right), \tag{6.8}$$

where $\delta_1^{\text{th}} = 2^{2r_p} - 1$, $\delta_2^{\text{th}} = 2^{2r_n} - 1$.

In case of $\upsilon = 1$, the outage probability of $x_p$ can be further given by

$$\begin{aligned}\mathrm{P}_{\mathrm{x}_p}^{\mathrm{ipSIC}} = 1 &- \underbrace{\Pr\left(\frac{\gamma_p \rho \left|h_p\right|^2}{\left(\vartheta - 1\right)\gamma_p \rho \left|h_p\right|^2 + \gamma_n \rho \vartheta |h_n|^2 + \beta_1 + 1} > \delta_1^{\text{th}}\right)}_{\triangleq \varphi_1} \\ &\times \underbrace{\Pr\left(\begin{array}{l}\frac{\lambda_n \rho \left|h_q\right|^2}{\lambda_n \rho(\vartheta-1)\left|h_q\right|^2 + \vartheta\lambda_p\rho\left|h_q\right|^2 + \theta_2\vartheta\rho\left|h_q\right|^2 + 1} > \delta_2^{\text{th}}, \\ \frac{\lambda_p \rho \left|h_q\right|^2}{\upsilon\rho|g|^2 + \lambda_p\rho(\vartheta-1)\left|h_q\right|^2 + \theta_2\vartheta\rho\left|h_q\right|^2 + 1} > \delta_1^{\text{th}}\end{array}\right)}_{\triangleq \varphi_2}\end{aligned} \tag{6.9}$$

**Proposition 1** *The closed-form of outage probability for signal $x_p$ in ipSIC case is given by*

$$\begin{aligned}\mathrm{P}_{x_p}^{\mathrm{ipSIC}} &= 1-\varphi_1\times\varphi_2\\ &= 1-\exp\left(-\frac{\omega}{\Omega_p}\right)\prod_{i=1}^{3}\lambda_i\left(\frac{\varepsilon_1\Omega_p}{\Omega_p\lambda_1+\omega}-\frac{\varepsilon_2\Omega_p}{\Omega_p\lambda_2+\omega}+\frac{\varepsilon_3\Omega_p}{\Omega_p\lambda_3+\omega}\right)\\ &\quad\times\left(\exp\left(-\frac{\Phi_3}{\Omega_q}\right)-\frac{1}{\Omega_q}\Phi_4^{-1}\exp\left(\left(\upsilon\rho\Omega_q\frac{\delta_1^{\mathrm{th}}}{\Phi_2}\right)^{-1}\right)\exp\left(-\Phi_4\Phi_3\right)\right),\end{aligned} \tag{6.10}$$

*where* $\Phi_1 \triangleq \rho\left(\lambda_n-\delta_2^{\mathrm{th}}\left(\lambda_n\left(\vartheta-1\right)+\vartheta\lambda_p+\theta_2\vartheta\right)\right)$, $\Phi_2 \triangleq \rho\left(\lambda_p-\left(\lambda_p\left(\vartheta-1\right)-\theta_2\vartheta\right)\right)$, $\Phi_3 \triangleq \max\left(\frac{\delta_2^{\mathrm{th}}}{\Phi_1},\frac{\delta_1^{\mathrm{th}}}{\Phi_2}\right)$, $\Phi_4=\frac{1}{\upsilon\rho\Omega_q\frac{\delta_1^{\mathrm{th}}}{\Phi_2}}+\frac{1}{\Omega_q}$.

***Proof*** See in Appendix.

Based on Eq. (6.10), the OP of $x_p$ for TWR-NOMA with pSIC for the special case $\upsilon=0$ is formulated by

$$\mathrm{P}_{x_p}^{\mathrm{pSIC}}=1-\exp\left(-\frac{\omega}{\Omega_p}-\frac{\Phi_3}{\Omega_q}\right)\prod_{i=1}^{3}\lambda_i\left(\frac{\varepsilon_1\Omega_p}{\Omega_p\lambda_1+\omega}-\frac{\varepsilon_2\Omega_p}{\Omega_p\lambda_2+\omega}+\frac{\varepsilon_3\Omega_p}{\Omega_p\lambda_3+\omega}\right). \tag{6.11}$$

### 6.3.2 Outage Probability of $x_n$

The OP in this case occurs in the following cases. The case is that $R$ can first decode the information $x_p$ and then detect $x_n$. We also consider the case that either of $D_3$ and $D_4$ can detect $x_n$ successfully. The OP of $x_n$ is computed by

$$\begin{aligned}\mathrm{P}_{x_n}^{\mathrm{ipSIC}} &= 1-\Pr\left(\gamma_{x_n}^{r}\geq\delta_2^{\mathrm{th}},\gamma_{x_p}^{r}\geq\delta_1^{\mathrm{th}}\right)\\ &\quad\times\Pr\left(\gamma_{q\to x_n}^{d}\geq\delta_2^{\mathrm{th}}\right)\times\Pr\left(\gamma_{m\to x_n}^{d}\geq\delta_2^{\mathrm{th}}\right).\end{aligned} \tag{6.12}$$

Substituting (6.3), (6.4), (6.5), (6.7) into Eq. (6.12), the OP of $x_n$ is expressed as

$$\begin{aligned}
\mathrm{P}_{x_n}^{\mathrm{ipSIC}} = 1 - \underbrace{\Pr\left(\begin{array}{l}\frac{\gamma_n\rho|h_n|^2}{\upsilon\rho|g|^2+(\vartheta-1)\gamma_n\rho|h_n|^2+\beta_2+1} \geq \delta_2^{\mathrm{th}}, \\ \frac{\gamma_p\rho\left|h_p\right|^2}{(\vartheta-1)\gamma_p\rho\left|h_p\right|^2+\gamma_n\rho\vartheta|h_n|^2+\beta_1+1} \geq \delta_1^{\mathrm{th}}\end{array}\right)}_{\triangleq\Psi_1} \\
\times \underbrace{\Pr\left(\frac{\lambda_n\rho\left|h_q\right|^2}{\lambda_n\rho\left(\vartheta-1\right)\left|h_q\right|^2+\vartheta\lambda_p\rho\left|h_q\right|^2+\theta_2\vartheta\rho\left|h_q\right|^2+1} \geq \delta_2^{\mathrm{th}}\right)}_{\triangleq\Psi_2} \\
\times \underbrace{\Pr\left(\frac{\lambda_n\rho|h_m|^2}{\lambda_n\rho\left(\vartheta-1\right)|h_m|^2+\vartheta\lambda_p\rho|h_m|^2+\theta_2\vartheta\rho|h_m|^2+1} \geq \delta_2^{\mathrm{th}}\right)}_{\triangleq\Psi_3}.
\end{aligned} \tag{6.13}$$

To calculate the probability $\Psi_1$ in Eq. (6.13), let $u \triangleq \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta|h_m|^2$. After some variable substitutions and manipulations, it can be shown that

$$\begin{aligned}
\Psi_1 &= \Pr\left(\begin{array}{l}\frac{\gamma_n\rho|h_n|^2}{\upsilon\rho|g|^2+(\vartheta-1)\gamma_n\rho|h_n|^2+\beta_2+1} \geq \delta_2^{\mathrm{th}}, \\ \frac{\gamma_p\rho\left|h_p\right|^2}{(\vartheta-1)\gamma_p\rho\left|h_p\right|^2+\gamma_n\rho\vartheta|h_n|^2+\beta_1+1} \geq \delta_1^{\mathrm{th}}\end{array}\right) \\
&= \Pr\left(\begin{array}{l}|h_n|^2 \geq \frac{\delta_2^{\mathrm{th}}}{\rho\left(\gamma_n-\delta_2^{\mathrm{th}}(\vartheta-1)\gamma_n\right)} \\ \times\left(\upsilon\rho|g|^2+\theta_1\gamma_q\rho\vartheta\left|h_q\right|^2+\theta_1\gamma_m\rho\vartheta|h_m|^2+1\right), \\ \left|h_p\right|^2 \geq \frac{\delta_1^{\mathrm{th}}}{\rho\left(\gamma_p-(\vartheta-1)\gamma_p\delta_1^{\mathrm{th}}\right)} \\ \times\left(\gamma_n\rho\vartheta|h_n|^2+\theta_1\gamma_q\rho\vartheta\left|h_q\right|^2+\theta_1\gamma_m\rho\vartheta|h_m|^2+1\right)\end{array}\right) \\
&= \Pr\left(|h_n|^2 \geq \tau_1\left(\upsilon\rho|g|^2+u+1\right), \left|h_p\right|^2 \geq \tau_2\left(\gamma_n\rho\vartheta|h_n|^2+u+1\right)\right).
\end{aligned} \tag{6.14}$$

Then, it is rewritten by

$$\begin{aligned}
\Psi_1 &= \frac{1}{\varphi_n\Omega_n}\int_0^\infty\int_0^\infty \exp\left(-\frac{\tau_2\left(u+1\right)}{\Omega_p}\right) f_u\left(u\right) \\
&\quad\times\exp\left(-\tau_1\left(\upsilon\rho y+u+1\right)\varphi_n\right) f_{|g|^2}\left(y\right)\mathrm{d}y\mathrm{d}u \\
&= \frac{1}{\varphi_n\Omega_n\left(1+\upsilon\rho\tau_1\varphi_n\Omega_I\right)}\exp\left(-\frac{\tau_2}{\Omega_p}-\tau_1\varphi_n\right) \\
&\quad\times\int_0^\infty \exp\left(-\frac{\left(\tau_2+\tau_1\varphi_n\Omega_p\right)u}{\Omega_p}\right) f_u\left(u\right)\mathrm{d}u \\
&= \Delta_1\exp\left(-\frac{\tau_2}{\Omega_p}-\tau_1\varphi_n\right)\prod_{i=1}^{2}\lambda'\left(\Delta_2-\Delta_3\right),
\end{aligned} \tag{6.15}$$

where $\tau_1 = \frac{\delta_2^{\text{th}}}{\rho\left(\gamma_n - \delta_2^{\text{th}}(\vartheta-1)\gamma_n\right)}$, $u \triangleq \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta|h_m|^2$, $\tau_2 = \frac{\delta_1^{\text{th}}}{\rho\left(\gamma_p - (\vartheta-1)\gamma_p\delta_1^{\text{th}}\right)}$, $\varphi_n = \frac{\Omega_p + \gamma_n\rho\vartheta\tau_2\Omega_n}{\Omega_p\Omega_n}$, $\nabla = \frac{1}{\theta_1\gamma_q\rho\vartheta}$, $f_u(u) = \prod_{i=1}^{2}\lambda'_i\left(\nabla\left(\frac{e^{-\frac{\nabla u}{\gamma_q\Omega_q}}}{\gamma_q\Omega_q - \gamma_m\Omega_m} - \frac{e^{-\frac{\nabla u}{\gamma_m\Omega_m}}}{\gamma_m\Omega_m - \gamma_q\Omega_q}\right)\right)$, $\Delta_1 = \frac{\nabla}{\varphi_n\Omega_n(1+\upsilon\rho\tau_1\varphi_n\Omega_l)\left(\gamma_m\Omega_m - \gamma_q\Omega_q\right)}$, $\Delta_2 = \frac{\gamma_q\Omega_p\Omega_q}{\tau_2\gamma_q\Omega_q + \tau_1\varphi_n\Omega_p\gamma_q\Omega_q + \Omega_p\nabla}$, $\Delta_3 = \frac{\gamma_m\Omega_p\Omega_m}{\tau_2\gamma_m\Omega_m + \tau_1\gamma_m\varphi_n\Omega_p\Omega_m + \Omega_p\nabla}$ and $\lambda' = \frac{1}{\lambda_n\rho(\vartheta-1)}$.

We continue compute the remaining component as below

$$\begin{aligned}\Psi_2 &= \Pr\left(\frac{\lambda_n\rho\left|h_q\right|^2}{\lambda_n\rho\left(\vartheta-1\right)\left|h_q\right|^2 + \vartheta\lambda_p\rho\left|h_q\right|^2 + \theta_2\vartheta\rho\left|h_q\right|^2 + 1} \geq \delta_2^{\text{th}}\right)\\ &= \Pr\left(\left|h_q\right|^2 \geq \frac{\delta_2^{\text{th}}}{\rho\left(\lambda_n - \delta_2^{\text{th}}\left(\lambda_n\left(\vartheta-1\right) + \vartheta\lambda_p + \theta_2\vartheta\right)\right)}\right)\\ &= \exp\left(-\frac{\delta_2^{\text{th}}}{\vartheta_1\times\Omega_q}\right),\end{aligned} \tag{6.16}$$

and

$$\begin{aligned}\Psi_3 &= \Pr\left(\frac{\lambda_n\rho|h_m|^2}{\lambda_n\rho\left(\vartheta-1\right)|h_m|^2 + \vartheta\lambda_p\rho|h_m|^2 + \theta_2\vartheta\rho|h_m|^2 + 1} \geq \delta_2^{\text{th}}\right)\\ &= \Pr\left(|h_m|^2 \geq \frac{\delta_2^{\text{th}}}{\rho\left(\lambda_n - \delta_2^{\text{th}}\left(\lambda_n\left(\vartheta-1\right) + \vartheta\lambda_p + \theta_2\vartheta\right)\right)}\right)\\ &= \exp\left(-\frac{\delta_2^{\text{th}}}{\vartheta_1\times\Omega_m}\right),\end{aligned} \tag{6.17}$$

where $\vartheta_1 \triangleq \rho\left(\lambda_n - \delta_2^{\text{th}}\left(\lambda_n\left(\vartheta-1\right) + \vartheta\lambda_p + \theta_2\vartheta\right)\right)$ with $\lambda_n > \delta_2^{\text{th}}\left(\lambda_n\left(\vartheta-1\right) + \vartheta\lambda_p + \theta_2\vartheta\right)$. In case of $\upsilon = 1$, combining Eq. (6.15), (6.16) and (6.17), we have the closed-form expression for the OP of $x_n$ with ipSIC is given by

$$\begin{aligned}\mathrm{P}_{x_n}^{\text{ipSIC}} &= 1 - \Psi_1\Psi_2\Psi_3\\ &= 1 - \Delta_1\exp\left(-\frac{\tau_2}{\Omega_p} - \tau_1\varphi_n\right)\\ &\quad\times\prod_{i=1}^{2}\lambda'\left(\Delta_2 - \Delta_3\right)\exp\left(-\frac{\delta_2^{\text{th}}}{\vartheta_1\times\Omega_q} - \frac{\delta_2^{\text{th}}}{\vartheta_1\times\Omega_m}\right).\end{aligned} \tag{6.18}$$

With the specific case of pSIC, substituting $\upsilon = 0$ into Eq. (6.18), the OP of $x_n$ is calculated by

$$\begin{aligned} \mathrm{P}_{x_n}^{\mathrm{pSIC}} &= 1 - \Psi_1 \Psi_2 \Psi_3 \\ &= 1 - \Delta_1 \exp\left(-\frac{\tau_2}{\Omega_p} - \tau_1 \varphi_n\right) \\ &\quad \times \prod_{i=1}^{2} \lambda' \left(\Delta_2 - \Delta_3\right) \exp\left(-\frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_q} - \frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_m}\right). \end{aligned} \tag{6.19}$$

### 6.3.3 *Asymptotic Expressions of Outage Behavior in OMA in NOMA*

We first consider the asymptotic OPs of $x_p$. We consider specific case when $\rho \to \infty$, the asymptotic OP of $x_p$ for ipSIC/pSIC can be computed based on the fact that $\mathrm{e}^{-x} \approx 1 - x$. The asymptotic OPs are given for ipSIC, pSIC, respectively, by

$$\begin{aligned} \mathrm{P}_{x_p}^{\mathrm{ipSIC}} &= 1 - \left(1 - \frac{\omega}{\Omega_p}\right) \prod_{i=1}^{3} \lambda_i \left(\frac{\varepsilon_1 \Omega_p}{\Omega_p \lambda_1 + \omega} - \frac{\varepsilon_2 \Omega_p}{\Omega_p \lambda_2 + \omega} + \frac{\varepsilon_3 \Omega_p}{\Omega_p \lambda_3 + \omega}\right) \\ &\quad \times \left(\left(1 - \frac{\Phi_3}{\Omega_q}\right) - \frac{1}{\Omega_q} \Phi_4^{-1} \left(1 + \left(\upsilon \rho \Omega_q \frac{\delta_1^{\mathrm{th}}}{\Phi_2}\right)^{-1}\right) \left(1 - \Phi_4 \Phi_3\right)\right), \end{aligned} \tag{6.20}$$

and

$$\begin{aligned} \mathrm{P}_{x_p}^{\mathrm{pSIC}} &= 1 - \left(1 - \frac{\omega}{\Omega_p}\right) \left(1 - \frac{\Phi_3}{\Omega_q}\right) \\ &\quad \times \prod_{i=1}^{3} \lambda_i \left(\frac{\varepsilon_1 \Omega_p}{\Omega_p \lambda_1 + \omega} - \frac{\varepsilon_2 \Omega_p}{\Omega_p \lambda_2 + \omega} + \frac{\varepsilon_3 \Omega_p}{\Omega_p \lambda_3 + \omega}\right). \end{aligned} \tag{6.21}$$

Then, the asymptotic outage probabilities of $x_n$ can be further evaluated. In similar way of $x_p$ in high SNR regimes, the asymptotic OP of $x_n$ with ipSIC/pSIC is formulated by

$$\begin{aligned} \mathrm{P}_{x_n}^{\mathrm{ipSIC}} &= 1 - \Delta_1 \left(1 - \frac{\tau_2}{\Omega_p} - \tau_1 \varphi_n\right) \prod_{i=1}^{2} \lambda' \left(\Delta_2 - \Delta_3\right) \\ &\quad \times \left(1 - \frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_q} - \frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_m}\right), \end{aligned} \tag{6.22}$$

and

$$\begin{aligned} \mathrm{P}_{x_n}^{\mathrm{pSIC}} &= 1 - \Delta_1 \left(1 - \frac{\tau_2}{\Omega_p} - \tau_1 \varphi_n\right) \\ &\times \left(1 - \frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_q} - \frac{\delta_2^{\mathrm{th}}}{\vartheta_1 \times \Omega_m}\right) \prod_{i=1}^{2} \lambda' \left(\Delta_2 - \Delta_3\right). \end{aligned} \tag{6.23}$$

### 6.3.4 Throughput Analysis

The other metric is necessary to consider for such system such as throughput in delay-limited transmission scenario. In this manner, the BS transmits message to users with a fixed rate, then the threshold of throughout exists. Particularly, the corresponding throughput for considered cases is expressed as

$$\begin{aligned} \mathrm{R}_{dl}^{\psi} &= \left(1 - P_{x_1}^{\psi}\right) R_{x_1} + \left(1 - P_{x_2}^{\psi}\right) R_{x_2} \\ &+ \left(1 - P_{x_3}^{\psi}\right) R_{x_3} + \left(1 - P_{x_4}^{\psi}\right) R_{x_4}, \end{aligned} \tag{6.24}$$

where $\psi \in$ (ipSIC, pSIC).

In next section, we will the counterpart, i.e., OMA is considered as reasonable benchmark in this study.

## 6.4 Outage Performance Analysis of OMA Users' Signals

In OMA mode, we can compute the SINR at $R$ to detect $x_p$ as

$$\gamma_{x_p}^{\mathrm{OMA},r} = \frac{\rho \left|h_p\right|^2}{\left(\vartheta - 1\right) \rho \left|h_p\right|^2 + 1} \tag{6.25}$$

Similarly, SINR at $R$ to detect $x_n$ in OMA can be given as

$$\gamma_{x_n}^{\mathrm{OMA},r} = \frac{\rho |h_n|^2}{\left(\vartheta - 1\right) \rho |h_n|^2 + 1}. \tag{6.26}$$

### 6.4.1 Outage Probability of $x_p$ in OMA

The OP of $x_p$ with ipSIC for OMA-based system can be computed as

$$\mathrm{P}_{\mathrm{x}_p}^{\mathrm{OMA}} = 1 - \Pr\left(\gamma_{x_p}^{\mathrm{OMA},r} \geq \delta_1^{\mathrm{th}}\right), \tag{6.27}$$

where $\delta_{1,\mathrm{OMA}}^{\mathrm{th}} = 2^{4r_p} - 1$ and $\delta_{2,\mathrm{OMA}}^{\mathrm{th}} = 2^{4r_n} - 1$.

The OP of $x_p$ can be further given by

$$\mathrm{P}_{\mathrm{x}_p}^{\mathrm{OMA}} = 1 - \underbrace{\Pr\left(\frac{\rho\left|h_p\right|^2}{\left(\vartheta - 1\right)\rho\left|h_p\right|^2 + 1} \geq \delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right)}_{\triangleq\varphi_1}. \tag{6.28}$$

In next step, $\varphi_1$ is further given by

$$\begin{aligned}\varphi_1 &= \Pr\left(\frac{\rho\left|h_p\right|^2}{\left(\vartheta - 1\right)\rho\left|h_p\right|^2 + 1} \geq \delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right) \\ &= \exp\left(-\frac{\delta_{1,\mathrm{OMA}}^{\mathrm{th}}}{\Omega_p\left(\rho - \left(\vartheta - 1\right)\rho\delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right)}\right).\end{aligned} \tag{6.29}$$

Finally, it can be achieved he closed-form expression of OP for $x_p$ as

$$\mathrm{P}_{\mathrm{x}_p}^{\mathrm{OMA}} = 1 - \exp\left(-\frac{\delta_{1,\mathrm{OMA}}^{\mathrm{th}}}{\Omega_p\left(\rho - \left(\vartheta - 1\right)\rho\delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right)}\right). \tag{6.30}$$

### *6.4.2 Outage Probability of $x_n$*

In OMA situation, the OP of $x_n$ can be expressed as

$$\mathrm{P}_{x_n}^{\mathrm{OMA}} = 1 - \Pr\left(\gamma_{x_n}^{\mathrm{OMA},r} \geq \delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right). \tag{6.31}$$

The OP of $x_n$ is calculated by

$$\mathrm{P}_{x_n}^{\mathrm{OMA}} = 1 - \underbrace{\Pr\left(\frac{\rho|h_n|^2}{\left(\vartheta - 1\right)\rho|h_n|^2 + 1} \geq \delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right)}_{\triangleq\Psi_1}. \tag{6.32}$$

Here, it can be achieved $\Psi_1$ as

$$\begin{aligned}\Psi_1 &= \Pr\left(\frac{\rho|h_n|^2}{\left(\vartheta - 1\right)\rho|h_n|^2 + 1} \geq \delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right) \\ &= \exp\left(-\frac{\delta_{2,\mathrm{OMA}}^{\mathrm{th}}}{\rho\Omega_n\left(1 - \left(\vartheta - 1\right)\delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right)}\right).\end{aligned} \tag{6.33}$$

We have the closed-form expression for the outage probability of $x_n$ and it is given by

$$\mathrm{P}_{x_n}^{\mathrm{OMA}} = 1 - \exp\left(-\frac{\delta_{2,\mathrm{OMA}}^{\mathrm{th}}}{\rho\Omega_n\left(1-(\vartheta-1)\,\delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right)}\right). \tag{6.34}$$

### 6.4.3 Asymptotic Expressions of Outage Behavior in OMA

We first consider the asymptotic OPs of $x_p$. Based on the analytical result in (6.13) and (6.14) when $\rho \to \infty$, the asymptotic OPs of $x_p$ with $\mathrm{e}^{-x} \approx 1 - x$ are given by

$$\begin{aligned}\mathrm{P}_{x_p,\mathrm{asym}}^{\mathrm{OMA}} &= 1 - \left(1 - \frac{\delta_{1,\mathrm{OMA}}^{\mathrm{th}}}{\Omega_p\left(\rho-(\vartheta-1)\,\rho\delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right)}\right) \\ &= \frac{\delta_{1,\mathrm{OMA}}^{\mathrm{th}}}{\Omega_p\left(\rho-(\vartheta-1)\,\rho\delta_{1,\mathrm{OMA}}^{\mathrm{th}}\right)}.\end{aligned} \tag{6.35}$$

Then, the asymptotic OPs of $x_n$ can be further evaluated. By applying similar way for $x_p$, the asymptotic OP of $x_n$ in high SNR is formulated by

$$\begin{aligned}\mathrm{P}_{x_n,\mathrm{asym}}^{\mathrm{OMA}} &= 1 - \left(1 - \frac{\delta_{2,\mathrm{OMA}}^{\mathrm{th}}}{\rho\Omega_n\left(1-(\vartheta-1)\,\delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right)}\right) \\ &= \frac{\delta_{2,\mathrm{OMA}}^{\mathrm{th}}}{\rho\Omega_n\left(1-(\vartheta-1)\,\delta_{2,\mathrm{OMA}}^{\mathrm{th}}\right)},\end{aligned} \tag{6.36}$$

where $\Omega_p = \lambda_p$, $\Omega_n = \lambda_n$, $\Omega_q = \lambda_q$, $\Omega_m = \lambda_m$, $\Omega_I = \lambda_I$.

## 6.5 Numerical Results

In this section, a typical pair of users with random pairing in NOMA network is investigated. Monte Carlo simulation parameters used in this section are summarized. Monte Carlo simulations repeated $10^6$ iterations. The distance between $R$ and USER 1 or USER 3: $d_1 = d_3 = 2$ m and the distance between $R$ and USER 2 or USER 4: $d_2 = d_4 = 20m$. To exhibit performance the considered outage performance, we set targeted data rates as $r_p = 0.1$, $r_n = 0.01$, $p \in (1,\ 3)$ and $n \in (2,\ 4)$. Without loss of generality, the power allocation coefficients of $x_p$ and $x_n$ are set as $\gamma_n = 0.8$ and $\gamma_p = 0.2$. $\Omega_p$ and $\Omega_n$ are set to be $\Omega_p = d_p^{-\alpha}$ and $\Omega_n = d_n^{-\alpha}$, in which the pass-loss exponent is chosen as $\alpha = 2$. It is noted that other remaining key parameters are indicated in specific simulation result.

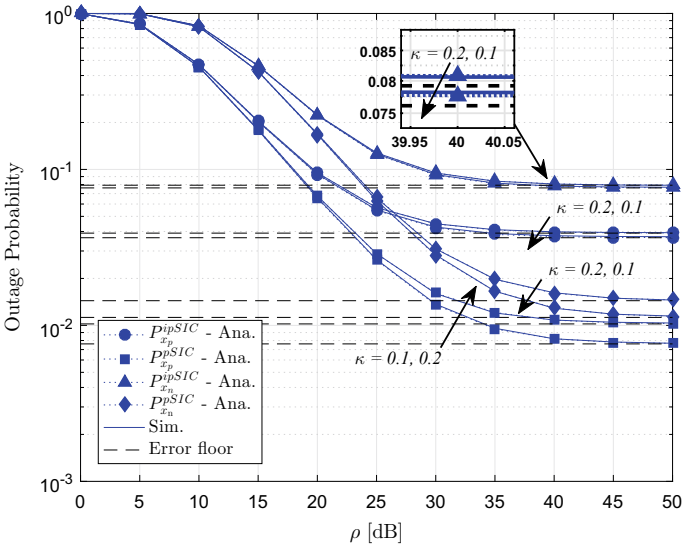



**Fig. 6.2** Outage probability versus the transmit SNR, with $\theta_1 = \theta_2 = 0.01$, $r_1 = 0.1$, $r_2 = 0.01$ and $\Omega_I = -20$ dB

Figure 6.2 plots the OP of a pair of NOMA users versus the transmit SNR. These lines are drawn to highlight impact of ipSIC/pSIC on system performance evaluation, where $\kappa = 0.1, 0.2$. In this scenario, the different values of threshold rates and power allocation factors are main factors to existence of performance gap between two NOMA users. At high SNR, as such value is greater than 35 dB, the outage event meets error floor. It can be observed that higher hardware impairment levels exhibit worse performance in whole range of transmit SNR, it is set to be from 0 to 50 dB. The exact analytical curves for the outage probability of these NOMA users match Monte Carlo results very well.

As a further advance, Fig. 6.3 plots impact of increasing power allocation factors to the OP versus SNR with different transmit SNR. More particularly, different trends in OP can be achieved in such case. It is observed that the outage behavior of each user vary as changing power allocation factor. One can observe that the asymptotic OPs are approximated to the analytical results in the whole regime of power allocation factors.

Figure 6.4 plots the throughput performance versus SNR with the different level of residual interference due to imperfect SIC. We call TWR in legend to indicate such NOMA in this study. It can be seen that the worse throughput is result from higher level of residual interference related to imperfect SIC. The throughput performance gap can be seen clearly as varying $\Omega_I$ with 3 cases $-25$ (dB), $-20$ (dB) and $-15$ (dB). It is obviously that $\Omega_I = -25$ (dB) exhibits best throughput performance.

In Fig. 6.5, the OPs are shown as a function of the transmit SNR. Reported from the impact of ipSIC, the residual residual interference signal is changed to examine

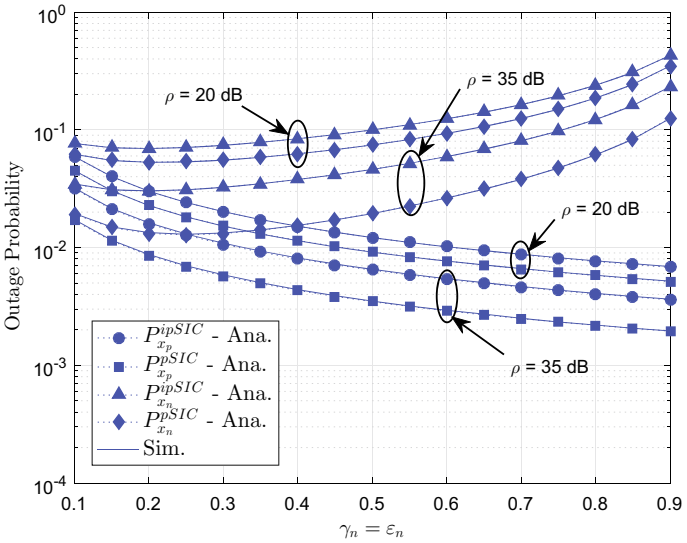



**Fig. 6.3** Outage probability comparison with respect to the power allocation coefficient $\gamma_n$ and $\varepsilon_n$, with $\theta_1 = \theta_2 = 0.01$, $\Omega_I = -20$ dB and $\kappa = 0.2$

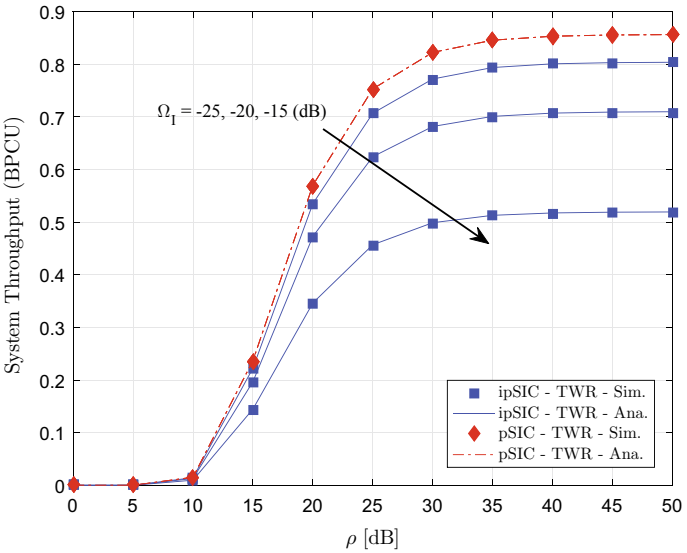



**Fig. 6.4** System throughput in delay-limited transmission mode versus the transmit SNR with different values of $\Omega_I$, with $\theta_1 = \theta_2 = 0.$ and $\kappa = 0.2$

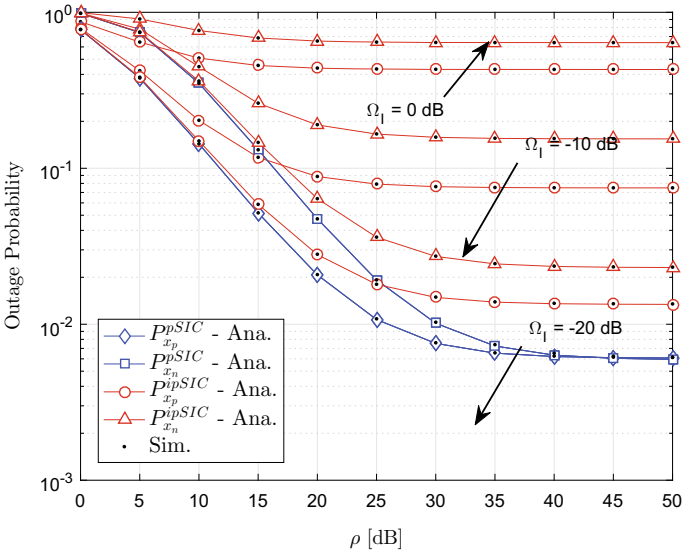



**Fig. 6.5** Outage probability versus SNR with different values of $\Omega_I$, with $\theta_1 = \theta_2 = 0.01$ and $\kappa = 0.2$

related impacts on outage event. Compared with the different level of residual interference signal, there is a decrease in outage probability for both users as increasing such level of residual interference signal. Furthermore, it is shown that the OP in pSIC scenario is lower than all cases of ipSIC scenarios at point as SNR is greater than 25 (dB). Such observation confirms impact of ipSIC in degrading performance related to outage event.

Now, location arrangement of these nodes contributes to varying outage performance as in Fig. 6.6. Interestingly, optimal location between $R$ and $D_3$ can be achieved for pSIC case of signal $x_n$ as $d_p = 2.5$. The worst outage event happens as $d_p = 10$.

Figure 6.7 confirms that asymptotic outage matches with exact outage performance of two NOMA users. It can be observed that outage performance in OMA is better than that of NOMA. The main reason is that power allocation factors for NOMA less than of OMA. However, the most important thing is that NOMA provides higher capacity of serving users in the same time with satisfaction in fairness.

## 6.6 Conclusion

This paper analyzed two-way NOMA and OMA network with hardware impairments. The imperfect SIC and perfect SIC were used in the receiver to provide comparison study. To examine system performance, we studied two main metrics such as the outage probability and throughput. Specially, we derived the outage probability in

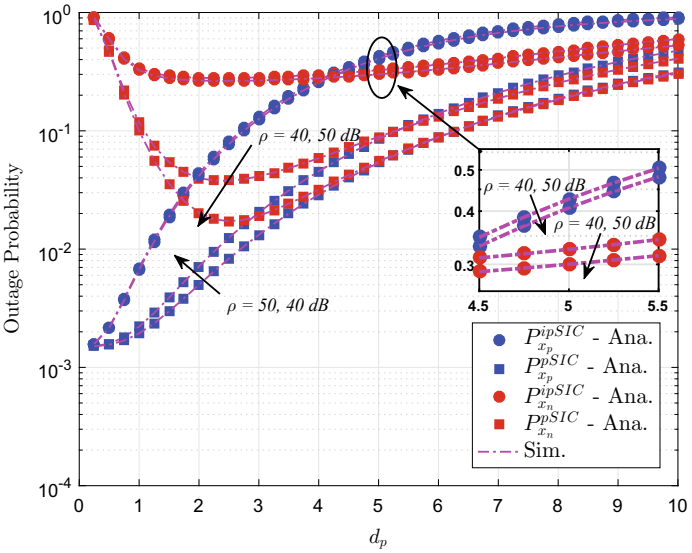



**Fig. 6.6** Outage probability comparison with respect to the distance between $R$ and $D_3$ ($d_p$), with $\theta_1 = \theta_2 = 0.01$, $\Omega_I = -20\,\text{dB}$ and $\kappa = 0.1$

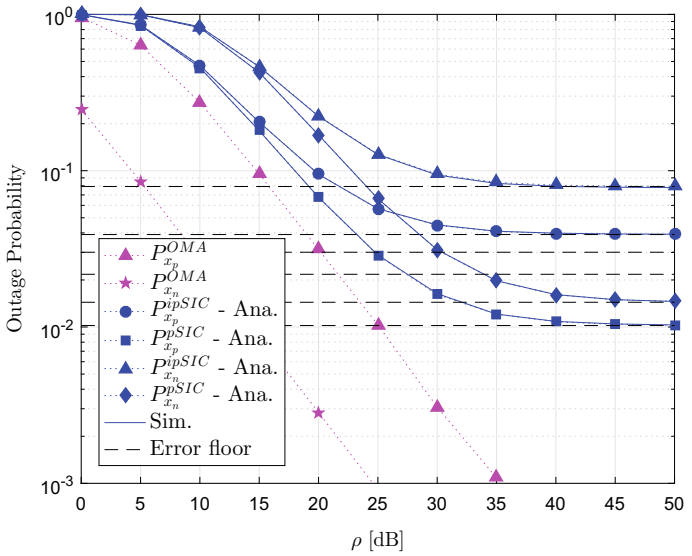



**Fig. 6.7** Comparison on OMA and NOMA scenarios, with $\theta_1 = \theta_2 = 0.01$, $r_1 = 0.1$, $r_2 = 0.01$ and $\Omega_I = -20\,\text{dB}$

the closed-form expressions. Then, the throughput in delay-limited transmissions scheme is further computed. The impact of hardware impairment on the system is evaluated in high SNR region. In particular, we derived asymptotic expressions of such outage probability. It is confirmed that the degraded outage occurred when existence of hardware impairment is raised. It can be confirmed that the worse system performance can be raised when the larger impairments level was observed. In addition, controlling hardware impairment level aims to achieve reasonable system performance.

## Appendix

### Proof of Proposition 1

***Proof*** To calculate the OP $\varphi_1$ in (6.9), let $z \triangleq \gamma_n\rho\vartheta|h_n|^2 + \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta|h_m|^2$. We denote that $z_1 = \gamma_n\rho\vartheta|h_n|^2$, $z_2 = \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2$, $z_3 = \theta_1\gamma_m\rho\vartheta|h_m|^2$ are also independent exponentially distributed random variables (RVs)with parameters $\lambda_1 = \frac{1}{\gamma_n\rho\vartheta}$, $\lambda_2 = \frac{1}{\theta_1\gamma_q\rho\vartheta}$, $\lambda_3 = \frac{1}{\theta_1\gamma_m\rho\vartheta}$, respectively.
It then further illustrates such result in following equation

$$\begin{aligned}\varphi_1 &= \Pr\left(\frac{\gamma_p\rho\left|h_p\right|^2}{(\vartheta-1)\,\gamma_p\rho\left|h_p\right|^2 + \gamma_n\rho\vartheta|h_n|^2 + \beta_1 + 1} > \delta_1^{\text{th}}\right)\\ &= \Pr\left(\begin{array}{l}\left|h_p\right|^2 \geq \frac{\delta_1^{\text{th}}}{\gamma_p\rho - (\vartheta-1)\gamma_p\rho\delta_1^{\text{th}}}\\ \times\left(\underbrace{\gamma_n\rho\vartheta|h_n|^2 + \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta|h_m|^2}_{\triangleq z} + 1\right)\end{array}\right)\\ &= \int_0^\infty \left(1 - F_{|h_p|^2}\left(\omega\left(z+1\right)\right)\right) f_z\left(z\right)\mathrm{d}z.\end{aligned} \tag{6.37}$$

Next, it can be rewritten as

$$\begin{aligned}\varphi_1 &= \int_0^\infty \exp\left(-\frac{\omega}{\Omega_p}\left(z+1\right)\right) f_z\left(z\right)\mathrm{d}z\\ &= \exp\left(-\frac{\omega}{\Omega_p}\right)\int_0^\infty \prod_{i=1}^{3}\lambda_i\left(\varepsilon_1\mathrm{e}^{-\lambda_1 z} - \varepsilon_2\mathrm{e}^{-\lambda_2 z} + \varepsilon_3\mathrm{e}^{-\lambda_3 z}\right)\mathrm{e}^{-z-1}\mathrm{d}z\\ &= \exp\left(-\frac{\omega}{\Omega_p}\right)\prod_{i=1}^{3}\lambda_i\left(\frac{\varepsilon_1\Omega_p}{\Omega_p\lambda_1+\omega} - \frac{\varepsilon_2\Omega_p}{\Omega_p\lambda_2+\omega} + \frac{\varepsilon_3\Omega_p}{\Omega_p\lambda_3+\omega}\right),\end{aligned} \tag{6.38}$$

where $\omega = \frac{\delta_1^{\text{th}}}{\gamma_p\rho - (\vartheta-1)\gamma_p\rho\delta_1^{\text{th}}}$, $z \triangleq \gamma_n\rho\vartheta|h_n|^2 + \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2 + \theta_1\gamma_m\rho\vartheta|h_m|^2$, $z_1 = \gamma_n\rho\vartheta|h_n|^2$, $z_2 = \theta_1\gamma_q\rho\vartheta\left|h_q\right|^2$, $z_3 = \theta_1\gamma_m\rho\vartheta|h_m|^2$, $\varepsilon_1 = \frac{1}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}$, $\varepsilon_2 = \frac{1}{(\lambda_3-\lambda_2)(\lambda_2-\lambda_1)}$, $\varepsilon_3 = \frac{1}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)}$, $\lambda_1 = \frac{1}{\gamma_n\rho\vartheta}$, $\lambda_2 = \frac{1}{\theta_1\gamma_q\rho\vartheta}$, $\lambda_3 = \frac{1}{\theta_1\gamma_m\rho\vartheta}$.
Similarly, $\varphi_2$ can be further calculated as follows

$$\begin{aligned}\varphi_2 &= \Pr\left(\begin{array}{l}\frac{\lambda_n\rho\left|h_q\right|^2}{\lambda_n\rho(\vartheta-1)\left|h_q\right|^2+\vartheta\lambda_p\rho\left|h_q\right|^2+\theta_2\vartheta\rho\left|h_q\right|^2+1} \geq \delta_2^{\text{th}},\\ \frac{\lambda_p\rho\left|h_q\right|^2}{\upsilon\rho|g|^2+\lambda_p\rho(\vartheta-1)\left|h_q\right|^2+\theta_2\vartheta\rho\left|h_q\right|^2+1} \geq \delta_1^{\text{th}}\end{array}\right)\\ &= \Pr\left(\left|h_q\right|^2 \geq \frac{\delta_2^{\text{th}}}{\Phi_1}, \left|h_q\right|^2 > \frac{\delta_1^{\text{th}}}{\Phi_2}, |g|^2 < \frac{\left|h_q\right|^2 - \frac{\delta_1^{\text{th}}}{\Phi_2}}{\upsilon\rho\frac{\delta_1^{\text{th}}}{\Phi_2}}\right).\end{aligned} \tag{6.39}$$

Next, it can be simplified as

$$\begin{aligned}\varphi_2 &= \Pr\left(\left|h_q\right|^2 \geq \underbrace{\max\left(\frac{\delta_2^{\text{th}}}{\Phi_1}, \frac{\delta_1^{\text{th}}}{\Phi_2}\right)}_{\triangleq\Phi_3}, |g|^2 < \frac{\left|h_q\right|^2 - \frac{\delta_1^{\text{th}}}{\Phi_2}}{\upsilon\rho\frac{\delta_1^{\text{th}}}{\Phi_2}}\right)\\ &= \Pr\left(\left|h_q\right|^2 \geq \Phi_3, |g|^2 < \frac{\left|h_q\right|^2 - \frac{\delta_1^{\text{th}}}{\Phi_2}}{\upsilon\rho\frac{\delta_1^{\text{th}}}{\Phi_2}}\right).\end{aligned} \tag{6.40}$$

Therefore, it can be solved as

$$\begin{aligned}\varphi_2 &= \int_{\Phi_3}^{\infty}\left(1 - \exp\left(-\frac{y - \frac{\delta_1^{\text{th}}}{\Phi_2}}{\upsilon\rho\Omega_q\frac{\delta_1^{\text{th}}}{\Phi_2}}\right)\right)\exp\left(-\frac{y}{\Omega_q}\right)\mathrm{d}y\\ &= \frac{1}{\Omega_q}\int_{\Phi_3}^{\infty}\left(\exp\left(-\frac{y}{\Omega_q}\right) - \exp\left(-\left(\frac{1}{\upsilon\rho\Omega_q\frac{\delta_1^{\text{th}}}{\Phi_2}} + \frac{1}{\Omega_q}\right)y + \frac{1}{\upsilon\rho\Omega_q\frac{\delta_1^{\text{th}}}{\Phi_2}}\right)\right)\mathrm{d}y\\ &= \exp\left(-\frac{\Phi_3}{\Omega_q}\right) - \frac{1}{\Omega_q}\Phi_4^{-1}\exp\left(\frac{1}{\upsilon\rho\Omega_q\frac{\delta_1^{\text{th}}}{\Phi_2}}\right)\exp\left(-\Phi_4\Phi_3\right),\end{aligned} \tag{6.41}$$

where $\Phi_1 \triangleq \rho\left(\lambda_n - \delta_2^{\text{th}}\left(\lambda_n(\vartheta-1) + \vartheta\lambda_p + \theta_2\vartheta\right)\right)$, $\Phi_2 \triangleq \rho\left(\lambda_p - \left(\lambda_p\left(\vartheta-1\right) - \theta_2\vartheta\right)\right)$, $\Phi_3 \triangleq \max\left(\frac{\delta_2^{\text{th}}}{\Phi_1}, \frac{\delta_1^{\text{th}}}{\Phi_2}\right)$, $\Phi_4 = \frac{1}{\upsilon\rho\Omega_q\frac{\delta_1^{\text{th}}}{\Phi_2}} + \frac{1}{\Omega_q}$.

Considering on $\upsilon = 1$ case, the closed-form expression for the outage probability of $x_p$ for TWR-NOMA with ipSIC is given by

$$\begin{aligned} \mathrm{P}_{x_p}^{\mathrm{ipSIC}} &= 1 - \varphi_1 \times \varphi_2 \\ &= 1 - \exp\left(-\frac{\omega}{\Omega_p}\right) \prod_{i=1}^{3} \lambda_i \left( \frac{\varepsilon_1 \Omega_p}{\Omega_p \lambda_1 + \omega} - \frac{\varepsilon_2 \Omega_p}{\Omega_p \lambda_2 + \omega} + \frac{\varepsilon_3 \Omega_p}{\Omega_p \lambda_3 + \omega} \right) \\ &\quad \times \left( \exp\left(-\frac{\Phi_3}{\Omega_q}\right) - \frac{1}{\Omega_q} \Phi_4^{-1} \exp\left( \left( \upsilon \rho \Omega_q \frac{\delta_1^{\mathrm{th}}}{\Phi_2} \right)^{-1} \right) \exp\left(-\Phi_4 \Phi_3\right) \right). \end{aligned} \tag{6.42}$$

## References


Bjornson E, Matthaiou M, Debbah M (2013) New look at dual-hop relaying: performance limits with hardware impairments. IEEE Trans Commun 61:4512–4525

Ding Z, Fan P, Poor HV (2016) Impact of user pairing on 5G non-orthogonal multiple-access downlink transmissions. IEEE Trans Veh Technol 65:6010–6023

Ding Z, Lei X, Karagiannidis GK et al (2017) A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends. IEEE J Sel Areas Commun 35:2181–2195

Do D-T (2015) Power switching protocol for two-way relaying network under hardware impairments. Radioengineering 24:765–771

Do D-T, Le C-B (2018) Application of NOMA in wireless system with wireless power transfer scheme: outage and ergodic capacity performance analysis. Sensors 18:3501

Do D-T, Nguyen H-S (2016) A tractable approach to analyze the energy-aware two-way relaying networks in presence of co-channel interference. EURASIP J Wirel Commun Netw 2016:271

Do D-T, Nguyen M-SV et al (2019) NOMA-assisted multiple access scheme for IoT deployment: relay selection model and secrecy performance improvement. Sensors 19:736

Islam SMR, Kwak KS (2016) Outage capacity and source distortion analysis for NOMA users in 5G systems. Electron Lett 52:1344–1345

Nguyen T-L, Do D-T (2017) A new look at AF two-way relaying networks: energy harvesting architecture and impact of co-channel interference. Ann Telecommun 72:669–678

Nguyen T-L, Do D-T (2018) Exploiting impacts of intercell iInterference on SWIPT-assisted non-orthogonal multiple access. Wirel Commun Mob Comput. https://doi.org/10.1155/2018/2525492

Shi S, Yang L, Zhu H (2016) Outage balancing in downlink nonorthogonal multiple access with statistical channel state information. IEEE Trans Wireless Commun 15:4718–4731

Song G, Wang X, Cheng J (2017) A low-complexity multiuser coding scheme with near-capacity performance. IEEE Trans Veh Technol 18:6775–6786

Studer C, Wenk M, Burg A (2010) MIMO transmission with residual transmit-RF impairments. In: 2010 international ITG workshop on smart antennas (WSA), pp 189–196

Tabassum H, Hossain E, Hossain J (2017) Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using poisson cluster processes. IEEE Trans Commun 65:3555–3570

Wang X, Jia M et al (2019) Exploiting full-duplex two-way relay cooperative non-orthogonal multiple access. IEEE Trans Commun 2019:2716–2729

Wei F, Zhou T, Xu T, Hu H (2019) Modeling and analysis of two-way relay networks: a joint mechanism using NOMA and network coding. IEEE J Sel Areas Commun 7:152679–152689

Yang Z, Ding Z et al (2016) On the performance of non-orthogonal multiple access systems with partial channel information. IEEE Trans Commun 64:654–667
Yang C, Xia B, Xie W et al (2018) Interference cancellation at receivers in cache-enabled wireless networks. IEEE Trans Veh Technol 67:842–846
Yue X, Liu Y et al (2018) Modeling and analysis of two-way relay non-orthogonal multiple access systems. IEEE Trans Commun 66:3784–3796
Zhang N et al (2016) Uplink non-orthogonal multiple access in 5G systems. IEEE Commun Lett 20:458–461
Zhang Y, Wang HM et al (2017a) Energy-efficient transmission design in non-orthogonal multiple access. IEEE Trans Veh Technol 66:2852–2857
Zheng B, Wang X et al (2017b) NOMA-based multi-pair two-way relay networks with rate splitting and group decoding. IEEE J Sel Areas Commun 35:2328–2341
Zheng B, Wen M et al (2018) Secure NOMA based full-duplex two-way relay networks with artificial noise against eavesdropping. In: IEEE international conference on communications (ICC), pp 1–6

# Chapter 7
# mmWave-Based 5G and Beyond Cellular Networks


**Mukesh Kumar Mishra, Aditya Trivedi, and Neelesh Mehra**

**Abstract** The exponential growth of wireless multimedia services and devices expect more spectrums to support the quality of service (QoS) of the fifth-generation (5G) and beyond cellular networks. The available large underutilized radio resources in millimeter wave (mmWave) have the potential to solve limited radio resource issues of the existing cellular bands. Recent studies suggest that densely deployed mmWave base stations with directional beamforming capability can be used for cellular systems. However, the exploitation of mmWave channels for the cellular network also introduces new challenges because of the inconsistent transmission behavior of the mmWave channel. Despite the problems raised for the mmWave signals, there are recent breakthroughs that make the mmWave viable for cellular communications, and the cellular network infrastructure should be expanded accordingly. In this regard, this chapter presents a review of issues associated with the mmWave-based cellular networks. Furthermore, key concepts and useful insights related to the mmWave-based cellular network design are also provided. Besides, a stochastic geometry-based theoretical framework is developed to account for the effect of the exploitation of mmWave for the cognitive-based multi-tier cellular networks. In this scheme, when conventional cellular channels are vacant, low power nodes operate in the traditional cellular band, and they are served by the mmWave band when channels are busy. Outage probability, area spectral efficiency, and energy efficiency are theoretically calculated for the system. Numerical results demonstrate that the proposed approach offers a significant improvement in outage performance as compared to cognitive radio-based multi-tier networks.

M. Kumar Mishra (✉) · N. Mehra
Samrat Ashok Technological Institute, Vidisha 464001, India
e-mail: mukesh8620@gmail.com

N. Mehra
e-mail: neeleshmehra@gmail.com

A. Trivedi
Atal Bihari Vajpayee-Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior 474015, India
e-mail: atrivedi@iiitm.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_7

## 7.1 Introduction

Over the last few years, with the success of smartphones, laptops, and the popularity of the various multimedia services (such as video calling and messaging, gaming, e-mail and instant messaging, social networking), the demand for high data rates has risen drastically (Hasan et al. 2011; Wang et al. 2014). The average mobile network link speed was 8.7 Megabits per second (Mbps) in 2017, and it is expected that it will meet 28.5 Mbps by 2022, i.e., threefold higher connection speed compared to 2017 (Cisco 2017). Besides, fast growth in the deployment of third-generation and fourth-generation cellular systems in the past few years increases the power consumption, capital, and operational cost of the cellular networks (Soh et al. 2013). Presently, most of the wireless standards operate at the licensed ultra-high-frequency (UHF) bands. Consequently, licensed bands are congested and expensive for network operators. On the other hand, explosively rising number of high data rate demanding multimedia devices and services expect broad spectrums to support the expected services of 5G wireless communication systems. Therefore, better connectivity, spectrum deficiency, and growing operational cost and capital cost are the critical challenges for the service providers. These challenges encourage both telecom industries and academia to search new paradigm for efficient utilization of radio spectrum, other possible frequency bands to enhance the network capacity and to rebuild the existing physical layer communication techniques in the best possible ways for the evolution of the 5G wireless networks. Recently, new techniques such as (i) deployment of small cells into existing cellular networks, (ii) opportunistic spectrum exploitation, (iii) utilization of underutilized millimeter wave (mmWave) frequency bands, and (iv) non-orthogonal multiple access enabled wireless networks are the prominent method for high data rates demanding 5G and beyond wireless networks.

## 7.2 Heterogeneous Deployment

With the consideration of practical deployment aspects and different operational characteristics of the cellular networks, theoretical modeling of the cellular network is always challenging for the research community. Besides, various channel impairments such as path loss, multipath fading, and shadowing, which considerably influence the system performance introduce additional complexity in the modeling. Furthermore, because of the increasing random placement of the low power nodes and diverse coverage range, spatial modeling of heterogeneous cellular is different from the traditional homogeneous cellular network (Andrews 2013; Chu et al. 2013). Hence, it is essential to investigate the modeling techniques that can be used to develop the theoretical framework to capture the properties of the cellular systems. Earlier, the deterministic grid model (e.g., hexagonal grid model) has been adopted to analyze and design the traditional cellular systems. This model assumes that base stations are placed at the cell center, and all base stations have homogeneous cell coverage area (Andrews 2013; Chu et al. 2013).

Furthermore, each cell exploits different sets of radio resources from adjacent cells to provide particular radio bandwidth in all cells. However, recent researches suggest that in the cellular network, including heterogeneous deployment, the coverage region of a cell depends on various factors such as capacity demand (i.e., network and link capacities) and viability of the base station deployment (Dhillon et al. 2012; ElSawy et al. 2017). Therefore, locations of the base stations concerning each other follow the random pattern, while in the hexagonal grid model fix pattern is considered for the modeling. It is clear that the traditional hexagonal model is not appropriate to capture the increasing topological randomness of the heterogeneous cellular system. In other words, low power nodes are not regularly placed and do not have similar cell coverage area. Moreover, the grid-based model gives either intractable results or inaccurate (but tractable) results because of unrealistic assumptions (ElSawy et al. 2017). In this regard, stochastic geometry gives a practical method to describe the random spatial pattern of large wireless systems. In recent years, tools from stochastic geometry have been successfully adopted to calculate the tractable and accurate performance bounds for the multi-tier wireless networks. Besides, it contributes useful insights related to network design (ElSawy et al. 2013, 2017; Andrews et al. 2011; Haenggi 2012; Haenggi and Ganti 2009). The vital benefit of the stochastic geometry-based modeling over conventional grid modeling is its potential to amalgamate different properties of the system (such as irregular placement of the base stations, channel characteristics, activities of MAC layer) into a mathematical model. Thus, the stochastic geometry, which captures topological randomness of the network using point process theory, is successfully adopted by the research community for modeling the multi-tier cellular systems (ElSawy et al. 2013, 2017; Andrews et al. 2011; Haenggi 2012; Haenggi and Ganti 2009). Additionally, it also provides more accurate and analytically tractable expressions for different performance metrics as compared to the hexagonal grid model. In the stochastic geometry-based system model, the spatial distribution of the transmitter and receiver is presented by an appropriate point process. So, the fundamental objects investigated in stochastic geometry are the point processes. A point process is a random collection of points in space (Haenggi 2012; Haenggi and Ganti 2009). The points generated from the spatial point processes denote the location of the points (base stations or mobile stations). The detailed overview of different types of point processes and their application for the cellular networks and cognitive radio networks are provided in (ElSawy et al. 2013; Haenggi and Ganti 2009; Haenggi 2012).

### *7.2.1 Stochastic Geometry-Based Heterogeneous Cellular Network Modeling*

Usually, homogeneous spatial Poisson point process (HSPPP) is used to model the cellular systems by the research community. HSPPP calculates tractable and accurate performance bounds for single-tier and multi-tier and provides useful insights related

to network design (Andrews et al. 2011; Dhillon et al. 2012). For instance, Andrews et al. (2011) calculated the coverage performance of the system over the Rayleigh fading channel. Authors assumed that the base stations are distributed using HSPPP, and all the base stations share the same radio resources. A typical mobile station is associated based on the received signal strength. Furthermore, authors compared HSPPP-based cellular networks with the square grid model and realistic cellular system. Key findings of this paper are: (a) HSPPP-based cellular network gives a tight lower bound for the coverage performance of the practical cellular network and (b) the large cellular networks can be considered as an interference-limited network, and its performance of the system is free from the density of the network. The work of Andrews et al. (2011) has been extended and analyzed for K-tier cellular systems in Dhillon et al. (2012). Authors derived the coverage and average transmission rate by considering that the mobile stations connected with the base station are offering the strongest average signal-to-interference ratio (SIR). Furthermore, they assumed that if the target SIR threshold is higher than 0 dB, only one network entity be able to meet the SIR condition. The work of Dhillon et al. (2012) has been extended in Mukherjee (2012), and they assumed that the mobile stations are connected to the tier providing the highest instantaneous signal-to-interference-plus-noise ratio (SINR). Coverage performance of multiple-input multiple-output (MIMO)-based multi-tier network has been studied in Dhillon et al. (2013) using beamforming and spatial division multiple access techniques. In ElSawy et al. (2017), authors characterized SINR distribution, bit error probability, and transmission rate for the cellular networks. Jo et al. (2012) used an association bias approach for the load balancing between the macro-base station (MBS) and small base stations (SBSs). A positive value of the bias indicates that a typical mobile station may be offloaded to SBS when the received signal power difference of the MBS and SBS falls below the specific bias value. The similar biasing concept, as used in Jo et al. (2012), is applied in Singh et al. (2013) to provide the tractable analytical model for offloading in a $K$-tier network with M radio access techniques. They also derived rate distribution and SINR distribution for analyzing the system.

In ElSawy et al. (2013), a detailed discussion is given related to modeling and performance evaluation of multi-tier cellular networks. Authors characterized SINR and associated performance measures such as outage probability and transmission rate. Besides, they also highlighted future research directions related to multi-tier cellular system modeling using stochastic geometry. In cellular network design, usually, the distance between two base stations is random (ElSawy et al. 2017). Further, in realistic cellular network deployment, two base stations of the same network operator cannot be placed arbitrarily close to each other. Therefore, in many research articles, it is argued that locations of the base stations should be characterized by the repulsive point process to maintain the minimum allowable separation between base stations (ElSawy et al. 2013, 2017; ElSawy and Hossain 2014, 2013). However, the probability generating functional (PGFL) of the repulsive point process such as Matérn hardcore point process is not available, which makes it analytically intractable and

complex (ElSawy et al. 2013; Haenggi 2012; Haenggi and Ganti 2009). Thus, HSPPP is usually adopted to model the cellular network by the research community because of its simplicity and tractability.

## 7.3 MmWave Bands for 5G and Beyond Networks

Most of the spectrums in the existing cellular bands (i.e., below 3 GHz bands) are already allocated to various dedicated wireless services. Therefore, it is becoming challenging to find free bands for new networks/services to improve the QoS of future cellular systems. Although the utilization of the existing cellular bands also depends on the scenarios. For instance, in the rural area, where base stations are sparsely deployed, most of the frequency bands are hardly exploited across time and space, whereas in the densely deployed urban scenarios, network operators demand more radio resources to provide a better QoS. Hence, obvious ways to boost the peak data rates are extending the bandwidth of the system or search for a new radio spectrum. Current measurements/trials confirm that lightly exploited mmWaves have potential for solving the limited radio resource issue of the 5G and beyond cellular networks. This band is already used by several other wireless standards such as IEEE 802.11ad for local area networking, IEEE 802.15.3c for personal area networking, and IEEE 802.16.1 for fixed-point access links (Akoum et al. 2012; Andrews et al. 2017). Recent researches based on theoretical and experimental results (Akdeniz et al. 2014; Rangan et al. 2014; Bai and Heath 2015; Singh et al. 2015; Sulyman et al. 2014) recommend that densely deployed mmWave base stations (mmWBSs) with directional beamforming capability can be used for the cellular networks. However, the exploitation of mmWave bands for the cellular network also introduces new challenges because of inconsistent transmission behavior of the mmWave signals and several limitations such as low cost, small size (especially for mobile devices), spectral, and energy efficient system (Andrews et al. 2017). The primary difference between the existing cellular networks and mmWave-based cellular network is susceptibility to blockages and the requirement of highly directional antennas at both transmitting and receiving node. Despite the challenges raised for the mmWave signals, there are recent breakthroughs that make the mmWave signals viable for cellular communications.

### *7.3.1 mmWave Signals Transmission Characteristics*

Transmission characteristics of mmWave bands are considerably diverse from below 3 GHz band because of the small wavelengths (ranging between 1 and 10 mm). Consequently, it should be carefully investigated to understand the mmWave channel characteristics. The basic idea behind the use of mmWave frequency for the cellular network was argued in 1956 by Fellers (1956). He investigated 30–300 GHz fre-

quency range and found that these frequency ranges can be used where high gain, huge bandwidth, and high directional antennas are required for the communication. However, research works based on the baseline Friis equation reported that high-frequency signals do not cover long distances. Though in Andrews et al. (2017), it is reported that the baseline Friis equation is applicable for omnidirectional radiation having an effective area of the antennas, i.e., $(wavelength\ of\ carrier\ signal)^2/4\pi$. Consequently, because of the small effective area of the antenna, small wavelength signals cannot be capable of transmitting and receiving much energy in free space (Andrews et al. 2017). Thus, directional antennas are necessary to alleviate the strong signal attenuation that occurred at mmWave signals. Moreover, an essential characteristic of mmWave-based communication system is its susceptibility to blockages. Conventional cellular bands (i.e., UHF band) also suffer from blockages, but its effects are considerably less as compared to the mmWaves. However, mmWaves experience substantially higher penetration losses when crossing by elements (such as ferroconcrete, tinted glass, and trees) due to the small wavelength and do not diffract well. In the past few years, many experiments/trials campaigns and studies have been carried out by industries and academia to estimate the transmission characteristics of mmWave bands in different scenarios such as urban, rural, indoor, and outdoor (Rangan et al. 2014; Singh et al. 2015; Bai and Heath 2015; Akdeniz et al. 2014; Rappaport et al. 2017; Hemadeh et al. 2018; Mishra and Trivedi 2019). These measurements and trials on mmWaves reveal some useful insights related to the mmWave channel characteristics for cellular networks. Some key points are listed as

- Transmission properties of the line of sight (LOS) and non-line concerning sight (NLOS) links are quite different.
- As compared to existing cellular band mmWaves experience higher attenuation and sparse multipath scattering.
- Wall penetration losses are also very large for mmWave bands as compared to the traditional cellular bands.
- Dense deployment of base stations with the directional antenna is needed to mitigate the substantial signal attenuation that occurred at mmWave signals.

#### 7.3.1.1 Stochastic Geometry-Based MmWave Cellular Network Modeling

Key observations of mmWave channel characteristics reported in Sect.7.3.1 indicate that stochastic geometry modeling used for the traditional cellular networks (discussed in Sect. 2.1) cannot be directly adopted for the mmWave-based cellular networks (Bai and Heath 2015; Andrews et al. 2017; Mishra and Trivedi 2019). Therefore, recent researches consider the propagation characteristics of the mmWave to model the system for realistic analysis using the tools from stochastic geometry. For instance, in Akoum et al. (2012), stochastic geometry-based modeling has been considered to investigate the SINR and rate distribution of the mmWave cellular network. However, they did not consider the blockage effect to examine the performance of

the network. In Bai et al. (2014), authors proposed HSPPP-based blockage model for the cellular network. Furthermore, they compared the performance of the mmWave network with the traditional cellular networks and found that it provides better performance. A similar blockage model has been adopted to analyze the cellular network (Bai and Heath 2015) and ad hoc network (Thornburg et al. 2015). Bai et al. Bai and Heath (2015) derived coverage probability and rate distribution by assuming that mmWave base stations are distributed by HSPPP on the two-dimensional plane. They simplified the exact random blockage model by approximating the random LOS region by considering fixed-sized ball with radius $R$. Also, using this simplification, they derived more tractable coverage probability expressions for mmWave cellular network. In Rappaport et al. (2013), it has been noticed that even with the low SINR condition, mmWave may provide comparable performance as compared to existing cellular networks due to the large underutilized bandwidth. In Mishra and Trivedi (2019), authors provided a generalized analytical framework for the mixed mmWave and UHF-based cellular network. Additionally, they included the impact of the various parameters (such as base station density, link distance, multipath fading, blockages, path loss, SINR threshold, and preference factor) on the performance of the network. Furthermore, they calculated the spectral efficiency and deployment cost efficiency of the network. The detailed discussion of mmWave-based cellular network and theoretical modeling using stochastic geometry is provided in Andrews et al. (2017); Rangan et al. (2014); Hemadeh et al. (2018); Rappaport et al. (2017).

## 7.4 System Design Considerations and Implications

In the previous sections, we have discussed that mmWave bands can provide better quality of services for 5G and beyond cellular networks. However, unplanned deployment of low power nodes with diverse operating characteristics, widely varying cell coverage, different types of backhaul connections, diverse operating frequency bands, environmental impairments, and transmission protocols make mmWave cellular network modeling and analysis challenging. Hence, major issues with mmWave-based communication system modeling are to build/develop frameworks that capture the characteristics of the mmWave bands and capable of carrying out the requirements of 5G and beyond networks. Also, sensitivity to blockages, higher wall penetration losses, poor diffraction characteristics, etc. are the issues related to mmWave signals. Therefore, the fundamental difference between a traditional cellular network and mmWave-based network is susceptibility to blockages and the requirement of highly directional antennas at transceivers. So, mmWave-based cellular system model should take into account these different aspects for realistic analysis. In this section, we briefly discussed some essential elements related to mmWave system modeling.

- **Identify the system scenario and operating frequency**: Transmission characteristics of mmWaves differ from one scenario to others, such as indoor, outdoor, or combination of both indoor-outdoor. Furthermore, its transmission characteristics also depend on the different operating frequency bands available in mmWaves, e.g., 28, 38, 60, 72 GHz, etc. Therefore, according to the chosen operating frequency and scenario channel model should be developed.
- **Blockages**: Blockages in the transmission medium affect wireless channels due to different phenomena such as reflection, refraction, scattering, diffraction, and absorption. Since, these effects are based on the environment/conditions. Therefore, received signal power from the transmitting node is characterized statistically as a function of the distance between transmitting and receiving node. Usually, log-normal distribution is considered in traditional cellular networks to model the shadowing effect. In a conventional homogeneous cellular system, blockages considered a secondary effect. This is because of the assumptions that most of the time, received signals are NLOS due to the long communication links (Andrews et al. 2017; 3GPP 2010). Moreover, omnidirectional antennas are considered for the analysis, which already includes the effect of blockages and other effects such as diffractions and ground reflections (Andrews et al. 2017). On the other hand, recent experiments on the outdoor mmWave system confirm that stationary obstructions such as buildings and trees cause a notable variation in the path loss law between LOS and NLOS links. Therefore, different path loss exponents are considered to model the LOS and NLOS link for mmWaves (Rappaport et al. 2014; 5G channel model for bands up to 100 GHz, Aalto Univ. 2015). Due to the reduced diffraction of mmWaves, a large portion of the signal energy is scattered; consequently, path loss exponent is much higher for the NLOS mmWave links (Rappaport et al. 2014; Zhang et al. 2015). However, blockages also affect the interference signals, so most of the interfering signals are likely to be blocked in the dense environment. Various blockage models adopted by the literature to model the mmWave-based wireless system are random shape theory model, 3GPP blockage model, LOS ball model, Poisson line model, etc. (Bai et al. 2014; Bai and Heath 2015; Thornburg et al. 2016; Andrews et al. 2017).
- **Multipath Fading** Various channel models are available in the literature to describe the statistics of the amplitude and phase of multipath fading signals. In this context, the Nakagami-$m$ fading model is an empirical model covering a wide range of multipath fading channels by varying the value of $m$ (fading parameter). Nakagami-$m$ is a generic fading channel that can model different fading environments. Therefore, this fading model is well accepted in literature for characterizing the small-scale fading for different frequency bands, including mmWaves. The behavior of the mmWave network depends on the various system parameters such as blockage model, path loss exponents, density, antenna gains and beamwidths, and bandwidth. For example, with large bandwidth and high directionality, the impact of the interference is less, whereas the influence of the interference increases with wider beamwidth and a smaller bandwidth (Andrews et al. 2017). These scenarios can be easily considered by varying the value of the $m$ in Nakagami-$m$ fading. On the other hand, going with a Rayleigh fading chan-

nel alone does not provide these wide ranges of analyses. Therefore, generally, in literature, firstly, authors carry out the analysis with the Nakagami-$m$ fading channel. After that, by substituting different values of $m$, they can calculate the performance of the network for various fading channels. In Gupta et al. (2016), authors found that taking a generalized fading model (e.g., Nakagami-$m$ distribution, Rician distribution, etc.) to characterize the small-scale fading does not give any extra design insights for the mmWave networks. However, it makes analysis considerably complicated. Therefore, they considered only Rayleigh fading for the performance analysis of the mmWave network.

- **mmWave Signals are Interference Limited or Noise Limited**: One of the major issues associated with the cellular network is interference management from the nearby cells utilizing the same time-frequency resources. In interference-limited network, rising transmission power does not improve the SINR. Another way to understand the behavior of the interference-limited network is that after a specific density, more increase in base station density does not considerably increase the performance of the system. However, SINR is balanced by the increased interference due to the increase in density. In the context of the mmWave band, features like beamforming and sensitivity to blockages alter the statistics of the interference encountered at the receiver. Various claims are drawn in the earlier experiments/trials about the sensitivity of mmWave band to interference. Some research claims that mmWave bands are noise limited, whereas some reports have argued that mmWave systems are interference-limited (Bai and Heath 2015; Thornburg et al. 2016; Mishra and Trivedi 2019). According to Andrews et al. (2017), the behavior of the mmWave network also depends on the critical network parameters, e.g., base station density, number of users, antenna gains, beamwidths, bandwidth of the system, blockage model, path loss exponents, etc. They considered 73 GHz and 28 GHz, two carrier frequency bands for the analysis. They found that the 73 GHz system favors being noise-limited, whereas a more considerable impact from interference is reported in the 28 GHz band. Numerically, they proved that the influence of interference rises with a smaller bandwidth and broad beamwidth in the beamforming. After studying the available literature carefully, it is hard to give a precise conclusion of whether mmWave cellular systems are noise-limited or interference-limited, although many works reported that mmWave-based networks could be LOS interference-limited with high density. In Bai and Heath (2015), Thornburg et al. (2016), Andrews et al. (2017), Mishra and Trivedi (2019), authors reported that increasing density of mmWave base station decreases the distance of the serving base station and increases the probability that serving mmWave base stations being LOS base station resulting in improved performance of the networks. However, after a certain density, there will be enough LOS interference for the typical outdoor user; as a result, outage performance decreases.

## 7.5 Analysis of Cognitive-Based Hybrid mmWave/UHF Cellular Network

Even with the considerably underutilized bandwidth available in mmWave band, it cannot be used solely for the mobile communication due to several limitations (such as sensitivity to blockages, higher path loss, and reduced wall penetration). Hybrid UHF and mmWave-based network may be one way for better exploitation of this band. Furthermore, UHF and mmWave are two distinct bands; consequently, cross-tier interference of the system can be effectively alleviated. So, in this section, using these facts of mmWave, and the capability of cognitive radio to identify the unoccupied bands, a spectrum access scheme is proposed for opportunistic spectrum access. This spectrum sharing scheme opportunistically exploits the UHF and mmWave band. In this scheme, when UHF channels are vacant, cognitive pico base station (CPBS) operates in the vacant UHF channels, and it serves by mmWave band when UHF channels are busy. In this work, outage probability, area spectral efficiency, and energy efficiency are theoretically calculated for the cognitive enabled hybrid cellular network (CHCN), and some useful observations are also drawn for CHCN.

### *7.5.1 System Model*

We consider a three-tier CHCN which comprises of MBSs, CPBSs, and mmWBSs. MBSs are considered as a licensed user network (LUN), and CPBSs constitute an unlicensed user network (UUN). MBS tier operates in the conventional cellular band (below 3 GHz band, a.k.a. UHF band), whereas the CPBS tier opportunistically exploits the UHF band used by the MBS network. mmWBS tier works on the mmWave band (e.g., 28 GHz or 38 GHz). It is utilized as a backup channel to provide coverage to CPBS users when free channels are not available for CPBSs during spectrum sensing. The MBSs are assumed to be distributed uniformly in the 2D plane as an HSPPP $\Upsilon_M$ with density $\lambda_M$. Similarly, CPBSs are modeled by HSPPP $\Upsilon_p$ with density $\lambda_p$ and mmWBSs follow another HSPPP $\Upsilon_{mw}$ with density $\lambda_{mw}$. All the point processes, i.e., $\Upsilon_M$, $\Upsilon_p$, and $\Upsilon_{mw}$ are independent of each other. The distance between typical mobile station to its serving base station,i.e., MBS or CPBS or mmWBS, is presented by $r_k$ where $k \in M, p, mw$. We have also assumed that that using cognitive radio capability, CPBS actively acquires information about access to the licensed resource blocks. In this work, the MBS and CPBS are analogous to LU and UU transmitters. Similarly, macro-mobile station (MMS) and pico mobile station (PMS) are equivalent to LU and UU receivers. Effect of path loss attenuation is considered by the standard distance-dependent path loss model, i.e., $r^{-\nu}$, where $\nu$ denotes path loss exponent. Nakagami-$m$ distribution with unit mean is considered to characterized the small-scale fading. It is assumed that a typical mobile station is outdoor. The mathematical modeling related to blockage and directional beamforming for mmWave and other assumptions associated with mmWave signal transmission is similar to Thornburg et al. (2016), Mishra and Trivedi (2019).

### 7.5.2 CPBS Network Activity Model

In cognitive radio network, it is assumed that UUs are permitted to transmit on the condition that the performances of the LUs are not affected by the interference created by the UUs. In other words, UUs may perform data transmission when LUs are inactive. In this work, we have divided CPBS/UU transmission period $T$ into two parts, i.e., sensing time and data transmission time with durations of $T_s$ and $T_d$, respectively. Each CPBS regularly detects MBS channels to know the free channels. During the sensing period, $T_s$, CPBS cannot transmit data. It is assumed that all the CPBSs are synchronized and having the same sensing time $T_s$ (Cheng et al. 2012; Panahi and Ohtsuki 2014). After sensing of the licensed spectrum within the sensing period $T_s$, identified vacant channels are allocated to CPBS for the data transmission in time $T_d$. We have assumed that the CPBS network uses a slotted ALOHA medium access protocol for scheduling the data transmission (Cheng et al. 2012; Zaidi et al. 2010). Hence, CPBSs perform data transmission with a probability of $p_a$. Furthermore, the probability of choosing the same channel from the pool of the free/unoccupied channel is $p_c$. It is assumed that CPBS knows the transmission parameters to set up the primary protection region (PPR) around MMS. The PPR is a circular disk or region around the typical MMS with a radius $R$ to reduce the consequence of the harmful interference for typical MMS generated by the CPBS tier. Hence, active CPBSs which are outside the PPR cause interference at the typical MMS. The activation of the CPBS depends on the PPR of the MMS and activities of the MMS. Due to the PPR, the CPBSs' activity depends on the MBS networks, which make accurate aggregate interference calculation difficult. The UUs/CPBSs inside the PPR of MMS cannot transmit during the communication session of the MMS.

### 7.5.3 Power Consumption Model

A power consumption model with embodied power is suggested to evaluate the energy efficiency of the system. In most of the literature, only operating power is considered for the analysis of the network power consumption while embodied power is neglected. The embodied power represents a considerable part of the total power consumption of the system, and it is 30–40 % for the MBSs and around 20% for the small cell network (Ge et al. 2014). The operating energy of the MBS, CPBS, and mmWBS may be separated into two parts, i.e., transmission power and constant part, and is defined as (Soh et al. 2013).

$$P^M = P_M \tau_M T_d + P_{SP}^M T_d \tag{7.1}$$

$$P^p = P_p \mathcal{P}_{OL} \tau_p(T_d) + P_{SP}^p(T_d + T_s) + P_{SS}^p(T_s) \tag{7.2}$$

$$P^{mw} = P_{mw} \tau_{mw} T_d + P_{SP}^{mw} T_d \tag{7.3}$$

where $P_M$, $P_p$, and $P_{mw}$ are the transmitted power of MBS, CPBS, and mmWBS. $P_{SP}^{M}$ denotes static power of the MBS. Similarly, $P_{SP}^{p}$ and $P_{SP}^{mw}$ describe the static power of CPBS and mmWBS, respectively. Load-dependent power consumption factor of respective base stations is denoted by $\tau_M$, $\tau_p$, and $\tau_{mw}$. $P_{SS}^{p}$ indicates the sensing power consumption of the CPBS. Further, $T_t = T_s + T_d$, where $T_t$ is the time interval of a frame (it is assumed that CPBS operates on frame by frame basis) which is the sum of $T_s$ sensing time and data transmission time $T_d$ (Mishra and Trivedi 2015).

Total power consumption model of MBS, CPBS, and mmWBS includes operating power and embodied power of the network. Embodied power consumption is the combination of initial embodied power $P_{in}$ and maintenance embodied power consumption $P_{mt}$. $P_{in}$ comprises power consumption in the initial making of the equipments, its assembling and installation. $P_{mt}$ includes power consumption of the maintenance of the base station. Thus, total power consumption for the MBS, CPBS, and mmWBS is expressed as

$$P_{\text{total}}^{M} = P^{M} + P_{E}^{M} = P^{M} + P_{in}^{M} + P_{mt}^{M} \tag{7.4}$$

$$P_{\text{total}}^{p} = P^{p} + P_{E}^{p} = P^{p} + P_{in}^{p} + P_{mt}^{p} \tag{7.5}$$

$$P_{\text{total}}^{mw} = P^{mw} + P_{E}^{mw} = P^{mw} + P_{in}^{mw} + P_{mt}^{mw} \tag{7.6}$$

where $P^{M}$, $P^{p}$, and $P^{mw}$ are given by the (7.1), (7.2), and (7.3), respectively. $P_{E}^{M}$, $P_{E}^{p}$, and $P_{E}^{mw}$ are the embodied power of MBS, CPBS, and mmWBS, respectively.

#### 7.5.3.1 Outage Analysis for MBS Tier

Aggregate interference at a typical MMS comprised of two parts, i.e., the co-tier interference created by other MBS, i.e., $\mathcal{I}_{M}^{m}$ and the cross-tier interference generated by CPBS, $\mathcal{I}_{M}^{p}$. However, the exact evaluation of the aggregate interference of the network is difficult due to the dependency between MBS and CPBS networks. It should also be noted that CPBS transmission depends on the locations and activities of the MMSs; therefore, CPBSs do not follow HSPPP (Lee and Haenggi 2012). The HSPPP of the MBSs/LUs and Poisson hole process (PHP) of the CPBSs/UUs are not independent. At this stage, it is essential to note that all the CPBS will not be considered as a potential interferer to calculate the aggregate interference at the typical MMS. Thus, before calculating the total interference at the typical MMS, it is required to calculate the density of the potential interferer. Assume that typical MMS is located at the origin to evaluate the aggregate interference of the network. For typical MMS, all the CPBS outside the PPR are considered as potential interferer for the typical MMS. Furthermore, it is assumed that CPBS network uses a slotted ALOHA medium access protocol for scheduling the data transmission. Hence, CPBS perform data transmission with probability $p_a$. Additionally, the probability of choosing the same channel from the pool of the vacant channel is given by $p_c$. To accommodate these two conditions $\Upsilon_p$ is thinned independently (Zaidi et al.

2010). Consequently, the density of the potential interferer in CPBS is expressed by $\lambda_p{}' = \lambda_p p_a p_c$. Using these facts, the outage of the MBS network over Nakagami-$m$ is expressed as

$$\mathcal{P}_o^M = \sum_{t=0}^{m} (-1)^t \binom{m}{t} e^{-s\mathcal{N}_M} \mathcal{L}_{I_M^m}(s) \mathcal{L}_{I_M^p}(s), \tag{7.7}$$

where $s = \frac{bt\theta_M r_M^{\nu}}{P_M}$ and $b = m(m!)^{-1/m}$. $m$ and $\nu$ represent the fading parameter and is the path loss exponent, respectively. $\theta_M$ and $\mathcal{N}_M$ are SINR threshold and noise power, respectively. $r_M$ indicates the distance between MBS to mobile station, respectively. $\Gamma(.)$ indicates Gamma function. Laplace transform $\mathcal{L}_{I_M^m}(s)$ and $\mathcal{L}_{I_M^p}(s)$ are expressed by

$$\mathcal{L}_{I_M^m}(s) = e^{\frac{-2\pi\lambda_M}{\nu} \sum_{k=1}^{m} \binom{m}{k} \frac{(P_M sT)^{\frac{2}{\nu}} \Gamma(m-k+\frac{2}{\nu})\Gamma(k-\frac{2}{\nu})}{\Gamma(m)}}, \tag{7.8}$$

$$\mathcal{L}_{I_M^p}(s) = e^{-2\pi\lambda_p{}' \int_R^{\infty} \left[ \frac{\sum_{k=1}^{m} \binom{m}{k} \left(sTP_p y^{-\nu}\right)^k}{\left(1+sTP_p y^{-\nu}\right)^m} \right] \mathrm{y}\mathrm{d}\mathrm{y}} \tag{7.9}$$

where $R$ is the radius of the PPR and $T = 1/m$. *Proof:* Similar to the proof of Eq. (7.3) of Swami et al. (2017).

### 7.5.4 Outage Analysis for CPBS Tier

In traditional CPBS networks, CPBS opportunistically exploits the unoccupied spectrum band of the LUN with probability $\mathcal{P}_{OL}$, and in this condition, SINR outage probability is expressed by $\mathcal{P}_{ol}^p$. Besides, when channels are occupied with the probability $1 - \mathcal{P}_{OL}$, then outage probability $\mathcal{P}_{ol}^p$ is denoted by $\mathcal{P}_{ol}^{busy}$ and becomes 1 due to the channel unavailability for the data transmission (Mishra et al. 2018). Consequently, outage probability of CPBS network $\mathcal{P}_{OL}^p$ is described as $\mathcal{P}_{OL}^p = \mathcal{P}_{OL}\mathcal{P}_{ol}^p + (1 - \mathcal{P}_{OL})\mathcal{P}_{ol}^{busy}$. In traditional opportunistic spectrum access, when MBS network is in a busy state, CPBS cannot communicate due to channel unavailability. Therefore, outage performance of the CPBS is degraded. In this work, this problem is solved by using another underutilized band, i.e., mmWave. It also improves the performance of the system. Thus, when the MBS tier is in occupied state CPBS serves by mmWave bands (e.g., 28 GHz or 38 GHz) with probability $1 - \mathcal{P}_{OL}$. As conventional cellular band and mmWave band are two distinct frequency bands, as a result, interference generated by CPBS may be adequately reduced for MBS network when it operates in mmWave band. Thus, the outage probability of CPBS tier is represented by

$$\mathcal{P}_{OM}^p = \mathcal{P}_{OL}\mathcal{P}_{ol}^p + (1 - \mathcal{P}_{OL})\mathcal{P}_o^{mw}, \tag{7.10}$$

Here $\mathcal{P}_{OL}$ is denoted by Mishra et al. (2018) and $\mathcal{P}_{o}^{mw}$ is expressed by Mishra and Trivedi (2019). $\mathcal{P}_{ol}^{p}$ is denoted by

$$\mathcal{P}_{ol}^{p} = \sum_{t=0}^{m} (-1)^{t} \binom{m}{t} e^{-s\mathcal{N}_p} \mathcal{L}_{I_P^m}(s) \mathcal{L}_{I_P^p}(s), \tag{7.11}$$

where $s = \frac{bt\theta_p r_p^{\nu}}{P_p}$. $\nu$ and $\mathcal{N}_p$ indicate the path loss exponent and noise power, respectively. $\theta_p$ and $r_p$ represent SINR threshold and the distance between CPBS to the mobile station, respectively. Laplace transform $\mathcal{L}_{I_P^m}(s)$ and $\mathcal{L}_{I_P^p}(s)$ are given by

$$\mathcal{L}_{I_P^m}(s) = e^{\frac{-2\pi\lambda_M}{\nu} \sum_{k=1}^{m} \binom{m}{k} \frac{(sTP_M)^{\frac{2}{\nu}} \Gamma(m-k+\frac{2}{\nu})\Gamma(k-\frac{2}{\nu})}{\Gamma(m)}}, \tag{7.12}$$

$$\mathcal{L}_{I_P^p}(s) = e^{\frac{-2\pi\lambda p'}{\nu} \sum_{k=1}^{\binom{m}{k}} \frac{(sTP_p)^{\frac{2}{\nu}} \Gamma(m-k+\frac{2}{\nu})\Gamma(k-\frac{2}{\nu})}{\Gamma(m)}}, \tag{7.13}$$

*Proof:* Proof of $\mathcal{P}_{ol}^{p}$ is similar to Eq. (7.7).

### 7.5.5 Total Outage Probability Analysis for CHCN

In conventional cognitive radio, LUN has a higher priority to access the spectrum, whereas in a multi-tier cellular system with an open access policy, a mobile station can be associated with the base station based on the tier association policy. In other words, there is no concept of higher priority, i.e., all the mobile stations are LUs, either it is associated with the MBS or CPBSs. In this work also, similar to Mishra and Trivedi (2019), weighted cost functions are used to determine the tier association probabilities. Therefore, tier selection probability of MBS tier $\mathcal{T}_M$ can be calculated as $\frac{\mathbb{T}_M}{\mathbb{T}_M+\mathbb{T}_p}$. Likewise, tier selection probability of CPBS tier $\mathcal{T}_p$ may be given as $\frac{\mathbb{T}_p}{\mathbb{T}_M+\mathbb{T}_p}$. Using the tier association probabilities, the total outage probability of CHCN is given by

$$\mathcal{P}_{OM}^{total} = \mathcal{T}_M \mathcal{P}_o^M + \mathcal{T}_p \mathcal{P}_{OM}^{p}, \tag{7.14}$$

where $\mathcal{T}_M$ and $\mathcal{P}_o^M$ indicate the tier association probability and tier outage probabilities for MBS tier, respectively. $\mathcal{T}_p$ and $\mathcal{P}_{OM}^{p}$ are the tier association probability and tier outage probabilities for CPBS tier, respectively.

### 7.5.6 Area Spectral Efficiency and Energy Efficiency Analysis

Using the definition of area spectral efficiency (Mishra et al. 2018; Mishra and Trivedi 2019) the area spectral efficiency for the MBS tier is expressed as

$$ASE_M = \lambda_M[1 - \mathcal{P}_o^M][\log_2(1 + \theta_M)] \tag{7.15}$$

Similarly, area spectral efficiency for CPBS tier when it operates in traditional UHF band and mmWave band are given by

$$ASE_p = \lambda_p[1 - \mathcal{P}_{ol}^p][\log_2(1 + \theta_p)] \tag{7.16}$$

$$ASE_{mw} = \lambda_{mw}[1 - \mathcal{P}_o^{mw}][\log_2(1 + \theta_{mw})] \tag{7.17}$$

where $\theta_M$ is SINR threshold for the MBS tier. $\theta_p$ and $\theta_{mw}$ are SINR thresholds for CPBS tier when it operates in traditional UHF band and mmWave band, respectively. $\mathcal{P}_o^M$ is outage probability of MBS tier. $\mathcal{P}_{ol}^p$ and $\mathcal{P}_o^{mw}$ denote SINR outage probability of CPBS tier when it operates in traditional UHF band and mmWave band, respectively.

Energy efficiency is given for CHCN as (Mishra et al. 2018)

$$EE = \frac{ASE_{\text{CHCN}}}{\lambda_M P_{\text{total}}^M + \lambda_p P_{\text{total}}^p + \lambda_{mw} P_{\text{total}}^{mw}}, \tag{7.18}$$

where $ASE_{CHCN}$ can be calculated by using the (7.15), (7.16), and (7.17). $P_{\text{total}}^M$, $P_{\text{total}}^p$, and $P_{\text{total}}^{mw}$ are given by (7.4), (7.5), and (7.6), respectively.

### 7.5.7 Results and Discussions

In this section, different numerical results are illustrated based on the previous analyses. It is assumed that mmWave link works in 28 GHz with 500 MHz bandwidth, and the traditional cellular link operates in 900 MHz with 20 MHz bandwidth. Similar assumptions have been considered in (Mishra and Trivedi 2019) also. The network parameters that are used to calculate the numerical results are given in Table 7.1.

Figure 7.1 illustrates the outage performance of CPBS tier with respect to the SIR threshold for different $\mathcal{P}_{idle}$. In this figure, we have plotted two types of cognitive-based networks, i.e., CHCN (when CPBS opportunistically exploits the MBS channels and underutilized mmWave channels) and traditional cognitive-based cellular network (CCN, in which CPBS opportunistically utilizes only the MBS channels for the communication). As expected, the outage performance of the network deteriorates with the increase in the SIR threshold of the CPBS tier. It may be seen that outage performance of the CPBS tier for CCN provides the worst performance as

**Table 7.1** Network Parameters

| Parameters | Values |
|---|---|
| $\lambda_M$, $\lambda_p$, $\lambda_{mw}$ | $10^{-6}m^{-2}$, $5 \times 10^{-5}m^{-2}$, $5 \times 10^{-5}m^{-2}$ |
| $\lambda_b$ | 0.001 $m^{-2}$ (Biswas et al. 2016) |
| $\theta_{mw}$, $\theta_M$, $\theta_p$ | 0 dB |
| $P_M$, $P_p$, $P_{mw}$ | 20 W (Soh et al. 2013), 1 W, 1 W (Tao et al. 2013) |
| $\tau_M$, $\tau_p$ | 4.7, 8.0 (Soh et al. 2013) |
| $\nu$ | 3.5 |
| $\nu_l$, $\nu_n$ | 2, 4 (Thornburg et al. 2015) |
| $P_{SP}^M$, $P_{SP}^p$, $P_{SP}^{mw}$ | 130.0 W, 4.8 W, 4.8 W (Soh et al. 2013) |
| $P_{SS}^p$ | 0.1 W |
| $r_M$, $r_p$, $r_{mw}$ | 300 m, 50 m, 50 m |
| $T_s$, $T_d$ | 0.2, 0.8 |

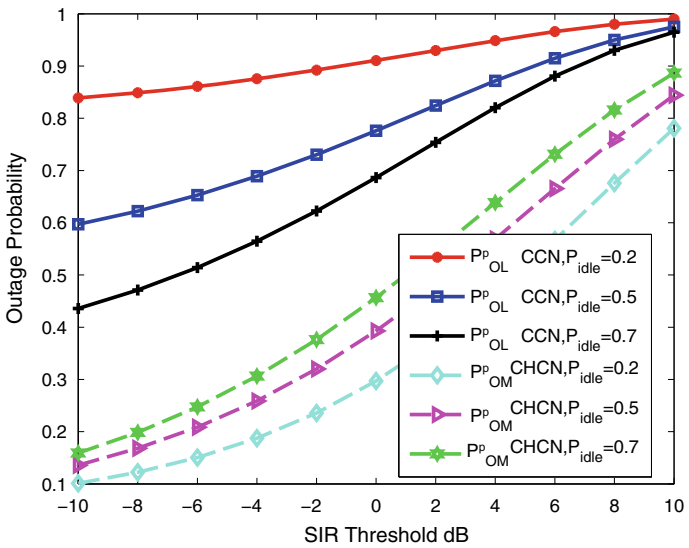



**Fig. 7.1** Outage probability versus SIR threshold for CPBS tier with $m = 1$ (Rayleigh fading)

compared to CHCN. However, in CCN, for the idle state of the MBS network, CPBS tier opportunistically utilizes the free channels of the MBS tier, whereas if MBS network is in the busy state (i.e., $\mathcal{P}_{idle} = 0$), then the SIR outage probability of CPBS network becomes one because of the channels unavailability; as a result, poor outage performance is noted as compared to the CCN. In CHCN, when free channels of MBS tier are not available for communication, CPBS exploits mmWave band to establish the communication link. Consequently, improved outage performance is achieved (as demonstrated graphically in Fig. 7.1 ). In CHCN, $\mathcal{P}_{idle} = 0.5$ indicates that 50%

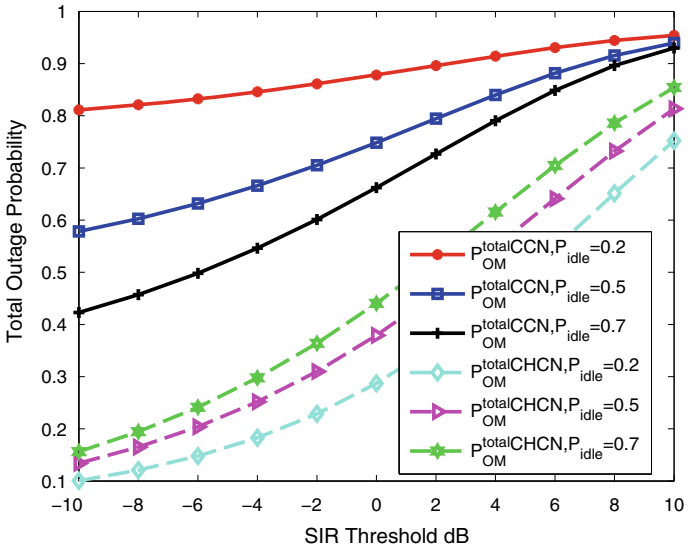



**Fig. 7.2** Total outage probability versus SIR threshold

time MBS network is in the idle condition, and CPBS is working with the cellular band and remaining 50% time when MBS network is in the busy state, CPBS utilizes mmWave band. Numerical values of outage probabilities of CPBS tier for CHCN and CCN under $\mathcal{P}_{idle} = 0.5$ and $\theta_p = 0dB$ are 0.3928 and 0.7761, respectively.

In Fig. 7.2, total outage probabilities of CHCN and CCN under open access mode are demonstrated for different $\mathcal{P}_{idle}$. The total outage performance of CHCN provides a better result as compared to CCN. Numerical values of outage probabilities for CHCN and CCN are 0.3791 and 0.7485, respectively. $\mathcal{P}_{idle} = 0.5$, and $\theta_p = 0dB$ are considered for calculating these results. Justifications of these results are similar to the results, as described in Fig. 7.1 with reference to CPBS tier outage performance. Figure 7.3 represents the relation between outage probabilities and CPBS densities for CHCN and CCN. Figure 7.3 reveals that interference created by the growing density of the CPBS noticeably impairs the outage probability of CHCN and CCN. As the density of the CPBSs increases, co-tier interference of the network severely affects the performance of the network. It is also seen that CHCN offers better outage performance. For instance, outage probability of the CHCN and CCN for SIR threshold of 0dB and $\lambda_p = 3 \times 10^5$ m$^{-2}$ are 0.3086 and 0.6779, respectively.

Area spectral efficiencies of the CHCN, CCN, and UHF-based single-tier network (USTCN) are compared in Fig. 7.4. It is noted that ASE of CHCN provides better area spectral efficiency as compared to CCN, and USTCN. Explanation of these results is similar to the results of Fig. 7.1. For SIR threshold of 0 dB, numerical values of area spectral efficiency are $2.929 \times 10^{-5}$ bps/Hz/m$^2$, $1.083 \times 10^{-5}$ $bps/Hz/m^2$, and $5.941 \times 10^{-7}$ $bps/Hz/m^2$ for CHCN, CCN, and USTCN, respectively. It may be seen that the area spectral efficiency of the system improves with the increasing

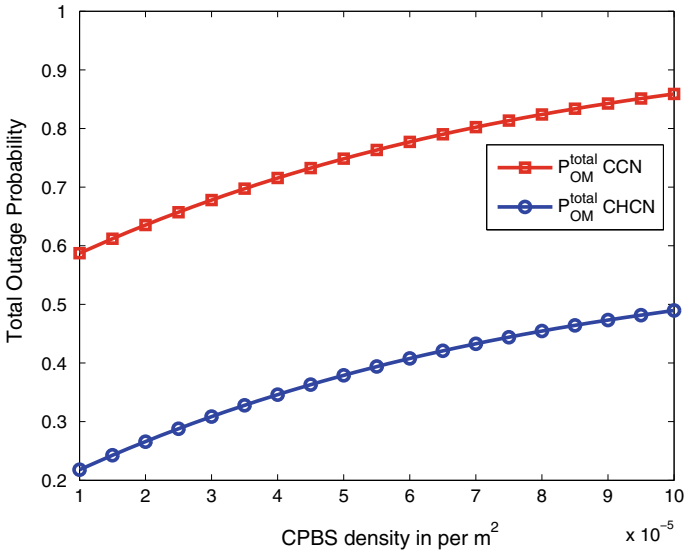



**Fig. 7.3** Total outage probability for CMCN with base station densities

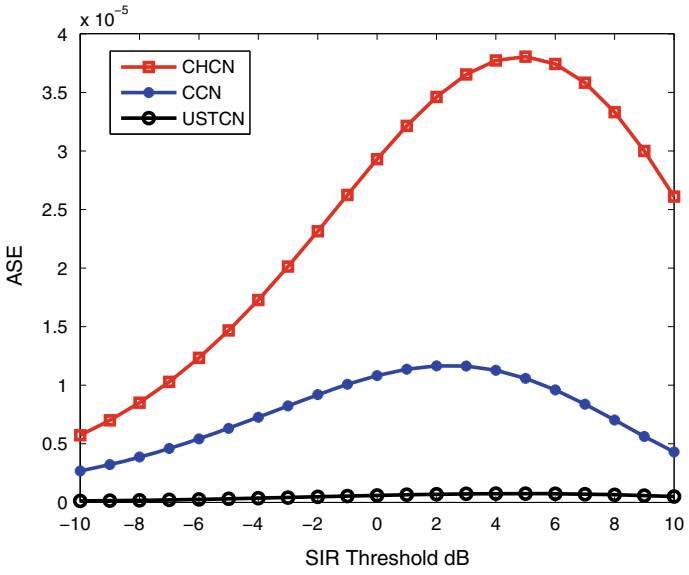



**Fig. 7.4** Area spectral efficiency versus SIR threshold

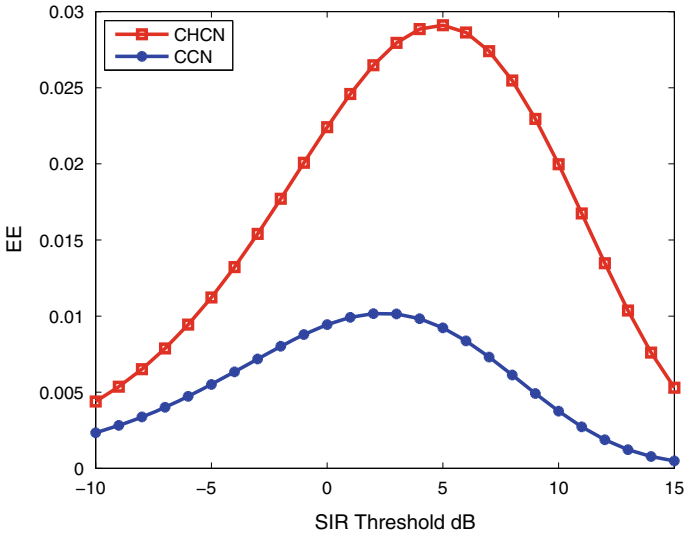



**Fig. 7.5** Energy efficiency versus SIR threshold

values of the SIR threshold. It is interesting to mention that a high SIR threshold corresponds to high link spectral efficiency. However, larger values of SIR threshold (beyond a particular SIR value) also result in increased outage probability; hence, area spectral efficiency diminishes after that value of the SIR threshold.

In Fig. 7.5, energy efficiency of the CHCN and CCN is studied. However, in Fig. 7.5, we consider only the operating power of the corresponding base stations to evaluate the energy efficiency of CHCN and CCN. This is because of the unavailability of the exact values of the embodied power for MBS, CPBS, and mmWBS in literature. It can be observed that similar to the area spectral efficiency (as shown in Fig. 7.4), energy efficiency of the network also improves with SIR threshold for specific value. It is also observed that energy efficiency of CHCN provides better energy efficiency performance as compared to CCN. For SIR threshold of 0 dB, the numerical values of energy efficiency are 0.02241 and 0.009437 bits/J for CHCN and CCN, respectively.

## 7.6 Future Research Directions

This section presents outlines for future work that are possible extensions of the work reported in the literature. We expect that these research directions may be explored to enhance the performance of the network and for better and more practical analysis. Suggestions for future research are listed below:

- In most of the existing research works, mmWave base stations are distributed by HSPPPs to develop the mathematical framework using distinct physical layer techniques. HSPPP is the general point process to characterize the cellular networks. However, to provide better modeling and more accurate analysis, it is essential to choose point processes that carry the properties of cellular networks with practical network models. Some other point processes, such as Matérn hardcore point process and Poisson cluster point process, offer more realistic and exact modeling for the mmWave-based cellular networks. Therefore, one fundamental research direction is to extend the analyses for a more practical point process, including Matérn hardcore point process, Poisson cluster point process, which can capture more accurately the characteristics of the cellular networks. Similarly, inhomogeneous Poisson point process (IPPP, also known as non-stationary PPP) is useful to model the network with spatially varying intensity. Hence, for a more realistic study, consideration of IPPP could be another new research problem.
- In the present works, locations of the mmWave transmitting nodes and blockages are distributed as independent HSPPP with different densities. However, in a realistic scenario, locations of some base stations and obstructions may be correlated (Bai and Heath 2015). Therefore, existing analysis can be extended by considering the correlation between mmWave base station and blockages.
- Hybrid sub-3GHz and mmWave-based network may be one approach for better exploitation of this band. However, some issues arise due to the propagation characteristics of mmWave and random deployment of the mmWave transmitters. These issues are: What should be the mobile station association strategy for the hybrid cellular network? What should be the spectrum sharing scheme? What should be the spatial modeling approach for the hybrid cellular network? These questions motivate to rethink and re-investigate the traditional policy of accessing the radio resources and association rules for such a hybrid system.
- mmWaves signal transmission characteristics depend on the operating frequency and scenarios. Even with the extensive experiments/trials performed in the 28 GHz, 38 GHz, E bands, the transmission mechanism of a broad mmWave band remains unidentified. Similarly, different scenarios are also needed to examine, for instance, crowded open market and streets, indoor/outdoor sports hall, airports, railway stations, etc. Therefore, theoretical and experimental analysis of different mmWave bands and scenarios are essential research direction. Furthermore, the transmission mechanism of mmWave with advanced physical layer techniques, such as NOMA, cognitive radio, intelligent reflecting surface, and full duplexing, needs to be investigated for mmWave bands.
- Coverage/outage probability is an essential performance metric in cellular networks for analyzing the performance of the system. However, recent researches suggest that in multi-tier mobile networks, user experience depends on the application-level data rate. In most of the literature, coverage/outage probability is determined based on the received signal strength at the tagged mobile user. This may be extended via analyzing coverage/outage of the system based on the aggregate data rate over some threshold, which is the more appropriate definition of coverage/outage for multi-tier cellular networks. Besides, another possible

extension is evaluation and investigation of the other necessary performance measures such as ad secrecy outage and performance measures related to the handover procedures such as cell boundary cross rate and cell dwell-time for mmWave systems.

- Spectral efficiency, energy efficiency, and deployment cost-efficiency optimization results available in the literature are mostly for the shared spectrum-based multi-tier network or cognitive-based ad hoc network. Hence, spectral efficiency, energy efficiency, and deployment cost efficiency of the mmWave-based cellular network and the relationship between spectral efficiency and energy efficiency, and spectral efficiency and deployment cost efficiency could be an interesting analysis.
- Optimal density analysis: Dense and unplanned deployment of low-power base stations raise essential issues about spectral and energy efficiency, whether they increase continuously or decrease or saturate at a certain point due to increasing interference and transmission impairments. Therefore, it is interesting to understand the consequence of the increasing density of the mmWave small base stations on system performance.
- MIMO has already proved its ability to gain better spectral efficiency for different wireless networks. Therefore, MIMO is a crucial part of the various wireless standards. However, due to the small dimension of the mobile devices, only a limited number of antenna elements may be placed at the receiving nodes. As a result, further improvement in the throughput of cellular systems is difficult with the traditional transmission frequency, i.e., the UHF band. On the other hand, recent findings indicate that some mmWave bands are feasible for cellular communications, and the cellular network infrastructure should be expanded correspondingly. Accordingly, using the small wavelength characteristics of the mmWave, we can fix a large number of antenna elements on a relatively small physical area at the transmitting and receiving nodes to employ massive MIMO and adaptive beamforming. In massive MIMO, a huge number of antennas are placed at the transmitting node and receiving node to transmit accurately concentrated energy to the receiving node resulting in lower propagation losses and higher spectral efficiency. Current research on mmW transmission suggests that array sizes of 32–256 antenna elements at the transmitting node and 4 to 16 antennas at the mobile station are suitable for mmWave-based MIMO systems (Andrews et al. 2017; Hong et al. 2014). These vast numbers of the antennas in a small physical area are possible because of the small wavelength of mmWave and recent findings in antenna circuit design (Hong et al. 2014). Although, these large arrays cannot be exploited for mmWave in a similar approach as they utilized for the conventional cellular bands. In the existing cellular networks, precoding and combining are achieved at the baseband using digital signal processing. But, such fully digital processing is difficult to perform at mmWave signals with broad bandwidth and a massive number of antenna elements (Andrews et al. 2017). Therefore, a typical MIMO system that assigns a radio frequency chain for all antennas is not easy to realize. Hence, one possible extension of existing works with a massive MIMO approach could be the analysis of different types of beamforming, such as analog beamforming, digital beamforming, and hybrid beamforming at receiving nodes.

This may give insights that MIMO architectures are suitable for mixed cellular networks. Furthermore, most existing works on mmWave networks assumed that perfect channel state information at the base station. Therefore, investigating the impact of imperfect channel information could be significant future work. Furthermore, incorporating a massive MIMO concept with a mixed cellular network and developing a mathematical framework could be another extension.

- In wireless communication, the optimization problem may be classified into two categories, i.e., statistical optimization and instantaneous optimization problem. Instantaneous optimization evaluates the optimal value of the instantaneous system parameters, such as allocation of the channel and power, and mode of operation. It assures optimal network performance at any time. On the other hand, statistical optimization determines the statistical system parameters, such as the base station density and distribution for channel gains, and gives the best-averaged network performance. Existing works could be extended for the instantaneous parameter optimization and statistical optimization to optimize the network from the perspective of network operators.
- In most of the current literature to keep the analysis tractable and straightforward, assumed that mobile users are either an indoor user or it is not moving much when the outdoor scenario is considered. However, handover is a necessary characteristic/process in the cellular network, and the existing work can be extended with the consideration of the handover condition for more accurate analysis.

## 7.7 Conclusion

This chapter provided a comprehensive overview of growing mmWave-based 5G and beyond cellular system. We reviewed mmWave channel transmission characteristics and challenges associated with mmWave-based cellular networks. Furthermore, we presented useful insights related to mmWave channel modeling and deployment based on the study of existing measurement/trials performed by different research groups and industries. Besides, we provided an overview of the relevant system design guidelines, e.g., identification of the system scenario, operating frequencies, blockages, considerations of multipath fading, etc. This chapter also proposed a hybrid cellular network using the cognitive radio approach and mmWave band for future mobile networks. An analytical framework is developed to determine the outage probability, area spectral efficiency, and energy efficiency of the system. As the analysis of the cellular networks by considering the random network topology is more practical than the conventional grid-based system, particularly for multi-tier deployment, therefore, theoretical expressions for outage probability are derived by using the tools from stochastic geometry. We expect that useful insights gained from this chapter will be helpful for solving some of the challenges associated with the mmWave-based 5G and beyond networks. It is also expected that the theoretical

framework developed in this work will continue to be improved and may be considered as a baseline model to establish more spectral and energy efficient systems by employing different strategies.

## References

5G channel model for bands up to 100 GHz, Aalto University, Espo, Finland, White Paper, Dec 2015

Akdeniz MR, Liu Y, Samimi MK, Sun S, Rangan S, Rappaport TS, Erkip E (2014) Millimeter wave channel modeling and cellular capacity evaluation. IEEE J Sel Areas Commun 32(6):1164–1179

Akoum S, El Ayach O, and Heath RW (2012) Coverage and capacity in mmwave cellular systems. In: 2012 conference record of the forty sixth Asilomar conference on signals, systems and computers (ASILOMAR)

Andrews J (2013) Seven ways that hetnets are a cellular paradigm shift. IEEE Commun Mag 51(3):136–144

Andrews J, Baccelli F, Ganti R (2011) A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. 59(11):3122–3134

Andrews JG, Bai T, Kulkarni MN, Alkhateeb A, Gupta AK, Heath RW (2017) Modeling and analyzing millimeter wave cellular systems. IEEE Trans Commun 65(1):403–430

Bai T, Heath RW (2015) Coverage and rate analysis for millimeter-wave cellular networks. IEEE Trans Wireless Commun 14(2):1100–1114

Bai T, Vaze R, Heath RW (2014) Analysis of blockage effects on urban cellular networks. IEEE Trans Wireless Commun 13(9):5070–5083

Biswas S, Vuppala S, Ratnarajah T (2016) On the performance of mmwave networks aided by wirelessly powered relays. IEEE J Sel Top Signal Process 10(8):1522–1537

Cheng S, Ao WC, Tseng F, Chen K (2012) Design and analysis of downlink spectrum sharing in two-tier cognitive femto networks. IEEE Trans Veh Technol 61(5):2194–2207

Chu X, Lopez-Perez D, Yang Y, Gunnarsson F (2013) Heterogeneous cellular networks theory, simulation and deployment, 1st edn. Cambridge University Press, Cambridge

Cisco (2017) Cisco visual networking index: global mobile data traffic forecast update, 2016 – –2021. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf

Dhillon H, Ganti R, Baccelli F, Andrews J (2012) Modeling and analysis of k-tier downlink heterogeneous cellular networks. IEEE J Sel Areas Commun 30(3):550–560

Dhillon H, Kountouris M, Andrews J (2013) Downlink mimo hetnets: modeling, ordering results and performance analysis. IEEE Trans Wireless Commun 12(10):5208–5222

ElSawy H, Hossain E (2013) A modified hard core point process for analysis of random csma wireless networks in general fading environments. IEEE Trans Commun 61(4):1520–1534

ElSawy H, Hossain E (2014) Two-tier hetnets with cognitive femtocells: downlink performance modeling and analysis in a multichannel environment. IEEE Trans Mobile Comput 13(3):649–663

ElSawy H, Hossain E, Haenggi M (2013) Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey. IEEE Commun Surveys Tutorials 15(3):996–1019

ElSawy H, Sultan-Salem A, Alouini MS, Win MZ (2017) Modeling and analysis of cellular networks using stochastic geometry: a tutorial. IEEE Commun Surveys Tutorials 19(1):167–203

Fellers RG (1956) Millimeter waves and their applications. Electr Eng 75(10):914–917

Further advancements for E-UTRA physical layer aspects (release 9), document 3GPP TR 36.814, 3GPP

Ge X, Cheng H, Guizani M, Han T (2014) 5G wireless backhaul networks: challenges and research advances. IEEE Network 28(6):6–11
Gupta AK, Andrews JG, Heath RW (2016) On the feasibility of sharing spectrum licenses in mmwave cellular systems. IEEE Trans Commun 64(9):3981–3995
Haenggi M (2012) Stochastic geometry for wireless networks, 1st edn. Cambridge University Press, Cambridge
Haenggi M, Ganti RK (2009) Interference in large wireless networks. Now Publishers Inc
Hasan Z, Boostanimehr H, Bhargava V (2011) Green cellular networks: a survey, some research issues and challenges. IEEE Commun Surveys Tutorials 13(4):524–540
Hemadeh IA, Satyanarayana K, El-Hajjar M, Hanzo L (2018) Millimeter-wave communications: physical channel models, design considerations, antenna constructions, and link-budget. IEEE Commun Surveys Tutorials 20(2):870–913
Hong W, Baek KH, Lee Y, Kim Y, Ko ST (2014) Study and prototyping of practically large-scale mmwave antenna systems for 5G cellular devices. IEEE Commun Mag 52(9):63–69
Jo HS, Sang YJ, Xia P, Andrews JG (2012) Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis. IEEE Transactions on Wireless Communications 11(10):3484–3495
Lee C-H, Haenggi M (2012) Interference and outage in poisson cognitive networks. IEEE Trans Wireeless Commun 11(4):1392–1401
Mishra MK, Trivedi A (2015) Performance analysis of two tier cognitive heterogeneous cellular network. In: Twenty first national conference on Communications (NCC) 2015, IEEE, pp. 1–6
Mishra MK, Trivedi A, Pattanaik KK (2018) Outage and energy efficiency analysis for cognitive based heterogeneous cellular networks. Wireless Netw 24(3):847–865. [Online]. Available: https://doi.org/10.1007/s11276-016-1371-3
Mishra MK, Trivedi A (2019) Spectral efficiency and deployment cost efficiency analysis of mmw/uhf-based cellular network. IEEE Trans Veh Technol 68(7):6565–6577
Mukherjee S (2012) Distribution of downlink sinr in heterogeneous cellular networks. IEEE Wireless Commun Sel Areas Commun 30(3):575–585
Panahi FH, Ohtsuki T (2014) "Stochastic geometry based analytical modeling of cognitive heterogeneous cellular networks. In: IEEE international conference on communications (ICC), pp 5281–5286
Rangan S, Rappaport TS, Erkip E (2014) Millimeter-wave cellular wireless networks: potentials and challenges. Proc IEEE 102(3):366–385
Rappaport TS, Sun S, Mayzus R, Zhao H, Azar Y, Wang K, Wong GN, Schulz JK, Samimi M, Gutierrez F (2013) Millimeter wave mobile communications for 5g cellular: it will work!. IEEE Access 1:335–349
Rappaport JRCDTS, Heath RW, Murdock JN (2014) Millimeter wave wireless communications. Prentice-Hall, Englewood Cliffs
Rappaport TS, Xing Y, MacCartney GR, Molisch AF, Mellios E, Zhang J (2017) Overview of millimeter wave communications for fifth-generation (5G) wireless networks-with a focus on propagation models. IEEE Tran Antennas Propag 65(12):6213–6230
Singh S, Dhillon H, Andrews J (2013) Offloading in heterogeneous networks: modeling, analysis, and design insights. IEEE Trans Wireless Commun 12(5):2484–2497
Singh S, Kulkarni MN, Ghosh A, Andrews JG (2015) Tractable model for rate in self-backhauled millimeter wave cellular networks. IEEE J Sel Areas Commun 33(10):2196–2211
Soh YS, Quek T, Kountouris M, Shin H (2013) Energy efficient heterogeneous cellular networks. IEEE Wireless Commun Sel Areas Commun 31(5):840–850
Sulyman AI, Nassar AT, Samimi MK, MacCartney GR, Rappaport TS, Alsanie A (2014) Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands. IEEE Commun Mag 52(9):78–86
Swami P, Mishra MK, Trivedi A (2017) Analysis of downlink power control and cooperation scheme for two-tier heterogeneous cellular network. Int J Commun Syst 30(13):e3282 [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.3282

Tao X, Zhao Z, Zhang H (2013) Location information based interference control for cognitive radio network in tv white spaces. In: IEEE Wireless communication and network conference (WCNC), 2013. IEEE, pp 3614–3619

Thornburg A, Bai T, Heath RW Jr (2015) Mmwave ad hoc network coverage and capacity. In: IEEE international conference on communications (ICC), pp 1310–1315

Thornburg A, Bai T, Heath RW (2016) Performance analysis of outdoor mmwave ad hoc networks. IEEE Trans Signal Process 64(15):4065–4079

Wang C-X, Haider F, Gao X, You X-H, Yang Y, Yuan D, Aggoune H, Haas H, Fletcher S, Hepsaydir E (2014) Cellular architecture and key technologies for 5G communication networks. IEEE Commun Mag 52(2):122–130

Zaidi SAR, McLernon DC, Ghogho M (2010) Outage probability analysis of cognitive radio networks under self-coexistence constraint. In: 44th annual conference on information sciences and systems (CISS), pp 1–6

Zhang Z, Ryu J, Subramanian S, Sampath A (2015) Coverage and channel characteristics of millimeter wave band using ray tracing. In: IEEE international conference on communications (ICC), pp 1380–1385

# Chapter 8
# Multi-user Full-Duplex Two-Way Relaying Systems with User Mobility


**Prasanna Raut and Prabhat Kumar Sharma**

**Abstract** In the last decade, the wireless inband full-duplex (FD) communications have been extensively studied to make a better use of available spectral resources. The FD communication allows the simultaneous transmission and reception on the same frequency band within the same device, which theoretically doubles the spectral efficiency. Moreover, FD multiple user communication assisted by spectral efficient bidirectional relays provides better coverage, better diversity gains, and improved capacity performance. This chapter examines the recent works done in the area of mobile multi-user FD relaying systems from the physical layer perspective. This chapter begins with the comparison of traditional half duplex techniques with the recently developed FD radios and the modeling of residual self-interference at the FD node. Further, recent works related to the half-duplex and/or FD, two-way and/or one-way relaying where the impact of user mobility is taken into account are discussed in detail. Moreover, this chapter surveys the multi-user systems, and the performances of various multi-user scheduling schemes are compared in detail. Finally, this chapter includes some applications of FD relaying with short-packet communication for its possible applications in mission-critical applications such as virtual reality, autonomous driving, telemedicine, etc.

**Keywords** Full-duplex · Multi-user systems · Two-way relaying · Residual self-interference · Mission-critical applications

## 8.1 Introduction

Wireless devices are increasingly spreading in modern-day life. With the increase in mobile phones, streaming live videos through base stations, computers, and laptops

P. Raut (✉) · P. Kumar Sharma
Department of ECE, Visvesvaraya National Institute of Technology, Nagpur, India
e-mail: rautprasannad@gmail.com

P. Kumar Sharma
e-mail: prabhatmadhavec1@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_8

connecting to the wireless fidelity (Wi-Fi) network, the upcoming wireless communication systems (WCS) are demanding higher speed and higher network availability (Ericsson 2019). Despite of the plethora of research and the advancements in the design of WCS over the last few decades, current WCS still experience the problems such as poor connection and performance loss. Having a strong signal on the Wi-Fi connection of a laptop or a cellular phone and still unable to load the requested page in a Web browser or mobile application can be a good example of such problems. Moreover, in the crowded wireless environments, such problems are much more frequent and noticeable. Hence, design and analysis of WCS are of utmost importance from the implementation perspectives.

Electromagnetic wave-based wireless radio transmission has been invented back in eighteenth century (Maxwell 1865). The first wireless telegraphy system constructed by Guglielmo Marconi in 1800s initiated the modern wireless communication (Klooster 2009). The demonstration of wireless radio communications by the scientist *N*. Tesla triggered the era of radio development. The way of information transmission and the applications of radio has been changed drastically right from the first telegraph to the first wired/wireless phone, and even to the current and upcoming digital cellular devices. The key application of wireless radio earlier was to ensure the transmission of information at far distance and with wide range. The role of wireless devices is increasing drastically in our day-to-day life. The key application of wireless radio earlier was to ensure the transmission of information at larger distances emphasizing on voice messages or voice signals. In contrast, with the rapid growth of wireless communication in recent past, data traffic has taken over the voice traffic and has increased in the recent years exponentially (Ericsson 2019).

The current and upcoming WCS [fifth generation (5G) and beyond] are expected to deal with the challenges faced by the existing systems. These challenges mainly include the high data rate transmission due to immense hike in the deployment of multimedia applications and the way people are looking forward to the Internet. Moreover, the upcoming WCS are also expected to fulfill the requirements of various practical use cases. These cases mainly include enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC) (Al-Fuqaha et al. 2015; Abedin et al. 2019; Liu et al. 2017; Sutton et al. 2019; Bockelmann et al. 2016). eMBB is an extended version of conventional MBB in terms of enhanced or improved capacity, data rates and coverage (Abedin et al. 2019; Liu et al. 2017). For coming mission-critical applications with high reliability and low-latency requirements such as virtual reality, industrial Internet, smart grids, infrastructure protection, tele-medicines, intelligent transportation systems (Sutton et al. 2019), the URLLC has become the basic requirement and key enabler. Lastly, mMTC is need to deal with the upcoming 5G Internet-of-things (IoT) scenario demanding higher network density with billions of devices and sensors connected to each other (Al-Fuqaha et al. 2015; Bockelmann et al. 2016).

In most of the WCS, duplex techniques are used as available resource (frequency, time, etc.)-sharing solutions. Currently deployed WCS employ devices which use

either a time-division duplex (TDD) or frequency-division duplex (FDD) approach to differentiate the wireless transmission and reception of signals.

- FDD-enabled WCS use two different channels or frequency bands which are separated by a sufficient amount of guard band. The guard bands help to rectify the interference that exists between the transmitter and receiver.
- In contrast, a single frequency band for transmission and reception is used in TDD. The TDD shares that single frequency band by assigning different time slots to transmitter and receiver.

Given the scarcity and expense of spectrum, FDD requires double the spectrum compared to TDD. However, FDD is widely adopted in cellular communication (e.g., the global system for mobile (GSM) system). Few system employs the 869–894 MHz band, i.e., 25-MHz band as the downlink (DL) spectrum (from the base station to the phone) and the 824–849 MHz band as the uplink (UL) spectrum. TDD is adopted in most of the wireless data transmission like WiMAX (Worldwide Interoperability for Microwave Access), Wi-Fi, Bluetooth, or even ZigBee. Though TDD seems to be the better choice, FDD is comprehensively implemented due to the prior frequency spectrum assignments and its applications in current and earlier wireless technologies. However, when it comes to the radio resource utilization and medium access, FDD or TDD is ineffective. FDD which demands the paired spectrum for UL and DL is incompetent to fulfill the demand of flexible spectrum access. TDD which requires isolated time slots for UL and DL connection is ineffective in real-time physical communications. One possible way to deal with this issue of uneven channel use in TDD without the requirement of extra bandwidth requirement with FDD can be the design of a radio that can transmit and receive information using the same carrier frequency at the same time instant.

## 8.2 Practical Full-Duplex

Compared to the currently deployed WCS which operates in half-duplex (HD) mode using either TDD or FDD technique, the full-duplex (FD)-enabled radios are twice efficient in terms of spectral efficiency. The reason behind this fact is that there is no need to share the resources between transmission and reception, and hence, there is full availability of time–frequency resource for both. This effectively doubles the resources. The FD mode ensures that one allocated spectrum can be used for transmission and reception at the time instant in such a way that the demands of utilization of resources and flexible access to the medium can be achieved unlike the case of either FDD or TDD. Thus, recently developed FD radio which can double the spectral efficiency has started gaining the research attention in the recent past.

### *8.2.1 Key Challenge*

The wireless transmission fades and attenuates significantly over large distance. The signal received from a own transmitting antenna of a node (local signal) is very much stronger, practically thousands of times, than the signal received from other nodes transmitting antenna using the same frequency. This reveals that the key challenge in design and implementation of a single channel FD wireless radio is mitigating or canceling the self-interference (SI) signal received from its own transmitting antenna.

#### 8.2.1.1 The Full-Duplex Self-Interference

The FD radios have been widely investigated in the plethora of recent literature due to the ability of improving the spectral efficiency and to compensate the capacity loss by the factor of 1/2 compared to HD systems. The FD radios, however, are more prone to the SI between the local transmit and receive antennas. This significantly limits the system achieving double performance. The strength of SI in FD devices can be 60–90 dB (decibel-watts) higher than the signal of interest. With this high amount of SI, almost all the available SI cancelation techniques cannot minimize this huge amount of interference (Li et al. 2017; Sharma and Garg 2014). Despite of the advancement in SI cancelation techniques and algorithms, entirely attenuating the SI is an impossible task. Based on the empirical results in Duarte et al. (2012), around 70 dB of SI cancelation is insufficient for the interference to be in acceptable noise levels. Due to this fact, the modeling and characterization of residual self-interference (RSI) (SI that remains after SI cancelation techniques) have an critical role to play in the design and performance evaluation of FD enabled wireless communication systems.

#### 8.2.1.2 Self-interference Mitigation/Cancelation Techniques

A lot of efforts have been made to minimize or to cancel the SI imposed by the FD radio. Till date, researchers worldwide have proposed many schemes to mitigate or cancel SI at FD node to great extent. These schemes mainly include antenna cancelation (Choi et al. 2010), passive SI suppression (PSIS) (Everett et al. 2014; Aryafar et al. 2012), and active SI cancelation (ASIC) (Duarte et al. 2014).

- In the SI suppression method with antenna cancelation, multiple antennas are deployed to ensure the received RF signals add destructively, or to get a signal with internal phase shifting by 180° (an inverse version of the signal). The important two antenna cancelation schemes include antenna cancelation based on $\lambda/2$ spacing and based on symmetry (Choi et al. 2010).
- The electromagnetic isolation-based techniques are used in PSIS SI cancelation method to mitigate the power of the signal interfering from its local transmitter to the local receiver. These techniques primarily include absorptive shielding,

directional antenna, antenna polarization, antenna separation, etc. (Everett et al. 2014; Aryafar et al. 2012). In the near-field wireless communications scenarios, the absorptive shielding and antenna polarization can provide almost 25 dB absorption (Everett et al. 2014) and 10 dB SI isolation (Aryafar et al. 2012), respectively. Antenna separation can result in approximately 40 dB of interference suppression with 30 cm of transmitter and receiver antenna separation. Directional antenna can achieve around 38 dB of interference suppression and the amount of suppression critically depends on the configuration of antenna (Everett et al. 2014).

- ASIC technique highly relies on the knowledge of its own transmitted signal. Based on the scenario of adoption of SI cancelation in the radio receiver (i.e., before or after the analog-to-digital converter), the ASIC is primarily divided into two categories: active analog SI cancelation and active digital SI cancelation (Duarte et al. 2012, 2014).

#### 8.2.1.3 Characterization of the Residual Self-interference

The modeling and characterization of RSI at FD node primarily depend on the interference cancelation capabilities and algorithms. Based on the proposed empirical results on FD wireless communication system (Duarte et al. 2012), the SI channel can be modeled as rice distribution (having small Rician parameter $K$ values, i.e., $K \leq 7$ dB before active SI cancelation and $K \leq 3$ dB after active SI cancelation) than Rayleigh distribution. This is due to the fact that the Kullback–Leibler distances observed from results are lesser for the Rice distribution.

One more reason to characterize the RSI as a Rice distribution is the significantly large power difference between the RSI signal and the signal of interest. This power difference occurs possibly due to the presence of line-of-sight (LOS) component between local transmit and local receive antennas. Even with the availability of better and efficient SI suppression techniques, the fact of presence of multipath component cannot be neglected. In the absence of LOS, the SI channels can be modeled as Rayleigh distributed random variable. Contrary, in case of presence of LOS component, Rician distribution is better fit.

### *8.2.2 Opportunities*

Cellular, Wi-Fi, Bluetooth, and femto-cell networks are undoubtedly the most commonly used wireless networks these days. Out of these, cellular network was the first deployed technology. It operates in the range of few kilometers and uses low transmit powers [almost 30 dBm (decibel-milliwatts)] for mobile devices. For these values of transmit powers and distances between communicating mobile devices, canceling the SI enough to enable FD wireless communications is infeasible. Thus, the practical deployment of FD wireless communications remained the paradigm for design and development of current and upcoming WCS.

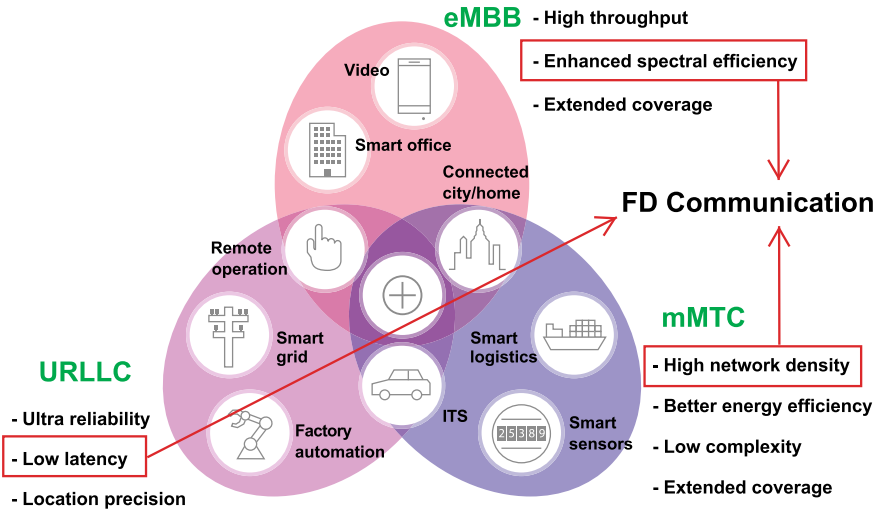



**Fig. 8.1** Practical use cases and examples of related applications where FD communication can be possibly deployed (Ericsson 2019)

In contrast, Wi-Fi, Bluetooth, and femto cells require lower transmission powers and less distance between communicating devices with the same size and equal or sometimes increased computation power of mobile devices compared to cellular networks. This implies that the power differential between the SI and the expected signal will be lower for FD communications in Wi-Fi, Bluetooth, or femto-cell networks compared to the FD cellular network. Thus, practical FD systems are feasible for various applications in Wi-Fi systems and have many advantages in Bluetooth and femto-cell networks. Moreover, as can be seen from Fig. 8.1, FD communication can possibly be applied for important used cases of upcoming WCS which include enhancing the spectral efficiency for eMBB, reducing the latency requirement for URLLC, and providing dense connectivity for mMTC.

## 8.3 Multi-user Full-Duplex Relaying

### *8.3.1 Literature Review*

In recent past, the limited spectrum resources are used to satisfy the increasing data rate demands. Current and upcoming wireless applications such as fourth generation (4G) and 5G insist high data rates, larger network capacity, higher spectral efficiency, better energy efficiency, multi-user diversity gains, and more mobility. To fulfill such purpose of upcoming WCS, many promising technologies such as two-way (TW) relaying (Liang et al. 2013; Zhang et al. 2016; Raut et al. 2019a, b), FD radios (Li

et al. 2017; Sharma and Garg 2014; Raut et al. 2019a, b), multi-user communication (Li et al. 2017; Zhang et al. 2012; Lei et al. 2013; Jang and Lee 2010; Joung and Sayed 2010; Raut et al. 2019a, b) have been recommended in the literature (and the references therein). The traditional duplex techniques use frequency or time as a resource-sharing solutions. However, recently, FD radio was introduced which can achieve transmission and reception on the same carrier frequency, simultaneously, which in result doubles the spectral efficiency unlike the traditional duplex options. Currently, to halve the number of time slots required for information exchange compared to one-way relaying (OWR), TWR has been emerged as efficient technique for exploiting spatial diversity with non-line-of-sight communication and has been adopted in literature (Zhang et al. 2016; Liu and Kim 2010; Li et al. 2017; Raut et al. 2019a, b) widely to improve the performance of the WCS. Further, in 4G wireless communications networks, relay-aided communication has emerged as an effective way to provide the information exchange between the cell-edge users (Hoymann et al. 2012; Xia et al. 2016). The field test results provided in Gan et al. (2012) demonstrate that a relay station can significantly increase the indoor received signal power, and hence is a good way to improve outdoor to indoor coverage.

In literature, amplify-and-forward (AF) and decode-and-forward (DF) are the two most common adopted relaying strategies (Liang et al. 2013; Zhang et al. 2012, 2016; Sharma and Garg 2014; Li et al. 2017; Lei et al. 2013; Jang and Lee 2010; Joung and Sayed 2010; Raut et al. 2019a, b). In AF relaying, relay simply broadcasts the amplified version of received signal with proper amplification factor. In contrast, the received signal is decoded first in DF-based relaying strategy. Later, relay broadcast the re-encoded version of the decoded signal. Sharing the antennas among users via relaying, the channel capacity can be increased and full diversity gain can be achieved. Recently, most of the works done in FD relay-aided communication considered single-source single destination scenario (Liang et al. 2013; Zhang et al. 2016; Sharma and Garg 2014). However, in the modern WCS, there is always possibility of multi-user environment (Li et al. 2017; Zhang et al. 2012; Lei et al. 2013; Jang and Lee 2010; Joung and Sayed 2010) to exploit the multi-user diversity gains. The multi-user diversity also helps to improve the performance of highly dense cellular networks that have limited resources. Because of these numerous benefits, the multi-user relay-aided FD communication has received great interest of the researchers worldwide in the design of the upcoming wireless communications networks.

### *8.3.2 Work Done and Results*

In Cui et al. (2014) and Yang et al. (2015), the two-way (Cui et al. 2014) and one-way (Yang et al. 2015) FD relay systems with AF relaying protocol are analyzed and are optimized with multi-relay scenario. The performance of the system was analyzed by evaluating bit error rate, outage probability, and system capacity. In Sharma and Garg (2014), the bit error rate and system outage probability performances of a DF-based two-hop relaying system is analyzed. The results in Cui et al. (2014), Yang et al.

(2015), and Sharma and Garg (2014) infer that the RSI at FD node can significantly impact the performance of FD-based relaying systems. The closed-form expressions and the approximations for a multi-user multi antenna AF relaying system with the assumption of co-channel interference, feedback delay, etc., were derived in terms of the outage probability and average symbol error rate Huang et al. (2014). The impact of various parameters was studied, and the findings of Huang et al. (2014) give insightful information related to the impact of feedback delay on the diversity order of the system.

Multi-user system scenario consists of multiplex users exchanging information with the aid of a one or more bidirectional relays. Multi-user communication assisted by the spectral efficient bidirectional relays provides diversity gains and improved capacity performance (Li et al. 2017; Ding et al. 2011; Lei et al. 2013; Jang and Lee 2010). Numerous multi-user scheduling schemes are proposed for the purpose of selection of a user pair in all the present $N$ pairs (Li et al. 2017; Ding et al. 2011; Jang and Lee 2010). The choice of a scheduling scheme depends on the trade-off between its performance and fairness of its service to the users. Generally, a relay serves a selected (or scheduled) user pair with traditional HD transmission in four time slots, whereas the number of time slots with FD transmission mode reduces to two. Thus, FD relay can serve twice more number of user information exchanges in a given time slot in multi-user communication.

The multi-user scenario was initially studied in cellular networks (Li et al. 2017; Ding et al. 2011; Lei et al. 2013; Jang and Lee 2010), where $N$ users compete for the relay node to send the data to a base station. The authors in Ding et al. (2011) and Lei et al. (2013) studied the user selection schemes by exploiting channel correlation to select one among $N$ user pairs. In Jang and Lee (2010), authors analyzed a scheduling scheme for two-way relaying scenario where single-source $N$ destination system was applied in cellular communication to enhance the quality of the communication path between the users and the intermediate stations. The outage performance of $N-N$ FD multi-user TWR system was evaluated in Raut and Sharma (2018) with optimal scheduling scheme and Rayleigh faded RSI. The system average rate and system outage probability of one-way and two-way relaying systems were evaluated in Zhang et al. (2016). Authors in Li et al. (2017) evaluated performance of various multi-user scheduling schemes such as random scheduling, maximum-minimum scheduling, optimal scheduling. The performance was analyzed in terms of the outage probability (Li et al. 2017).

The IoT has emerged as an efficient technology that allows the devices, various objects to receive and send the data using software, sensors, etc. IoT has gained a significant interest in the recent past as it can connect various objects and devices with computer-based systems to improve their efficiency. The coming era of wireless communication will be highly dependent on the evolution of IoT technology. IoT will play a significant role where the reliability, and the latency will be of utmost importance. One such example can be of mission-critical applications such as factory automation, tele-medicines, virtual reality. For these reasons, the IoT-enabled ultra-reliable and low-latency communication (URLLC) has attracted significant interest of the research community (Hu et al. 2016a, b, 2019; Gu et al. 2018; Wang et al.

2019). In URLLC, the short-packets with finite blocklength codes (FBC) are adopted to ensure the low-latency (LL) and ultra-reliability (UR) requirements (Hu et al. 2016a, b; Gu et al. 2018; Wang et al. 2019).

The challenging aspect of IoT-enabled mission-critical communication is the battery constraints of smart nodes. Particularly, replacing or recharging the batteries is inconvenient and demands high cost for IoT scenarios. Thus, the energy harvesting (EH) has emerged as a viable solution to this issue (Hu et al. 2019; Boshkovska et al. 2015). The EH employs simultaneous wireless information and power transfer (SWIPT) through extracting the ability of radio frequency (RF) signals to carry both information and energy (Hu et al. 2019; Boshkovska et al. 2015; Haghifam et al. 2016). In the linear EH method, the amount of energy harvested by a node linearly increases with the input power (Hu et al. 2019); in contrast, the nonlinear EH model considers the constraints on minimum and maximum energy to be harvested (Boshkovska et al. 2015). Thus, the nonlinear EH models are more practical.

The mobility of users, which leads to the time-varying channels (Khattabi and Matalgah 2016, 2015), plays an important role in the performance analysis of any practical communication system. In such situations, the estimation of channel coefficients at the receiver becomes a critical task. The impact of the imperfect channel state information on the error and outage performance of a multi-relay system with lower user mobility was investigated in Khattabi and Matalgah (2016) for HD mobile nodes. Moreover, the impact of outdated channel estimation was studied in Khattabi and Matalgah (2016) and Khattabi and Matalgah (2015). Authors in Khattabi and Matalgah (2016) considered the time-selective nature of communication links due to the mobility of relay. Based on pilot signal transmission method, a channel estimation method was derived. The outage performance of the FD multi-user communication system with user mobility was evaluated in Raut and Sharma (2019), and various insights regarding the impact of RSI and mobility were obtained.

With the help of Jake's mobility model, the correlation parameter of the time-varying channel links $h_{A_j}$ and $g_{A_j}$ is given by $\rho_{A_j} = \mathcal{J}_0\left(\frac{2\pi f V_{A_j}}{cR_s}\right)$. In a similar way, for channels $h_{B_j}$ and $g_{B_j}$, $\rho_{B_j} = \mathcal{J}_0\left(\frac{2\pi f V_{B_j}}{cR_s}\right)$. Here $V_{A_j}$, $V_{B_j}$ represents the mobility, $f$ is carrier frequency, $c$ is speed of light, $R_s$ is transmission symbol rate, and $\mathcal{J}_0$ represents zeroth-order Bessel function of first kind. With the assumption of time-selective fading, the node estimates the channel coefficients over first signaling period as $\hat{h}_{A_j}(1)$, $\hat{h}_{B_j}(1)$, $\hat{g}_{A_j}(1)$, $\hat{g}_{B_j}(1)$, with the help of a pilot signal. Hence, $h_{A_j}$, $h_{B_j}$, $g_{A_j}$, and $g_{B_j}$ for $m$-th signaling position are represented as

$$\begin{aligned} h_{U_j}(m) &= \rho_{U_j}^{m-1}\left(\hat{h}_{U_j}(1) + h_{\epsilon_{U_j}}(1)\right) + \sqrt{1-\rho_{U_j}^2}\sum_{l=1}^{m-1}\rho_{U_j}^{m-l-1}e_{U_j}^{h}(l), \\ g_{U_j}(m) &= \rho_{U_j}^{m-1}\left(\hat{g}_{U_j}(1) + g_{\epsilon_{U_j}}(1)\right) + \sqrt{1-\rho_{U_j}^2}\sum_{l=1}^{m-1}\rho_{U_j}^{m-l-1}e_{U_j}^{g}(l), \end{aligned} \tag{8.1}$$

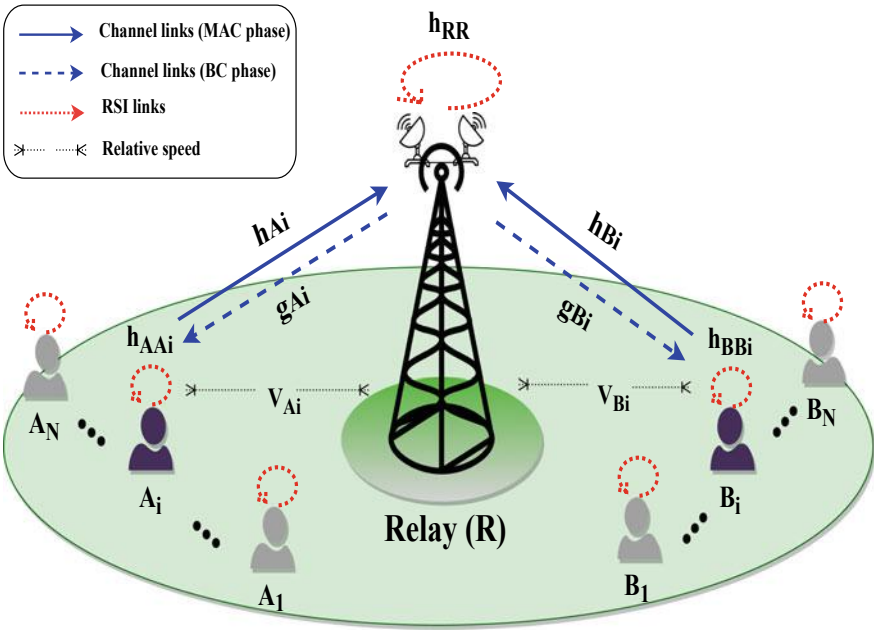



**Fig. 8.2** Considered multi-user FD TWR system. A scheduled FD user pair $A_i$ and $B_i$ exchanges the information with help of FD relay (Raut et al. 2019a)

where $\{U\} \in \{\{A\}, \{B\}\}$, $\hat{h}_{U_j}$ and $\hat{g}_{U_j}$ are the estimated channel coefficients and are distributed as zero mean circularly symmetric complex Gaussian (ZM-CSCG) with variances $\sigma^2_{\hat{h}_{U_j}}$, $\sigma^2_{\hat{g}_{U_j}}$, respectively. $h_{\epsilon_{U_j}}$ and $g_{\epsilon_{U_j}}$ are the estimation errors and are distributed as ZM-CSCG with variances $\sigma^2_{h_{\epsilon_{U_j}}}$, $\sigma^2_{g_{\epsilon_{U_j}}}$, respectively. The random processes $e^h_{U_j}(l)$, $e^g_{U_j}(l)$ are the varying components of the associated links $h_{U_j}$, $g_{U_j}$ and are distributed as $e^h_{U_j}(l) \sim \mathcal{CN}\left(0, \sigma^2_{e^h_{U_j}}\right)$, $e^g_{U_j}(l) \sim \mathcal{CN}\left(0, \sigma^2_{e^g_{U_j}}\right)$, respectively.

Consider a system shown in Fig. 8.2. The users $A_1$ to $A_N$ are exchanging the information with the paired users $B_1$ to $B_N$, $j \in \{1, 2, \ldots, N\}$, with the aid of a two-way FD relay $R$. All the user nodes are assumed to be mobile with the relative speed between $A_j$ and $R$ ($B_j$ and $R$) denoted by $V_{A_j}$ ($V_{B_j}$). The channel coefficients between the node $A_j$ and $R$ ($R$ and $A_j$) are represented as $h_{Aj}$ ($g_{Aj}$). Similarly, the channel coefficients between the nodes $B_j$ and $R$ ($R$ and $B_j$) are represented as $h_{Bj}$ ($g_{Bj}$). We assume that all the block-fading channels are subject to independent and non-identically distributed (i.ni.d.) complex Gaussian fading (Rayleigh envelope). Finally, the channel coefficients $h_{AA_j}$, $h_{BB_j}$, $h_{RR}$ represent the Rician distributed RSI channels at the nodes $A_j$, $B_j$, and $R$, respectively. Based on scheduling criterion, the relay selects a $i$-th user pair ($A_i$−$B_i$) in every time slot. The entire relaying operation is divided into two phases. First in the multiple access channel (MAC) phase, the users $A_i$ and $B_i$ from the $i$-th selected user pair simultaneously transmit their data to

$R$. Then, after some signal processing (discussed in detail in following sections), $R$ broadcasts the received signal to $A_i$ and $B_i$ in the broadcast (BC) phase.

In other words, the users $A_i$ and $B_i$ simultaneously send signals $X_{A_i}$ and $X_{B_i}$, respectively, to the relay node $R$ (MAC phase). We denote the transmission power of user nodes $A_i$, $B_i$ as $P_{A_i} = \mathbb{E}\{|X_{A_i}|^2\}$, $P_{B_i} = \mathbb{E}\{|X_{B_i}|^2\}$, respectively. Using (8.1), the received signal at the relay node in $k$-th time slot is given by

$$\begin{aligned} Y_R[k] = &\sqrt{P_{A_i}}\rho_{A_i}^{m-1}\hat{h}_{A_i}(1)X_{A_i}[k] + \sqrt{P_{B_i}}\rho_{B_i}^{m-1}\hat{h}_{B_i}(1)X_{B_i}[k] \\ &+ \sqrt{P_{A_i}}\rho_{A_i}^{m-1}h_{\epsilon A_i}(1)X_{A_i}[k] + \sqrt{P_{B_i}}\rho_{B_i}^{m-1}h_{\epsilon B_i}(1)X_{B_i}[k] \\ &+ \sqrt{P_{A_i}}\phi_{A_i}^{h}(m)X_{A_i}[k] + \sqrt{P_{B_i}}\phi_{B_i}^{h}(m)X_{B_i}[k] \\ &+ n_R[k] + \sqrt{P_R}h_{RR}(m)\, t_R[k], \end{aligned} \tag{8.2}$$

where $t_R[k]$ is the signal transmitted by $R$ with instantaneous power $P_R$ in $k$-th time slot.

Based on the relaying scheme (AF or DF), the relay $R$ broadcasts the re-encoded or amplified signal $t_R$ to users $A_i$ and $B_i$ with transmission power $P_R = \mathbb{E}\{|t_R|^2\}$ (BC phase). Using (8.1), the signals received at nodes $A_i$, $B_i$ in $k$-th time slot can be written as

$$\begin{aligned} Y_{A_i}[k] = &\sqrt{P_R}\rho_{A_i}^{m-1}\hat{g}_{A_i}(1)t_R[k] + \sqrt{P_R}\rho_{A_i}^{m-1}g_{\epsilon_{A_i}}(1)t_R[k] \\ &+ \sqrt{P_R}\phi_{A_i}^{g}(m) + n_{A_i}[k] + \sqrt{P_{A_i}}h_{AA_i}(m)\, t_{A_i}[k], \\ Y_{B_i}[k] = &\sqrt{P_R}\rho_{B_i}^{m-1}\hat{g}_{B_i}(1)t_R[k] + \sqrt{P_R}\rho_{B_i}^{m-1}g_{\epsilon_{B_i}}(1)t_R[k] \\ &+ \sqrt{P_R}\phi_{B_i}^{g}(m) + n_{B_i}[k] + \sqrt{P_{B_i}}h_{BB_i}(m)\, t_{B_i}[k], \end{aligned} \tag{8.3}$$

respectively, where the terms $\phi_{A_i}^{g}(m)$ and $\phi_{B_i}^{g}(m)$ are defined in (8.1), and $n_{A_i}[k] \sim \mathcal{CN}(0, \sigma_{A_i}^2)$, $n_{B_i}[k] \sim \mathcal{CN}(0, \sigma_{B_i}^2)$ represent the AWGN at users $A_i$, $B_i$, respectively. $t_{A_i}[k]$, $t_{B_i}[k]$ are the signals transmitted by users $A_i$, $B_i$, respectively.

Using the statistics of SINR of the physical layer signals, the symbol error probability (SEP) of the $i$-th scheduled user pair for different modulation schemes is given by

$$\text{SEP} = a\mathbb{E}\left\{Q\left(\sqrt{b\,\gamma_{\text{end}}^{i}}\right)\right\} = \frac{a}{\sqrt{2\pi}}\int_0^{\infty}\mathcal{F}_{\gamma_{\text{end}}^{i}}\left(\frac{y^2}{b}\right)\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y, \tag{8.4}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\mathrm{e}^{-y^2/2}\mathrm{d}y$ is the Gaussian function, $\gamma_{\text{end}}^{i}$ represents the end-to-end SINR of the $i$-th scheduled user pair, and the parameters $a$ and $b$ depend on the type of modulation scheme used. $\mathcal{F}_{\gamma_{\text{end}}^{i}}(\cdot)$ is the CDF of $\gamma_{\text{end}}^{i}$.

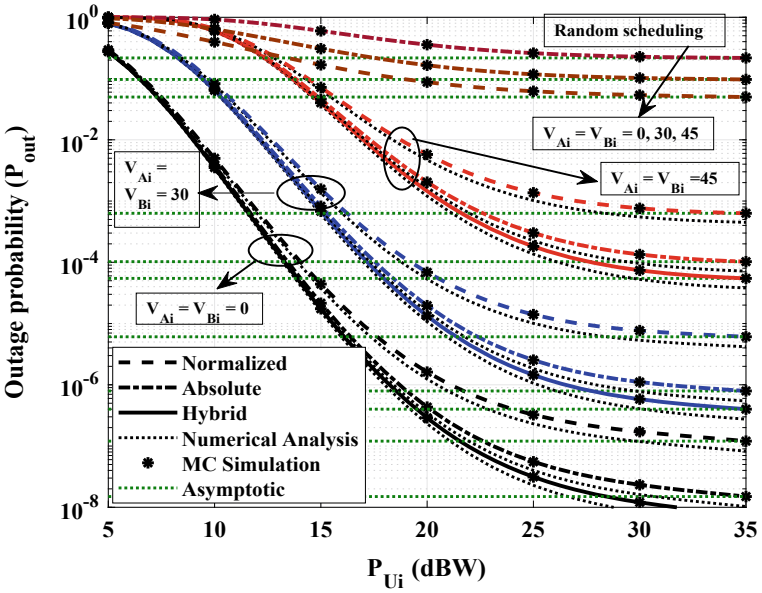



**Fig. 8.3** Outage probability versus user transmit power $P_{Ui}$ for various multi-user scheduling schemes (Raut et al. 2019a)

Further, the outage probability of the FD multi-user two-way relay system is the probability that both the data information from users $A_i$ and $B_i$ are not decoded successfully. Mathematically, outage probability of the considered system is evaluated as

$$P_{out} = \mathcal{F}_{\gamma^i_{\text{end}}}(\gamma_{\text{th}}) = \Pr\left\{\gamma^i_{\text{end}} \leq \gamma_{\text{th}}\right\}. \tag{8.5}$$

As can be seen from Fig. 8.3, the outage performance of the considered system decreases first with power. Later, it saturates to a constant value shown with the asymptotic floors. The reason behind this saturated performance is the fact that higher transmit power strengthens the RSI at FD node which leads to no improvement in the SINR of the system beyond this saturated value. Hence, the outage performance saturates to a constant value achieving the irreducible outage floors. Further, the HYB scheduling scheme outperforms the ABS and NRM channel power-based (CPB) scheduling. The NRM CPB scheduling produces the best system performance in terms of scheduling fairness, and fair utilization of resources as each pair has equal probability to get selected. The error performance of the considered system degrades with transmit power. Contrary to its half-duplex counterpart, it saturates and achieves the irreducible error floors as shown in Fig. 8.4. This happens due to the strengthening of RSI at high transmit power which limits the performance of the system. Moreover, it is observed that dense network with higher number of user pairs can improve the

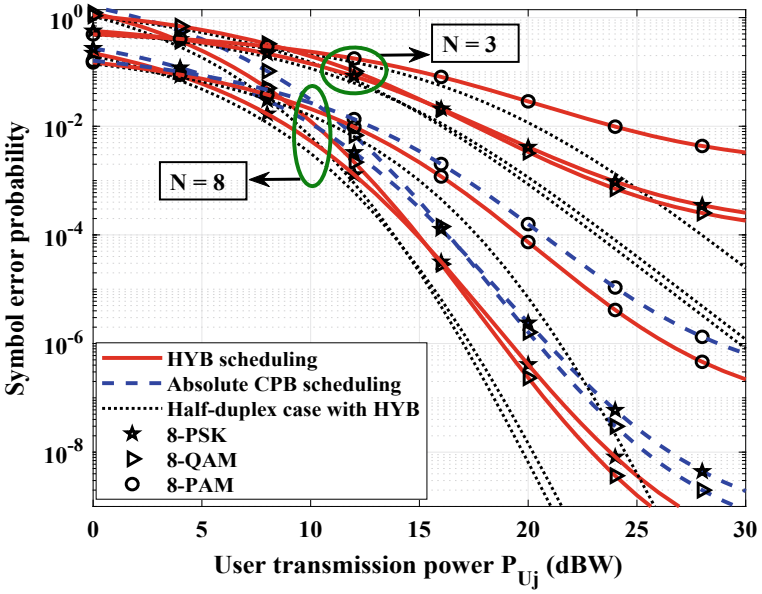



**Fig. 8.4** Error probability of the system versus $P_{Ui}$ with number of user pairs (Raut et al. 2019b)

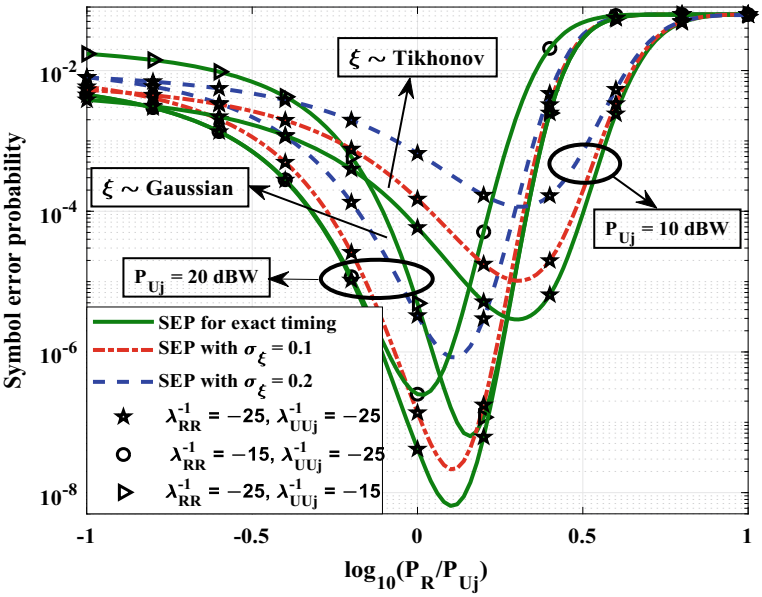



**Fig. 8.5** SEP of the system under impact of synchronization error (Raut et al. 2019b)

error performance significantly. From Fig. 8.5, it is observed that timing errors (for both Tikhonov and Gaussian distributed timing errors) critically affect the system performance compared to the perfect synchronization assumption. Further, for given $P_{Uj}$, the SEP of the system deceases first and then starts increasing beyond $P_R$ due to the strengthening of RSI as $P_R$ exceeds $P_{Uj}$ which results in degraded system performance. Moreover, it is observed that the SEP minima is critically dependent on the RSI, and this minimum value decreases with higher RSI.

## 8.4 Conclusion and Future Directions

This chapter studied the mobile multi-user FD relaying systems from the physical layer perspective. Key challenges for the deployment of FD systems and the modeling of RSI at FD node were examined. Moreover, the works related to the multi-user HD and/or FD, two-way and/or one-way relaying systems are discussed in detail. Furthermore, this chapter highlighted the performances of various multi-user scheduling schemes and concluded with applications of FD relaying for mission-critical communication. With the increasing demand of data rates and higher network density with better connectivity, URLLC, eMBB, and mMTC communication has been emerged as the eye-catching technologies from the future perspective. The performance analysis of modern communication systems based on these communication technologies could of great interest.

## References

Abedin SF, Alam MGR, Kazmi SMA, Tran NH, Niyato D, Hong CS (2019) Resource allocation for ultra-reliable and enhanced mobile broadband IoT applications in Fog network. IEEE Trans Commun 67(1):489–502. https://doi.org/10.1109/TCOMM.2018.2870888

Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surv Tutor 17(4):2347–2376. https://doi.org/10.1109/COMST.2015.2444095

Aryafar E, Khojastepour MA, Sundaresan K, Rangarajan S, Chiang M (2012) MIDU: Enabling MIMO full duplex. In: Proceedings of the 18th annual international conference on mobile computing and networking. Association for Computing Machinery, New York, NY, USA, pp 257–268. https://doi.org/10.1145/2348543.2348576

Bockelmann C, Pratas N, Nikopour H, Au K, Svensson T, Stefanovic C, Popovski P, Dekorsy A (2016) Massive machine-type communications in 5G: physical and MAC-layer solutions. IEEE Commun Mag 54(9):59–65. https://doi.org/10.1109/MCOM.2016.7565189

Boshkovska E, Ng DWK, Zlatanov N, Schober R (2015) Practical non-linear energy harvesting model and resource allocation for swipt systems. IEEE Commun Lett 19(12):2082–2085. https://doi.org/10.1109/LCOMM.2015.2478460

Choi JI, Jain M, Srinivasan K, Levis P, Katti S (2010) Achieving single channel, full duplex wireless communication. In: Proceedings of the sixteenth annual international conference on mobile computing and networking. Association for Computing Machinery, New York, NY, USA, pp 1–12. https://doi.org/10.1145/1859995.1859997

Cui H, Ma M, Song L, Jiao B (2014) Relay selection for two-way full duplex relay networks with amplify-and-forward protocol. IEEE Trans Wireless Commun 13(7):3768–3777. https://doi.org/10.1109/TWC.2014.2322607

Ding H, Ge J, da Costa DB, Guo Y (2011) Outage analysis for multiuser two-way relaying in mixed Rayleigh and Rician fading. IEEE Commun Lett 15(4):410–412. https://doi.org/10.1109/LCOMM.2011.022411.102238

Duarte M, Dick C, Sabharwal A (2012) Experiment-driven characterization of full-duplex wireless systems. IEEE Trans Wireless Commun 11(12):4296–4307. https://doi.org/10.1109/TWC.2012.102612.111278

Duarte M, Sabharwal A, Aggarwal V, Jana R, Ramakrishnan KK, Rice CW, Shankaranarayanan NK (2014) Design and characterization of a full-duplex multiantenna system for wifi networks. IEEE Trans Veh Technol 63(3):1160–177. https://doi.org/10.1109/TVT.2013.2284712

Ericsson (2019) Ericsson mobility report, Ericsson AB. Technical report, Nov 2019. https://www.ericsson.com/en/mobility-report

Everett E, Sahai A, Sabharwal A (2014) Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans Wireless Commun 13(2):680–694. https://doi.org/10.1109/TWC.2013.010214.130226

Gan J, Guo Z, Sandlund K, Liu J, Shen X, Fan R, Liu W, Wang H, Liu G (2012) LTE in-band relay prototype and field measurement. In: Proceedings of the IEEE 75th vehicular technology conference (VTC Spring), pp 1–5. https://doi.org/10.1109/VETECS.2012.6239938

Gu Y, Chen H, Li Y, Vucetic B (2018) Ultra-reliable short-packet communications: half-duplex or full-duplex relaying? IEEE Wireless Commun Lett 7(3):348–351. https://doi.org/10.1109/LWC.2017.2777857

Haghifam M, Makki B, Nasiri-Kenari M, Svensson T, Zorzi M (2016) Wireless-powered relaying with finite block-length codes. arXiv preprint arXiv:1611.05995

Hoymann C, Chen W, Montojo J, Golitschek A, Koutsimanis C, Shen X (2012) Relaying operation in 3GPP LTE: challenges and solutions. IEEE Commun Mag 50(2):156–162. https://doi.org/10.1109/MCOM.2012.6146495

Hu Y, Gross J, Schmeink A (2016a) On the capacity of relaying with finite blocklength. IEEE Trans Veh Technol 65(3):1790–1794. https://doi.org/10.1109/TVT.2015.2406952

Hu Y, Schmeink A, Gross J (2016b) Blocklength-limited performance of relaying under quasi-static Rayleigh channels. IEEE Trans Wireless Commun 15(7):4548–4558. https://doi.org/10.1109/TWC.2016.2542245

Hu Y, Zhu Y, Gursoy MC, Schmeink A (2019) Swipt-enabled relaying in iot networks operating with finite blocklength codes. IEEE J Sel Areas Commun 37(1):74–88. https://doi.org/10.1109/JSAC.2018.2872361

Huang Y, Al-Qahtani F, Zhong C, Wu Q, Wang J, Alnuweiri H (2014) Performance analysis of multiuser multiple antenna relaying networks with co-channel interference and feedback delay. IEEE Trans Commun 62(1):59–73. https://doi.org/10.1109/TCOMM.2013.112213.130390

Jang Y, Lee YH (2010) Performance analysis of user selection for multiuser two-way amplify-and-forward relay. IEEE Commun Lett 14(11):1086–1088. https://doi.org/10.1109/LCOMM.2010.100410.101004

Joung J, Sayed AH (2010) Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems. IEEE Trans Signal Process 58(3):1833–1846. https://doi.org/10.1109/TSP.2009.2038668

Khattabi YM, Matalgah MM (2016) Performance analysis of multiple-relay AF cooperative systems over Rayleigh time-selective fading channels with imperfect channel estimation. IEEE Trans Veh Technol 65(1):427–434. https://doi.org/10.1109/TVT.2015.2399295

Khattabi Y, Matalgah MM (2015) Conventional and best-relay-selection cooperative protocols under nodes-mobility and imperfect-CSI impacts: Ber performance. In: 2015 IEEE wireless communications and networking conference (WCNC), pp 105–110. https://doi.org/10.1109/WCNC.2015.7127453

Klooster J (2009) Icons of invention: the makers of the modern world from Gutenberg to gates. Greenwood Press, Greenwood Icons

Lei X, Fan L, Michalopoulos DS, Fan P, Hu RQ (2013) Outage probability of TDBC protocol in multiuser two-way relay systems with Nakagami-m fading. IEEE Commun Lett 17(3):487–490. https://doi.org/10.1109/LCOMM.2013.020413.122338

Li C, Xia B, Shao S, Chen Z, Tang Y (2017) Multi-user scheduling of the full-duplex enabled two-way relay systems. IEEE Trans Wireless Commun 16(2):1094–1106. https://doi.org/10.1109/TWC.2016.2637340

Liang X, Jin S, Gao X, Wong KK (2013) Outage performance for decode-and-forward two-way relay network with multiple interferers and noisy relay. IEEE Trans Commun 61(2):521–531. https://doi.org/10.1109/TCOMM.2012.122112.110778

Liu P, Kim IM (2010) Performance analysis of bidirectional communication protocols based on decode-and-forward relaying. IEEE Trans Commun 58(9):2683–2696. https://doi.org/10.1109/TCOMM.2010.080310.090292

Liu G, Hou X, Jin J, Wang F, Wang Q, Hao Y, Huang Y, Wang X, Xiao X, Deng A (2017) 3-D-MIMO with massive antennas paves the way to 5G enhanced mobile broadband: from system design to field trials. IEEE J Sel Areas Commun 35(6):1222–1233. https://doi.org/10.1109/JSAC.2017.2687998

Maxwell JC (1865) Viii. A dynamical theory of the electromagnetic field. Philos Trans R Soc Lond 155:459–512. https://doi.org/10.1098/rstl.1865.0008. https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1865.0008

Raut P, Sharma PK, Kothari A (2019) FD multi-user mobile system with timing errors. IEEE Commun Lett 23(12):2394–2397. https://doi.org/10.1109/LCOMM.2019.2939247

Raut P, Kaple T, Sharma PK (2019a) Outage and average rate performances of full-duplex multiuser AF relay systems with time-selective fading. IEEE Syst J 1–9. https://doi.org/10.1109/JSYST.2019.2953949

Raut P, Sharma PK (2018) Full-duplex multi-user two-way relay systems with optimal scheduling. In: Proc. IEEE communication systems and networks (COMSNETS), pp 443–446. https://doi.org/10.1109/COMSNETS.2018.8328239

Raut P, Sharma PK (2019b) Full-duplex multi-user pair scheduling with time-selective fading and imperfect CSI. In: Proceedings of the eleventh national conference on communications, pp 1–6. https://doi.org/10.1109/NCC.2019.8732242

Sharma PK, Garg P (2014) Performance analysis of full duplex decode-and-forward cooperative relaying over Nakagami-m fading channels. Trans Emerg Telecommun Technol 25(9):905–913. https://doi.org/10.1002/ett.2720

Sutton GJ, Zeng J, Liu RP, Ni W, Nguyen DN, Jayawickrama BA, Huang X, Abolhasan M, Zhang Z, Dutkiewicz E, Lv T (2019) Enabling technologies for ultra-reliable and low latency communications: from PHY and MAC layer perspectives. IEEE Commun Surv Tutor 21(3):2488–2524. https://doi.org/10.1109/COMST.2019.2897800

Wang H, Yang Q, Ding Z, Poor HV (2019) Secure short-packet communications for mission-critical IoT applications. IEEE Trans Wireless Commun 18(5):2565–2578. https://doi.org/10.1109/TWC.2019.2904968

Xia B, Yang C, Cao T (2016) Modeling and analysis for cache-enabled networks with dynamic traffic. IEEE Commun Lett 20(12):2506–2509. https://doi.org/10.1109/LCOMM.2016.2607165

Yang K, Cui H, Song L, Li Y (2015) Efficient full-duplex relaying with joint antenna-relay selection and self-interference suppression. IEEE Trans Wireless Commun 14(7):3991–4005. https://doi.org/10.1109/TWC.2015.2415809

Zhang X, Peng M, Ding Z, Wang W (2012) Multi-user scheduling for network coded two-way relay channel in cellular systems. IEEE Trans Wireless Commun 11(7):2542–2551. https://doi.org/10.1109/TWC.2012.051712.111377

Zhang Z, Ma Z, Ding Z, Xiao M, Karagiannidis GK (2016) Full-duplex two-way and one-way relaying: average rate, outage probability, and tradeoffs. IEEE Trans Wireless Commun 15(6):3920–3933. https://doi.org/10.1109/TWC.2016.2531050

# Chapter 9
# PAPR Reduction of Filter Bank Techniques for 5G Communication Systems


**Srinivas Ramavath, Amitkumar V. Jha, Umesh Chandra, Bhargav Appasani, Yasin Kabalci, and Ersan Kabalci**

**Abstract** The stringent criteria of the fifth-generation (5G) radio networks have sparked research on waveforms beyond the fourth-generation mobile radio networks (4G). Recently, filtered multicarrier (FMC) transmission-based techniques have been proposed as the waveforms for 5G systems. These techniques have their own merits and demerits. Some major demerits still remain unsolved in the design of 5G waveforms. The high peak-to-average power ratio (PAPR) of the transmitted FMC signals is one such demerit. The peak-to-average power ratio is the ratio of the peak power to the average power in the waveform. An extensive amount of research has been carried out in literature to reduce the PAPR value for the filtered multicarrier system. This chapter presents a comprehensive summary of the various filter techniques as well as methods for reducing the PAPR. Further, a comparative performance analysis of these techniques is presented to elucidate their merits. Non-orthogonal waveforms such as generalized frequency division multiplexing (GFDM), filter bank multicarrier (FBMC), biorthogonal frequency-division multiplexing (BFDM), and universal

S. Ramavath · A. V. Jha · U. Chandra · B. Appasani (✉)
School of Electronics Engineering,
Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India
e-mail: bhargav.appasanifet@kiit.ac.in

S. Ramavath
e-mail: srinivas2012iitg@gmail.com

A. V. Jha
e-mail: amit.jhafet@kiit.ac.in

U. Chandra
e-mail: umesh.samalfet@kiit.ac.in

Y. Kabalci
Electrical and Electronics Engineering, Faculty of Engineering,
Nigde Omer Halisdemir University, Nigde-Turkey, Turkey
e-mail: yasinkabalci@ohu.edu.tr

E. Kabalci
Faculty of Engineering and Architecture,
Nevsehir Haci Bektas Veli University, Nevsehir, Turkey
e-mail: kabalci@nevsehir.edu.tr



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_9

filtered multicarrier (UFMC) have been considered. The results assist in making the decision of proper waveforms among many to suit different next-generation mobile communication systems.

**Keywords** PAPR · PSD · BER · 5G · OFDM · GFDM · FBMC · UFMC · BFDM

## 9.1 Introduction

Wireless communication systems help to connect people and exchange valuable information among themselves across the globe due to its various advantages such as mobility and portability of the user equipments, which helps to send any type of valuable message to the end user or the group of end users easily at anytime and anywhere. Moreover, wireless communication is vital to many applications such as Internet of things (IoT), wireless sensor networks (WSN), cloud computing, object tracking, and many more. Hence, future generation of communication system is moving fast toward wireless system.

Since the power and spectrum are limited, hence, an efficient use of both power and spectrum is crucial in success of any communication systems. However, wireless communication system is limited by both power and bandwidth. Achieving low bit-error-rate (BER) for high data rate applications with efficient use of power and spectrum is a challenging issue for wireless communication system (Sahin et al. 2014; Schaich et al. 2015; Osseiran et al. 2014). Moreover, low latency, reliability, quality of the received information, interconnection of devices, communicating between devices when they are moving at high velocity, and cost-effectiveness are some of the added requirements which may not be ignored.

In order to transmit data over the channel efficiently various modulation techniques have been proposed by various researchers, out of them, orthogonal frequency-division multiplexing (OFDM) is a promising modulation technique to suit the requirements of the 4G cellular communication systems and its predecessor 3.5G. However, the OFDM may not be employed for 5G applications as a result of following hindrances:

1. Cyclic prefix (CP) used in OFDM reduces the inter-symbol interference (ISI). However, it consumes additional power and spectrum.
2. The requirement of sophisticated synchronization scheme at the receiver since OFDM is extremely sensitive to carrier frequency offset (CFO).
3. High requirement of peak-to-average power ratio (PAPR) makes it power inefficient.
4. It may be difficult to use OFDM for cognitive radio application as in OFDM orthogonality among the subcarriers must be maintained, which may be difficult if the spectrum is discontinuous.
5. Due to large side-lobe of the OFDM inter-carrier interference (ICI) is more severe. Hence, it is required to design the waveforms so that the above limitations experi-

enced by the OFDM can be minimized and suitable for 5G applications (Vasudevan 2010).

Some of the key waveforms that can be used in 5G communication systems are generalized frequency division multiplexing (GFDM), filter bank multicarrier (FBMC), biorthogonal frequency-division multiplexing (BFDM), and universal filtered multicarrier (UFMC). Most of them are non-orthogonal waveforms. Hence, the BER performance may not be better than OFDM. However, less value out-of-band (OOB) radiation as compared to OFDM removes the need of stringent synchronization scheme. But like OFDM, high peak-to-average power ratio is one of the challenging issues of these waveforms, which needs to be addressed. Hence, various techniques have been proposed in literature to reduce PAPR for the contenders of 5G waveforms, which will be discussed and analyzed in this chapter (Aminjavaheri et al. 2015; Vakilian et al. 2013).

The rest of the chapter is organized into the following sections. Some of the key technology and components of 5G communication will be discussed in Sect. 9.2. In Sect. 9.3, different modulation techniques are discussed. Section 9.4 describes various key multicarrier modulation schemes for 5G technologies. The result and analysis are presented in Sect. 9.5 followed by conclusion in Sect. 9.6.

## 9.2 Key Technology and Components

In this section, some of the key requirements of 5G standards for wireless communication systems will be discussed.

The features and the aim of 5G standards are much more different than that of the existing 4G standards. In last few years, an unprecedented growth has been observed in number of users with different mobile devices, which requires more spectrum allocation in order to accommodate all of them. Moreover, high data rate or throughput is essential for the streaming of videos in real-time with low latency. Hence, to avoid the spectrum crunch, we have to move for the higher frequency band and also use the unused spectrum using cognitive radio techniques. Another advantage is that at high frequency, massive MIMO can be implemented easily. Massive MIMO can be used either for improving the throughput or for decreasing the probability of deep fade. Multicarrier modulation schemes (around 28 GHz) used in 5G applications may be unsuitable for millimeter wave (mm Wave) which uses frequency bands at around 60 GHz. Hence, for mm wave applications, single-carrier modulation schemes are suitable rather than multicarrier schemes. However, at high frequency, the cell coverage area are less. Hence, frequent hand-off should be handled carefully for achieving uninterrupted communication. Synchronization and channel estimation are also difficult to estimate properly. So, the modulation scheme (OFDM) that is used in 4G technology, which is orthogonal in nature may not be used in 5G as it requires stringent synchronization scheme at high frequency (Andrews 2014; Vasudevan 2015, 2017; Chen and Zhao 2014).

Some of the key requirements of 5G standards as compared to 4G are:

1. Aggregate data rate will need to be increased roughly up to 10000 times from 4G to 5G.
2. Edge rate is 100Mbps for 5G, which is 100 times than 4G system.
3. Peak data rate will be 10Gbps.
4. Round trip latency for 4G is 15ms, whereas for 5G it will need to 1ms.
5. Energy consumption and cost must be less per-link for 5G as compared to 3G and 4G.
6. Must support large number of divers type of devices (Yunzheng et al. 2015).

## 9.3 5G Technology Waveforms

In this section, some of the well-accepted waveforms for 5G applications will be discussed along with their advantages and disadvantages.

The selection of an acceptable waveform technique has long been a significant investigation concern in the 5G technology design. This work discusses the pros and cons of the various waveforms such as the FBMC, UFMC, and GFDM and draws comparisons among them based on key performance indicators (KPIs), such as PAPR, complexity, latency, filtering level, spectrum performance, spectral coexistence, and power spectral density. These key performance indicators play crucial role in selecting suitable 5G waveforms for advanced applications. The comparison process for the waveform candidate is as illustrated in Fig. 9.1 (Vasudevan 2010; Aminjavaheri et al. 2015; Vakilian et al. 2013; Andrews 2014; Vasudevan 2015).

### *9.3.1 Filter Bank Multicarrier (FBMC)*

In order to remove the side-lobes of the spectrum, FBMC passes the subcarriers through a filter bank similar to OFDM (Renfors et al. 2010). The suppression of the side-lobes makes FBMC band-limited with low OOB radiation when compared to

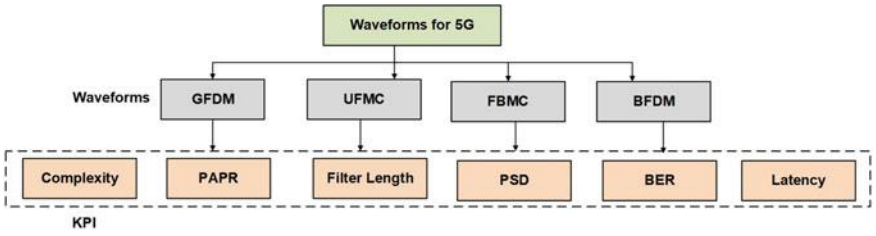



**Fig. 9.1** 5G waveform candidates

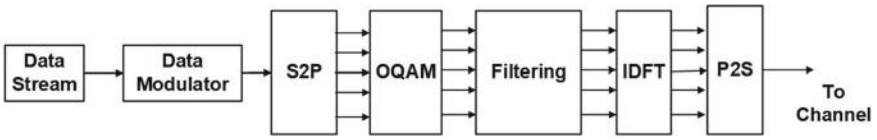



**Fig. 9.2** Block diagram of FBMC modulator

other waveforms and thus, making it spectral efficient. The transceiver for FBMC can be designed using polyphase filters as well as blocks of FFT/IFFT (Shaat et al. 2010). Like GFDM and UFMC, FBMC waveform is also non-orthogonal which can be used in both the uplink and downlink in the wireless communication with lower synchronization and equalization complexity (Payaró et al. 2010). It also supports cognitive radio networks with lower overheads and latency (Schellmann et al. 2014; Behrouz 2011) (Fig. 9.2).

In OFDM, for reducing the PAPR, individual data blocks are considered. However, in FBMC, several overlapping data blocks are considered because of symmetric nature of the FBMC-OQAM signal. Thus, the concept of peak average power ratio for FBMZC-OQAM is different from that used in the OFDM. FBMC-OQAM signal's structural characteristics are taken into consideration based on the description given by the PHYDYAS in burst transmission. There exist two transition phases in FBMC-OQAM burst: initial transition and final transition. During an FBMC-OQAM burst, the signal is very low somewhere in the transition. Further, the peak power of the signal occurs primarily in the middle region. Hence, we would evaluate the value of PAPR in the middle region of FBMC-OQAM waveform.

#### 9.3.1.1 Advantages and Disadvantages

FBMC has highly spectral efficient than others. Side-lobe suppression is done by the use of filter banks and hence presents lower OOB radiation. It is suitable for asynchronous data transmission and also for high mobility. It can be implemented in cognitive radio applications. Better gain can be achieved but it will increase the complexity. However, the major disadvantages of FBMC are the ICI and ISI. PAPR is also high for FBMC (Renfors et al. 2010; Shaat et al. 2010; Payaró et al. 2010; Schellmann et al. 2014; Behrouz 2011). FBMC cannot be used in short burst data transmission.

### *9.3.2 Universal Filtered Multicarrier (UFMC)*

The UFMC basically combines both FBMC and filtered-OFDM (F-OFDM). It will perform in the presence of asynchronous data and also overcomes the demerits of OFDM (Hwang et al. 2009; Samal and Vasudevan 2013a, b; Banelli et al. 2014; Ramavath and Kshetrimayum 2012) and FBMC (Bochechka et al. 2017; Xi et al. 2016). Here, a group of subcarriers are filtered instead of separately filtering the

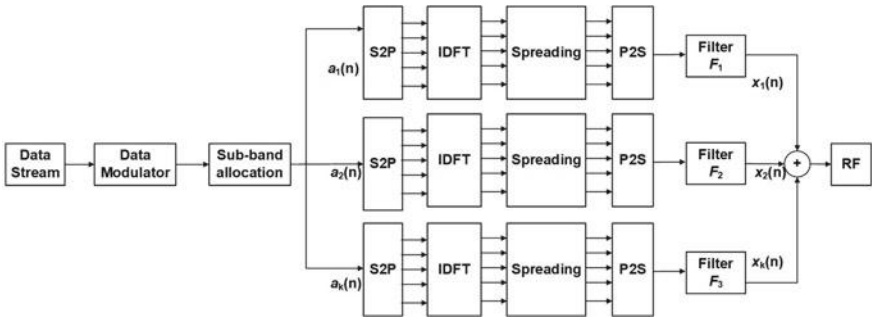



**Fig. 9.3** Block diagram of UFMC modulator

subcarriers. It also has low latency as compared to FBMC and has low OOB radiation than OFDM (Wang et al. 2014; Mukherjee et al. 2015) (Fig. 9.3).

In one sense, OFDM is considered as a special case of UFMC. However, the greatest benefit of choosing UFMC as the key candidate waveforms for 5G is that it can be marginally and specifically applied to a large number of research findings in OFDM. Around the same time, UFMC does not use cyclic prefix to prevent inter-symbol interference, which implicitly increases the system's spectral efficiency and effectively smooths the important spectral capital.

#### 9.3.2.1 Advantages and Disadvantages

High spectral efficiency, low overhead, and low latency are the major advantages of UFMC with respect to other waveforms. For asynchronous data transmission, it performs well. However, an important limitation of UFMC is the presence of higher OOB radiation than FBMC. It is a non-orthogonal waveform, and multi-tap equalizers are required for high data rate. Moreover, high PAPR is a major concern for UFMC (Chen et al. 2014).

### *9.3.3 Generalized Frequency Division Multiplexing (GFDM)*

The GFDM is also a multicarrier scheme based on non-orthogonal techniques and can be implemented using filter banks (Fettweis et al. 2009). Like OFDM it is also strong contender of 5G mobile applications. GFDM spreads the assigned spectrum into multiple unequal sub-bands. Like OFDM, cyclic prefix (CP) is also used with tail biting concept, which helps for the mitigation of ISI (Michailow et al. 2014). Tail biting concept makes the CP shorter in length and improves the efficiency of the spectrum than that of OFDM. Unlike OFDM, here equalization is done in the frequency domain (Matth et al. 2017) (Fig. 9.4).

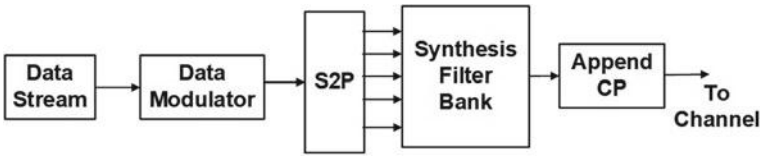



**Fig. 9.4** Block diagram of GFDM modulator

We should remember that interference (different from background noise) carries information about the data. The knowledge of each data symbol is effectively stored by GFDM's ISI and ICI through observations on more than M frequency domain. It offers an opportunity for the data transmission to leverage the frequency selectivity of the multipath fading channel.

#### 9.3.3.1 Advantages and Disadvantages

Like OFDM, it can be configured in a similar manner. The GFDM waveform can be designed so that the receiver complexity can be reduced for equalization and synchronization. Low OOB radiation, low interference, and unequal sub-bands make it suitable for cognitive radio. Use of filtering techniques in GFDM lowers the PAPR as compared to OFDM. Some of the major disadvantages can be sorted as: The GFDM waveform should be designed carefully by using some special types of filters otherwise it may lead to the improve the complexity of the receiver. Synchronization, interference cancellation, and PAPR reduction are some of the major challenges still persists in GFDM. Moreover, MIMO system implementation using GFDM is a difficult task (Fettweis et al. 2009; Michailow et al. 2014; Matth et al. 2017).

### *9.3.4 Biorthogonal Frequency-Division Multiplexing (BFDM)*

The BFDM is a typical CP-OFDM scheme and has the capability of providing lower ICI and lower ISI interference (Wang et al. 2017). These benefits can be achieved by adding additional degrees of freedom to the system. The traditional OFDM techniques use the principle of orthogonality, which states that the prototype filter's impulse response given by $g(t)$ should be orthogonal to its own time-frequency shift version (Kasparick et al. 2014). Unless a CP is used to add a guard time between different symbols, the orthogonality of the signals could be lost due to channel distortions. However, this additional overhead increases the length of the filter. The time dispersion and frequency dispersion are treated separately in dual dispersive channel, which makes the use of this suboptimal in the system (Siclet et al. 2002a). A way to overcome this issue is to observe that to obtain perfect demodulation condition

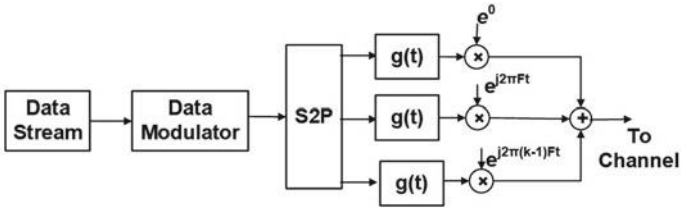



**Fig. 9.5** Block diagram of BFDM modulator

is only sufficient but not necessary. The use of various transmit and receive pulses is precisely the extra degree of freedom that the BFDM modulation scheme enables (Siclet et al. 2002b) (Fig. 9.5).

The use of traditional filter bank for FBMC is another challenge for its implementation. To overcome the issues of traditional filter bank which used use linear convolution, the DFT filter bank based on circular convolution is used in FBMC with PR modulation structure to obtain a biorthogonal frequency division multiple access (BFDMA) system. The designers can independently choose the prototype window of transmitter in BFDMA yielding optimal power spectral density. We are also implementing a function which reduces PAPR in BFDMA. In a nutshell, we obtain a BFDMA scheme satisfying the requirements of 5G communication standards, which are: high PSD, low PAPR, low sensitivity to CFO, and robustness.

#### 9.3.4.1 Advantages and Disadvantages

The designing requirements for transmit prototype is more flexible in BFDM because of biorthogonality of the sending pulses and receiving pulses. The BFDM is also more stable in transmission frequency offsets. The excellent and controllable trade-off between efficiency and BFDM's main advantage over traditional OFDM is due to the time and frequency offsets (Wang et al. 2017; Kasparick et al. 2014; Siclet et al. 2002a, b).

## 9.4 PAPR and Reduction Techniques

The IFFT polyphase network ((IFFT-PPN) is used to add several subcarrier components in multicarrier system. Thus, a transmitted signal in the multicarrier system can have peak amplitude resulting in high PAPR. Generally, negligible values of PAPR can be seen in single carrier system unlike multicarrier system. Further, the high PAPR degrades the system performance by decreasing the signal to quantization noise ratio (SQNR) since it reduces the signal power of transmitter power amplifier. In addition to this, the issues of PAPR become even more critical for uplink design since the mobile terminal is generally equipped with limited battery capacity. Moreover, because of saturation, even the linear amplifiers produce nonlinear distortions.

The nonlinear region of high-power amplifier (HPA) can be characterized by either input back-off (IBO) or output back-off (OBO) as given in Eq. (9.1) (Chen and Zhao 2014). Here, $P_{\text{in}}$ and $P_{\text{out}}$ represent input and output power.

$$\text{IBO} = 10\log_{10}\frac{P_{\text{in}}^{\max}}{P_{\text{in}}}, \quad \text{OBO} = 10\log_{10}\frac{P_{\text{out}}^{\max}}{P_{\text{out}}} \tag{9.1}$$

For the transmitted signal $s(t)$, the value of PAPR can be given by Eq. (9.2).

$$\text{PAPR}_{s(t)} = \frac{\max|s(t)|^2}{E\left[|s(t)|^2\right]} \tag{9.2}$$

Here, $\max|s(t)|^2$ is the maximum signal power and $E\left[|s(t)|^2\right]$ is the average signal power. Some of the well-known PAPR reduction techniques can be found in (Tiwari et al. 2016; Rahmatallah and Mohan 2013). In general, the PAPR reduction methods can be divided into four groups:

(i) Distortion techniques.
(ii) Distortion-less techniques.
(iii) Predistortion technique.
(iv) Spreading technique (Fig. 9.6).

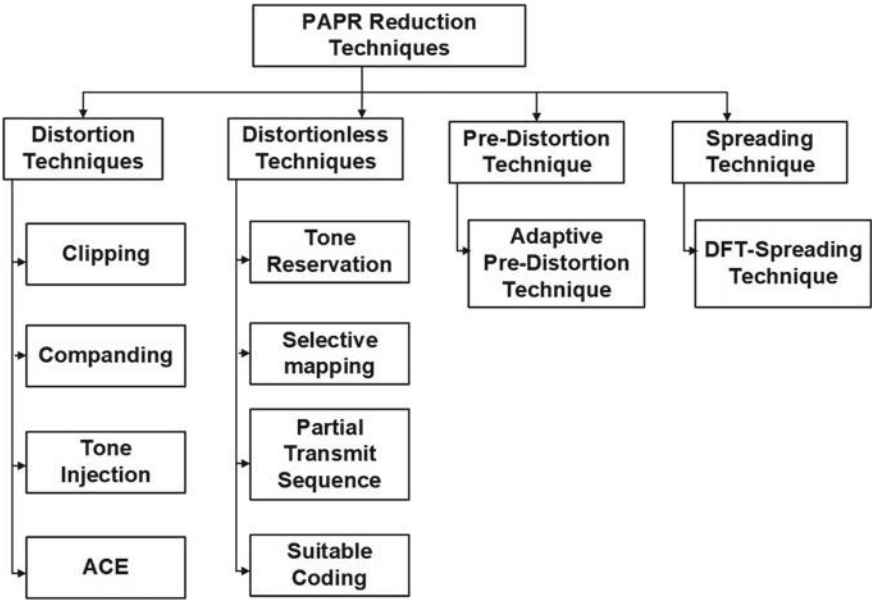



**Fig. 9.6** PAPR reduction taxonomy

### 9.4.1 Distortion Techniques

The techniques below reduce its PAPR while distorting the multicarrier signal:

1. **Clipping:** It is the simplest and most commonly used method, where an amplitude clipper is used to reduce the signal range to a predefined level. The function of amplitude clipping is given by

$$T_{x(n)} = \begin{cases} x\,(n)\,, & |x\,(n)| \leq V_{\mathrm{m}} \\ V_{\mathrm{m}} e^{j\phi}, & |x\,(n)| > V_{\mathrm{m}} \end{cases} \tag{9.3}$$

   where $V_{\mathrm{m}}$ denotes maximum signal amplitude, and $\phi$ denotes the phase of $x(n)$. The non-linearity introduces in the clipped signal results in both in-band as well as outside inconsistencies. The appropriate filtering can be applied at the output of clipper to remove the distortion. But this may force the signal to go beyond the clipping threshold sometime. To avoid this, a mechanism where repeated filtering and clipping are performed can be used. However, it can increase the implementation complexity of the system. Further, a nonlinear compression or clipping around the peaks can be used to decrease the value of PAPR. This techniques may disturb the orthogonality of the subcarriers due to interference between the in-band and out-of-band components.
2. **Companding:** The companding is an invertible nonlinear transformation which can be represented as a monotonously increasing function. The companding consists of compression and expansion of the signal in such a way that low amplitude signal is increased without altering the high amplitude signals. Thus, it reduces the PAPR by increasing the signal power. Among many available techniques, the most widely used companding techniques includes: exponential companding technique and polynomial companding technique. The companded signal can be recovered back at the receiver using inverse nonlinear transformation. This is worth of noting that the complexity of both clipping and companding methods is low at the cost of higher spectral emissions.
3. **Tone Injection:** The tone injection techniques use additional optimized (in amplitude and phase) subcarriers which are superimposed on the carrier carrying information. The additional tones $I$ are generally inserted in between the data subcarriers $S$ to get a signal $\tilde{S} = S + I$ having low PAPR. This process is analogous to the method in which constellation size is increased such that each point in enlarged constellation will map it to corresponding original constellation. In this, due to mapping of data symbols into equivalent constellation points, a unique degree of freedom is obtained. The transmitter reduces the PAPR based on the value of degree of freedom. The mapping of the redundant constellations to the original constellation must be done by the receiver (Ren et al. 2018).
4. **Active Constellation Extension:** If we ignore the outer constellation points, then this method is similar to that of tone injection. In this method, the outer points in the constellation are extended away from the original constellation. On optimizing these constellation points, we get additional degree of freedom. Thus, it can further

reduce the PAPR as compared to the tone injection. This can also decreases the BER due to extension of outer constellation points. However, this technique can only be applied in the modulation scheme having larger constellation size.

### *9.4.2 Distortion-Less Techniques*

In distortion-less techniques, the blocks of the multicarrier symbols are scrambled at the input side and the blocks with minimum PAPR are transmitted. Hence, this reduces the PAPR based on scrambling techniques. As the number of subcarrier increases, spectral efficiency decreases without incurring out-band power issue. The complexity of the system also increases with increasing subcarriers. Furthermore, PAPR cannot be obtained below a specified level. The following techniques are some of the distortion-less techniques that reduce the PAPR:

1. **Tone Reservation:** In this procedure, the PAPR reduction is reserved for a group of subcarriers within the multicarrier symbol. Such reserved subcarriers hold no data; they are modulated in such a way that a low PAPR signal is created by the addition of all subcarriers. The $R$ tones in the $S$ data subcarriers are reserved to get a $\tilde{S} = S + R$ reduced PAPR signal. The method for defining and maximizing the reserved tones is slightly complex. The number of reserved tones decides the effectiveness of this method (Ren et al. 2018).
2. **Selective Mapping:** This technique is based on the methods where the data blocks with least PAPR are transmitted out of many created data blocks. The various phase vectors are multiplied to modulated symbols before IFFT to generate the data blocks. The information pertaining to the phase vectors is also forwarded to the receiver. This additional requirement at the receiver reduces the spectral efficiency of the system. This is worth of noting that the PAPR reduction is directly proportional to the size of phase vectors and IFFTs.
   The basic idea of selective mapping (SLM) appears in Fig. 9.7. The original stream of data is copied into several parallel streams. These are then multiplied with various phases $U$. IFFT is used to receive $U$ individual multicarrier signals and finally, transmission is done for the signal having least PAPR. Here, input data block $a_{n,m}$ is multiplied by the various step sequences of $L$, $p^l = \left[p_0^l, p_1^l, \ldots, p_{N-1}^l\right]^{\mathrm{T}}$, where $p_i^l = e^{j\varphi_i^l}$ and $\varphi_i^l \epsilon\, ([0, 2\pi])$ for $i = 0, 1, \ldots, N-1$ and $l = 1, 2, \ldots L$. This generates a modulated data block $a_{n,m}^l$, which selects the lowest PAPR for transmission (Liu et al. 2019).
3. **Partial Transmit Sequence (PTS):** This technique is based on the method where input data block is divided into a series of several disjoint sub-blocks. Each sub-block is passed through IFFT block and then it is multiplied by a phase vector. Finally, a unique combination is made for transmission such that the combination has minimum PAPR (Ren et al. 2018). The creation of $L$ disjoint sub-blocks by partitioning $N$ input data block are done using Eq. (9.4).

$$a_{n,m} = \left[X^0, X^1, \ldots, X^{L-1}\right]^{\mathrm{T}} \tag{9.4}$$

Here, $a_{n,m}$ denotes consecutively located sub-blocks of equal size. After applying the scrambling to each sub-block, it is multiplied with respective phase factor $b^l = e^{j\phi l}$, $l = 1, 2, \ldots, L$, which is wisely chosen to reduce the PAPR.

4. **Suitable Coding:** This technique works on the principle to select the code words such that it reduces the PAPR of the transmitted signal. The code rate is vital because if the code rate decreases then bandwidth is wasted. Further, it neither produces out-of-band radiation nor distortion. The peak power is obtained when $N$ number of subcarriers having similar phases are superimposed. Thus, mean power is $1/N$ times of maximum power. Since, all codewords cannot lead into reduction of PAPR, hence selection of codewords is vital in achieving low PAPR. Consequently, finding a best codeword is one of the challenging task. One can use various techniques in this such as Reed-Muller code, Hadamard code, Golay complementary series, and M-sequence (Sharifian et al. 2016) (Figs. 9.7 and 9.8).

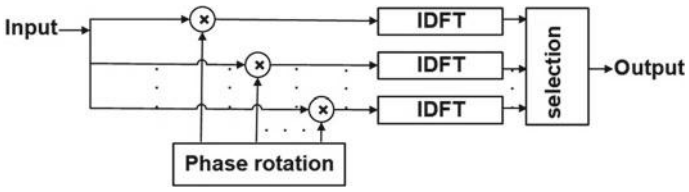



**Fig. 9.7** Structure of selective mapping (SLM) technique

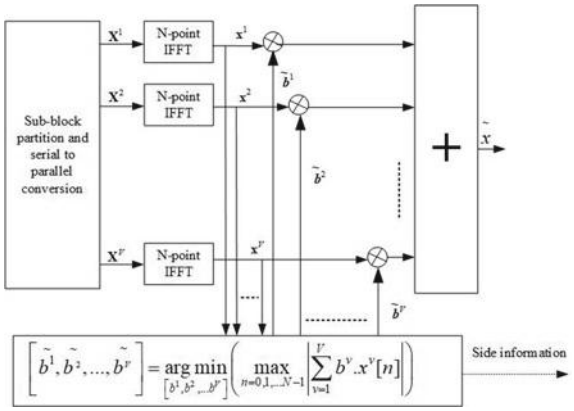



**Fig. 9.8** Structure of partial transmit sequence (PTS) technique

### 9.4.3 Techniques with Predistortion

1. **Adaptive Predistortion Technique:** The adaptive predistortion technique in multicarrier systems will counteract the nonlinear effect of a high-power amplifier. The problem of non-linearity in the HPA can be addressed by achieving dynamically configuration of the input parameter using minimum hardware requirement. The adaptive predistorter's convergence time and the mean square error can be minimized through the use of a broadcast technique and the implementation of appropriate training signals (Sharifian et al. 2016).

### 9.4.4 Techniques with DFT Spreading

1. **DFT-Spreading Technique:** The DFT-spreading method involves the spreading the signals at the input using DFT. The DFT-spreaded signal is sent to the IFFT block. This has an ability to reduce the multicarrier signal PAPR at par with that of single-carrier system. This mechanism has been shown much promising in case of uplink transmission of the mobile system (Das and Tiwari 2015).

## 9.5 Results Analysis and Discussion

This section gives a comparative analysis of several contending waveforms in the context of 5G communication system. The efficiency of the waveforms under consideration is obtained using Monte Carlo simulations. Table 9.1 presents the parameters that were considered for these waveforms.

Many strategies used for PAPR reduction improve the mean power of the transmitted signal. The signal must lie in the linear region of power amplifier (PA). Based on the signal characteristics, the linear region must be extended. Otherwise, it increases the distortion as well as BER if the signal crosses the linear region (Fig. 9.9).

It has been observed that complexity of techniques has better performance in terms of PAPR. But, at the same time, the complex techniques are costly in terms of hardware requirements, power, and implementation cost. Thus, for a practically

**Table 9.1** PAPR, PSD, and BER with different waveform candidates

| 5G waveforms | PAPR (dB) | PSD (dB) | BER = $10^{-2}$ (dB) |
|---|---|---|---|
| UFMC | 7.80 | −93 | 34 |
| FBMC | 11.10 | −98 | 17 |
| GFDM | 9.40 | −48 | 23 |
| BFMC | 9.10 | −70 | 28 |

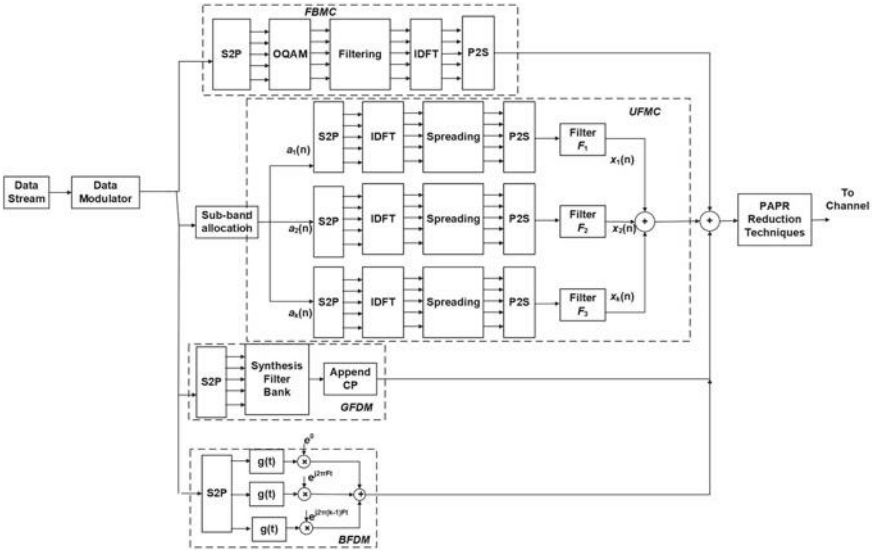



**Fig. 9.9** Block diagram of 5G waveforms modulator

feasible system, the hardware complexity must be as low as possible. Further, the time required to process the data should also be as low as possible.

Generally, side information is transmitted in order to reduce the PAPR in some of the techniques. This requires additional bandwidth which also decreases the data rate of the signal. On the other hand, PAPR is also controlled based non-informative data which further increases the bandwidth requirement. Moreover, several other factors must be taken into account in order to reduce the PAPR such as device non-linearity, cost, and performance.

The results can be mainly summarized using Figs. 9.10, 9.11, and 9.12. In Fig. 9.12, the experimental BERs from the different candidate modulations are shown. The biggest observation is that the obtained results match quite-precisely the theoretical performance of the different modulations. In particular, it is observed that FBMC gives minimum BER among all waveforms throughout the whole range of power received. At the same time, GFDM is the worst, while OFDM and UFMC are intermediate with OFDM performing slightly better than UFMC. The fact that GFDM is the one with worst results is not surprising. It is well known that the main advantage of GFDM is its flexibility and that it needs to sacrifice BER.

For all waveforms, the complementary cumulative distribution function (CCDF) of the peak average power ratio is as shown in Fig. 9.10. Here, $10^5$ frames were transmitted such that each frame is having four signal blocks. We have assumed threshold of 0.1% for PAPR comparison such that ($\text{PrPAPR} > \text{PAPRo} = 10^{-3}$). This has been observed from the simulation that UFMC, BFDM, and GFDM have the lowest PAPR of 3.5, 2.0, and 1.5 dB, respectively, which is less than FBMC. UFMC,

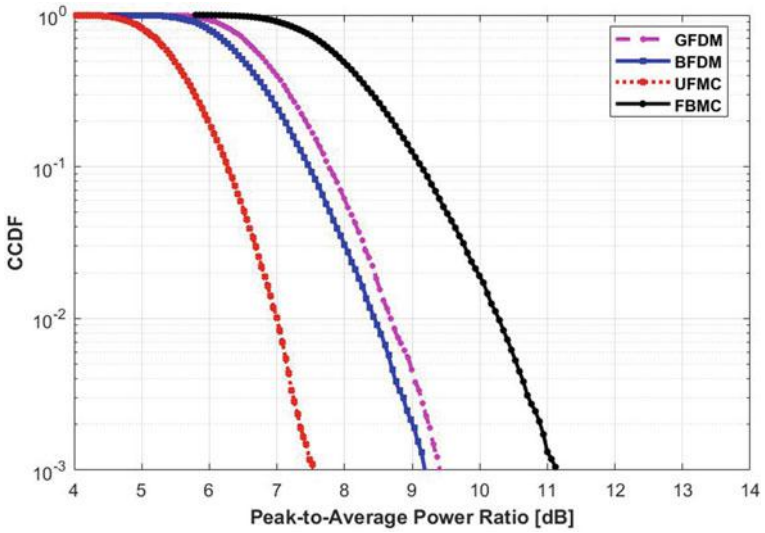



**Fig. 9.10** Comparison of PAPR

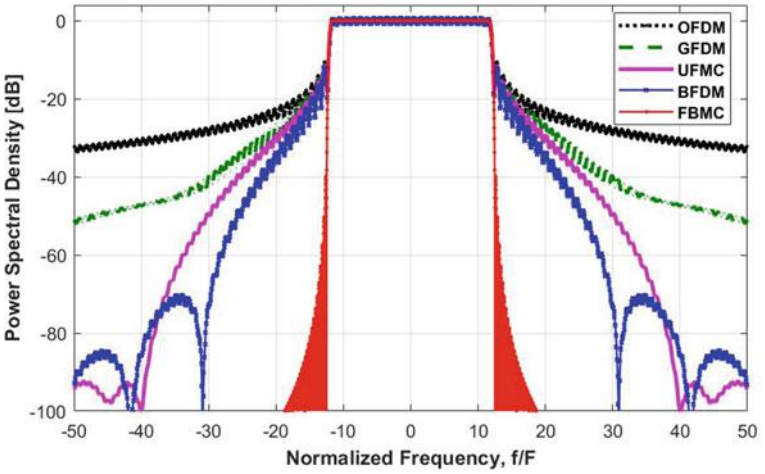



**Fig. 9.11** Comparison of out-of-band leakage

BFDM, and GFDM have equivalent PAPRs which are about 1.5 dB smaller than FBMC.

The power amplifier restricts the maximum peak of the 5G signal, which increase in in-band and out-of-band distortions. The second induces an unnecessary rise in the strength of the side-lobes of the PSD of the 5G signal. This effect is called spectral spread. It is also referred to as the spectral regrowth. From Fig. 9.11, it can be inferred that higher PA non-linearity results into lower IBO as well as higher

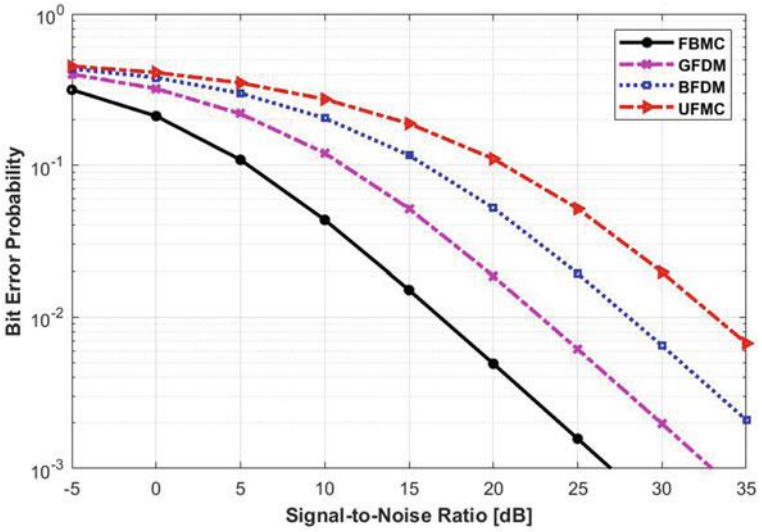



**Fig. 9.12** Comparison of BER for ideal channel

spectral distribution. Spectral regrowth results in higher interference among the sub-bands of the 5G signal, unless there is also improved frequency isolation between neighboring subcarriers to ensure orthogonality. Nevertheless, the drawback of this approach is that it reduces the spectral efficiency.

The plots for unilateral PSD are shown in Fig. 9.11 for the waveforms under considered. To observe OoB characteristics of the waveform, a 20 MHz system with 256 subcarriers is considered such that 128 subcarriers are switched off (64 on each edge). For each waveform considered, the PSD is averaged over $10^4$ transmitted symbols. It can be seen that the FBMC has the minimum attenuation stop band and the narrowest transition band. The GFDM, BFDM, and UFMC show an attenuation of 48, 70, and 93 dB higher as compared to the OFDM respectively. Further, the narrowest transition band is obtained for GFDM whereas the widest transition band is obtained for UFMC. Furthermore, GFDM gives 5 dB more attenuation stop band, but the smaller transition band when compared to OFDM.

To achieve a particular BER, one can represent the output of a modulation technique using SNR. The CCDF is the main focus of PAPR reduction strategies. There exists a trade-off between CCDF and the BER. The PA's high peaks can cause a major in-band distortion that results in higher BERs. Certain strategies may also involve the transmission of side information. In case of error in reception of side information, the entire symbol in the 5G waveform is recovered with error. This error degrades the performance of the BER drastically.

Figure 9.12 provides a comparison of BER versus $\frac{E_b}{N_0}$ of considered 5G waveforms modulation. The GFDM is more spectral efficient when compared to OFDM as it

uses systems with less CP length. It can be inferred that the UFMC outperforms its counterparts i.e., GFDM, FBMC, and OFDM vanilla types.

## 9.6 Conclusion

The LTE and its LTE-A development were regulated through the 3GPP. A radical move is needed by looking at the prospective of wireless communication and forecasting its challenges. This move from 4G to 5G involves a decline in performance backwards. Current 5G waveforms and its shared frame structure are important building blocks to achieve those objectives. Beginning with the main drivers of 5G systems, spectral and temporal heterogeneity, intermittent traffic, real-time software and hardware constraints, this chapter discusses the fundamentals of the 5G system. In the light of the Gabor principle, general waveform considerations are discussed. Then it describes in detail the four waveform candidates GFDM, UFMC, FBMC, and BFDM, summarizing the available performance. Such waveforms help and require the structure of the unified frame. The 5G frame structure unleashes the potential of multicarrier waveforms for future applications. Performance tests, including optimized and/or appropriate waveform parameters, are given for the candidate transceiver approaches. These new signal formats require standardization because on both ends of the link they need to be known. The telecommunications industry as a whole needs to have pragmatic view on 5G system and must reach to some conclusion in order to standardization of 5G.

## References

Aminjavaheri A, Farhang A, RezazadehReyhani A, Farhang Boroujeny B (2015) Impact of timing and frequency offsets on multicarrier waveform candidates for 5G. In: Proceedings of IEEE signal process. Signal process. Edu. Workshop (SP/SPE), pp 178–183

Andrews JG et al (2014) What will 5G be? IEEE J Sel Areas Commun 32(6):1065–1082

Banelli P et al (2014) Modulation formats and waveforms for 5G networks: who will be the heir of OFDM?: an overview of alternative modulation schemes for improved spectral efficiency. IEEE Sign Process Mag 31(6):80–93

Bochechka G et al (2017) Comparative analysis of UFMC technology in 5G networks. In: 2017 international Siberian conference on control and communications (SIBCON). IEEE, pp 1–6

Chen Y, Schaich F, Wild T. Multiple access and waveforms for 5G: IDMA and universal filtered multi-carrier. In: 2014 IEEE 79th vehicular technology conference (VTC Spring), 2014 IEEE, pp 1–5

Chen S, Zhao J (2014) The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication. IEEE Commun Mag 52(5):36–43

Das SS, Tiwari S (2015) Discrete Fourier transform spreading-based generalised frequency division multiplexing. Electron Lett 51(10):789-791

Farhang-Boroujeny B (2011) OFDM versus filter bank multicarrier. IEEE Sign Process Mag 28(3):92–112

Fettweis G, Krondorf M, Bittner S (2009) GFDM-generalized frequency division multiplexing. In: VTC Spring 2009-IEEE 69th vehicular technology conference 2009 . IEEE, pp 1–4

Hwang T et al (2009) OFDM and its wireless applications: a survey. IEEE Trans Veh Technol 58(4):1673–1694

Kasparick M, Wunder G, Jung P, et al (2014) Bi-orthogonal waveforms for 5G random access with short message support. In: Proceedings of European wireless 2014; 20th European wireless conference, pp 1–6

Liu Z, Xiao P, Hu S (2019) Low-PAPR preamble design for FBMC systems. IEEE Trans Veh Technol 68(8):7869–76

Matth M et al (2017) Generalized frequency division multiplexing: a flexible multi-carrier waveform for 5G. In: 5G mobile communications. Springer, New York, pp 223–259

Michailow N et al (2014) Generalized frequency division multiplexing for 5th generation cellular networks. IEEE Trans Commun 62(9):3045–61

Mukherjee M et al (2015) Reduced out-of-band radiation-based filter optimization for UFMC systems in 5G. In: 2015 international wireless communications and mobile computing conference (IWCMC)

Osseiran A et al (2014) Scenarios for 5G mobile and wireless communications: the vision of the METIS project. IEEE Commun Mag 52(5):26–35

Payaró M, Pascual-Iserte A, Nájar M (2010) Performance comparison between FBMC and OFDM in MIMO systems under channel uncertainty. In: 2010 European wireless conference (EW). Lucca 2010, pp 1023–1030

Rahmatallah Y, Mohan S (2013) Peak-to-average power ratio reduction in OFDM systems: a survey and taxonomy. IEEE Commun Surv Tutor 15(4):1567–1592

Ramavath S, Kshetrimayum RS (2012) Analytical calculations of CCDF for some common PAPR reduction techniques in OFDM systems. In: 2012 international conference on communications, devices and intelligent systems (CODIS), Kolkata, 2012, pp 393–396

Ren S, Deng H, Qian X, Liu Y (2018) Sparse PTS scheme based on TR schemes for PAPR reduction in FBMC-OQAM systems. IET Commun 12(14):1722-1727

Renfors M et al (2010) Filter banks for next generation multicarrier wireless communications. EURASIP J Adv Sign Process, pp 1–2. (Article Id. 314193)

Sahin A, Guvenc I, Arslan H (2014) A survey on multicarrier communications: Prototype filters, lattice structures, and implementation aspects. IEEE Commun Surv Tuts 16(3):1312–1338. (3rd Quart)

Samal UC, Vasudevan K (2013) Bandwidth efficient turbo coded OFDM systems. In: 13th International Conference on ITS telecommunications (ITST), 2013. IEEE, pp 493–498

Samal UC, Vasudevan K (2013) Preamble-based timing synchronization for OFDM systems. In: 2013 IEEE 3rd international advance computing conference (IACC). IEEE, pp 313–318

Schaich F et al (2015) FANTASTIC-5G: 5G-PPP project on 5G air interface below 6 GHz. In: Proceedings of European conference on networks and communications, pp 1–5

Schellmann M et al (2014) FBMC-based air interface for 5G mobile: challenges and proposed solutions. In: 2014 9th international conference on cognitive radio oriented wireless networks and communications (CROWNCOM), Oulu, 2014, pp 102–107

Shaat M, Bader F (2010) Computationally efficient power allocation algorithm in multicarrier-based cognitive radio networks: OFDM and FBMC systems. EURASIP J Adv Sign Process 1:1–13. (Article Id. 528378)

Sharifian Z, Omidi MJ, Saeedi-Sourck H, Farhang A (2016) Linear precoding for PAPR reduction of GFDMA. IEEE Wirel Commun Lett 5(5):520–523

Siclet C, Siohan P, Pinchon D (2002a) Analysis and design of OFDM/QAM and BFDM/QAM oversampled orthogonal and biorthogonal multicarrier modulations. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP), Orlando, USA, May 2002

Siclet C, Siohan P, Pinchon D (2002b) Oversampled orthogonal and biorthogonal multicarrier modulations with perfect reconstruction. In2002 14th International Conference on Digital Signal Processing Proceeding, vol 2. DSP 2002 (Cat. No.02TH8628), Santorini, Greece, pp 647–650

Tiwari S, Chatterjee S, Das SS (2016) Comparative analysis of waveforms for fifth generation mobile networks. In: 2016 IEEE international conference on advanced networks and telecommunications systems (ANTS), Bangalore, pp 1–6

Vakilian V, Wild T, Schaich F, Ten Brink S, Frigon JF (2013) Universal filtered multi-carrier technique for wireless systems beyond LTE. In: Proceedings of IEEE globecom workshops (GC Wkshps), pp 223–228

Vasudevan K (2010) Digital communications and signal processing, 2nd edn. Universities Press (India) Private Limited, Chennai

Vasudevan K (2017) Near capacity signaling over fading channels using coherent turbo coded OFDM and massive MIMO. (Online). Available: https://arxiv.org/abs/1711.10104

Vasudevan K (2015) Coherent detection of turbo-coded OFDM signals transmitted through frequency selective Rayleigh fading channels with receiver diversity and increased throughput. Wirel Pers Commun 82(3):1623–1642

Wang X et al (2014) Universal filtered multi-carrier with leakage-based filter optimization. In: Proceedings of 20th European wireless conference 2014. VDE, pp 1–5

Wang G, Shao K, Zhuang L (2017) Biorthogonal frequency division multiple access. IET Commun 11(6):773–82

Xi Siyu et al (2016) DFT-spread universal filtered multi-carrier for 5G. In: 2nd IEEE international conference on computer and communications (ICCC), 2016. IEEE, pp 1591–1596

Yunzheng T, Long L, Shang L, Zhi Z (2015) A survey: several technologies of non-orthogonal transmission for 5G. China Commun 12(10):1–15

Zhang X, Jia M, Chen L, Ma J, Qiu J (2015) Filtered-OFDM-enabler for flexible waveform in the 5th generation cellular networks. In: Proceedings of IEEE global communications conference (GLOBECOM)

# Chapter 10
# Optimization of Resources to Minimize Power Dissipation in 5G Wireless Networks


**Jyotsna Rani and Ganesh Prasad**

**Abstract** In today's modern communications, with the evolution of various applications, the demand for data rate is increasing exponentially at the cost of huge consumption of available resources. It has been recorded that the communication networks dissipate nearly 1% of the worldwide total power consumption which results in millions of tons of $CO_2$ emission due to their production, thereby causing various environmental health hazards. The optimal utilization of available resources that can balance the present coexisting problem without any compromise on the high throughput demand paves the way for the next-generation green 5G wireless networks. In this chapter, we study the minimization of total power consumption while satisfying the desired coverage of the user equipments (UEs) to provide the minimum throughput over the network. In this regard, the deployment of base stations (BSs), their number, and transmit power are optimized in two scenarios (i) when the number of UEs is large in 5G wireless network and (ii) when a moderate number of UEs are distributed over the field.

**Keywords** Binomial point process (BPP) · Deployment of base stations (BSs) · Coverage probability · Optimization of resources · Green communication · Joint optimization

## 10.1 Introduction and Background

Today, the digital domain of all applications has become an integral part of our daily life that results in the growth of data rate demand by 10 times in every 5 years. In order to fulfill it, the architecture is also increasing at the same rate in the next-generation

J. Rani · G. Prasad (✉)
Department of Electronics and Communication Engineering,
National Institute of Technology Silchar, Silchar, India
e-mail: gpkeshri@ece.nits.ac.in

J. Rani
e-mail: jyotsnarani0009@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_10

wireless network that consumes around 1% of the worldwide total electricity consumption, and it requires a huge production of electricity which emits 130 million tons of $CO_2$ per year Fettweis and Zimmermann (2008). Therefore, we need a reasonable strategy for optimal utilization of available resources to combat the coexistence problem while satisfying the desired coverage of UEs.

In this chapter, we investigate the optimal deployment of network architecture as well as the optimization of transmit power to minimize overall power consumption under the given constraint of desired coverage that gives the minimum throughput over the network. From the state of the art, authors in Andrews et al. (2011); Elsawy et al. (2013); Srinivasa and Haenggi (2010) have deployed the network nodes randomly with a given distribution. It has been described in Andrews et al. (2011) that deployment based on homogeneous Poisson point process (HPPP) is more tractable and satisfies the practical aspects than the conventional strategy where the nodes are uniformly placed on a grid. In Elsawy et al. (2013), for different types of network and MAC layer, various point processes like HPPP, binomial point process (BPP), hard core point process (HCPP), and Poisson cluster process (PCP) are used to measure the system performances. Further, it is analyzed in Srinivasa and Haenggi (2010) that BPP is a more realistic and tractable model than HPPP.

Based on the required hours for service of the BSs, the power dissipation over the network can be reduced by dynamically turning them on/off Wu et al. (2015). Different frameworks used to determine the sleeping mode are discoursed in Marsan et al. (2009); Oh and Krishnamachari (2010). In Marsan et al. (2009), the on/off the BSs is determined by the traffic profile; besides, traffic profile and BSs density are used for deciding the sleeping mode in Oh and Krishnamachari (2010). Energy-saving algorithm based on on/off of the BSs that achieve saving up to 80% is discoursed in Oh et al. (2013). The trade-off between energy efficiency and spectral efficiency is investigated in Peng et al. (2013) over a switching-based network; thereafter, a power control technique is described to optimize the trade-off. Authors in Zhou et al. (2009) describe the centralized and decentralized power reduction mechanism while satisfying the outage constraint. The blockage in services and delay in delivery of the data due to switching operation are analyzed in Jie et al. (2012); Wu et al. (2013).

Recently, the optimization of deployment and transmit power have been investigated in Perabathini et al. (2014); Sarkar et al. (2014); Verenzuela et al. (2016); González-Brevis (2011). In Perabathini et al. (2014) and Sarkar et al. (2014), the area power consumption of the network is minimized by optimizing the density of the base stations (BSs) under the constraint of users' coverage and data rate demand. Authors in Verenzuela et al. (2016) optimize the multiple parameters like transmit power, density of BSs, number of antennas, and users served per BS to reduce the overall power dissipation while satisfying a given data rate demand. A different strategy for energy saving is described in González-Brevis (2011) where the number of BSs and their location are optimized. However, from the recent works, the optimization of transmit power, number of nodes and their location is further required to be explored in two scenarios that are: (i) when large UEs are associated in a 5G wireless network and (ii) when moderate UEs are distributed in a rural area.

## 10.2 System Assumptions

Here, we describe the network topology and channel models along with system assumptions considered for the proposed framework.

### *10.2.1 Network Topology*

We consider a circular and a square field where the $N_u$ UEs are distributed using BPP and $N_b$ BSs are deterministically deployed as shown in Fig. 10.1. The size of the cells over the field is determined by Dirichlet regions using Voronoi tessellation Stojmenovic et al. (2006) as depicted in Fig. 10.2. Here, the boundary of a cell is determined by the bisector line between the two BSs (denoted by $\times$). Based on it, the generated cells in Fig. 10.1 are square or rectangular in shape over the square field as the BSs are located at regular distance along the length and width. Similarly, the generated cells over the circular field have the shape of arc or triangular as the BSs are placed at a regular intervals in the angular direction from the center. A UE lying in a particular cell is assumed to be completely associated with the underlying BS in the cell. The interference over the downlink communication is assumed to be absent as BSs use orthogonal multiple-access techniques Mhiri et al. (2013).

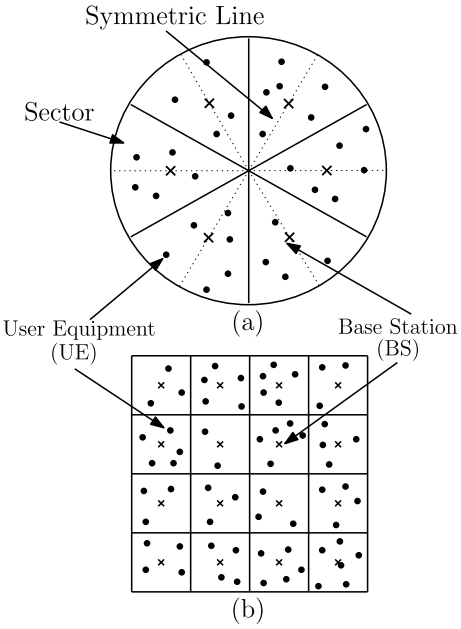



**Fig. 10.1** Randomly distributed UEs by BPP lying under the coverage of deterministically deployed BSs in **a** a circular and **b** a square field

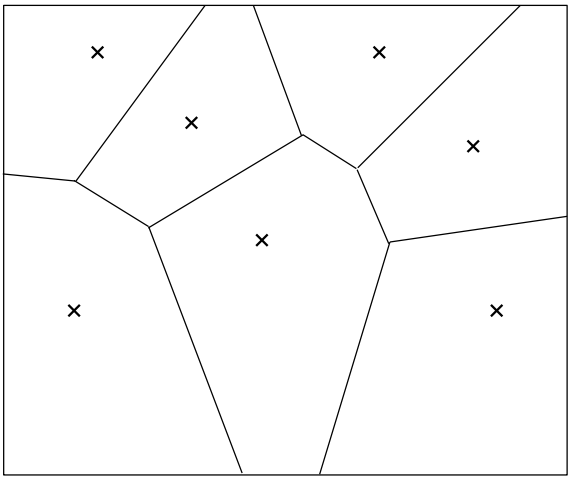


**Fig. 10.2** Cells' generations using Voronoi tessellation about the BSs (represented by $\times$) in a bounded square field

### *10.2.2 Channel Model*

We assume that the channel is flat where the channel power coefficient $h$ between an UE and a BS is a variate with exponential distribution. Therefore, $h \sim \exp(\mu)$ with mean $\mu = 1$. The signal-to-noise ratio (SNR) $\gamma$ received at a UE from a BS at a distance $r$ is given by $\gamma = \frac{P_t h r^{-\alpha}}{\sigma^2}$, where $P_t$ is the transmit power from the BS, $\alpha$ is the path loss exponent, and $\sigma^2$ is the noise power of the additive noise at the receiver Rappaport (1996). In the framework, we assume that the network is homogeneous as all the BSs transmit the same power $P_t$.

### *10.2.3 Power Dissipation in a BS*

While a downlink transmission, the power dissipation $P_D$ in a BS is given as:

$$P_D = \epsilon_1 P_t + \epsilon_2, \tag{10.1}$$

where $\epsilon_1$ accounts for the scaling of the power transmitted by the BS and $\epsilon_2$ is power dissipation due to signal processing, power supply, and battery backup. Total power consumption over $N_b$ BSs deployed over the field is $N_b P_D$ that is required to be minimized by optimizing the available resources.

## 10.3 Problem Formulation and Optimization of Transmit Power

In this section, we study the coverage probability of a UE by a BS. Thereafter, coverage probability of the farthest UE from the BS is described in two cases: (i) when a single BS and (ii) when multiple BSs are deployed over the field. Lastly, a constrained optimization problem is formulated to minimize the total power consumption $N_{\mathrm{b}}P_{\mathrm{D}}$.

### *10.3.1 Coverage Probability of an UE*

A UE is said to be in coverage from its BS when the SNR $\gamma \geq \mathcal{T}$, where $\mathcal{T}$ is threshold SNR at the UE to successfully detect the received signal. If the UEs are distributed using BPP, coverage probability of a UE from the BS is given as:

$$P_{\mathrm{cov}} = \mathbb{E}_r\left[\Pr\left(\frac{P_{\mathrm{thr}}^{-\alpha}}{\sigma^2} \geq \mathcal{T}\right)\right] = \mathbb{E}_r\left[\Pr\left(h \geq \frac{\mathcal{T}\sigma^2 r^{\alpha}}{P_{\mathrm{t}}}\right)\right]$$
$$= \int_r \exp\left(-\frac{\mathcal{T}\sigma^2 r^{\alpha}}{P_{\mathrm{t}}}\right) f\,(r, x, y)\,\mathrm{d}r, \tag{10.2}$$

where $\mathbb{E}_r[\,\cdot\,]$ is the expectation in distance $r$ of the UE from the BS and $f\,(r, x, y)$ is probability density function (PDF) of $r$ when the BS is located at $(x, y)$.

### *10.3.2 Coverage Probability of the Farthest UE*

In order to assure the coverage of all UEs, we need to investigate the coverage of the farthest UE from the BS in a cell. This can be analyzed in two cases when a cell has: (i) a single BS and (ii) multiple BSs.

#### 10.3.2.1 Case (i)

If $N_{\mathrm{u}}$ UEs are distributed using BPP over a cell consisting a single BS, then the cumulative density function (CDF) $F_{\mathrm{far}}(r, x, y)$ and PDF $f_{\mathrm{far}}(r, x, y)$ of the farthest UE distance from the BS are given as Thompson (1956):

$$F_{\mathrm{far}}(r, x, y) = [F(r, x, y)]^{N_{\mathrm{u}}};\ f_{\mathrm{far}}(r, x, y) = N_{\mathrm{u}}[F(r, x, y)]^{N_{\mathrm{u}}-1} f\,(r, x, y) \tag{10.3}$$

Further, using (10.2), the coverage probability of farthest UE is given as:

$$P_{\mathrm{cov}}^{\mathrm{far}} = \int_r \exp\left(-\frac{\mathcal{T}\sigma^2 r^{\alpha}}{P_{\mathrm{t}}}\right) f_{\mathrm{far}}\,(r, x, y)\,\mathrm{d}r. \tag{10.4}$$

#### 10.3.2.2 Case (ii)

For $N_u$ UEs distributed using BPP over a cell consisting $N_b$ ($< N_u$) BSs, the CDF and PDF of farthest UE distance from the $i$th BS are given by (10.5) and (10.6), respectively.

$$F^i_{\text{far}}(r, \mathbf{x}, \mathbf{y}) = \sum_{k=0}^{N_u} \binom{N_u}{k} \left(\frac{A_i}{W}\right)^k \left(1 - \frac{A_i}{W}\right)^{N_u - k} \left[F_i(r, \mathbf{x}, \mathbf{y})\right]^k, \tag{10.5}$$

$$f^i_{\text{far}}(r, \mathbf{x}, \mathbf{y}) = \sum_{k=0}^{N_u} \binom{N_u}{k} \left(\frac{A_i}{W}\right)^k \left(1 - \frac{A_i}{W}\right)^{N_u - k} k\left[F_i(r, \mathbf{x}, \mathbf{y})\right]^{k-1} f_i(r, \mathbf{x}, \mathbf{y}), \tag{10.6}$$

where $W$ is the total area of the field, $A_i$ is the area of the $i$th cell, $(\mathbf{x}, \mathbf{y}) = \{(x_i, y_i);\ i \in \{1, 2, \ldots, N_b\}\}$ is the coordinate of locations of the $N_b$ BSs, and $F_i(r, \mathbf{x}, \mathbf{y})$ and $f_i(r, \mathbf{x}, \mathbf{y})$ are CDF and PDF of UE's distance from the BS in $i$th cell, respectively. Note that the $F_i(r, \mathbf{x}, \mathbf{y})$ and $f_i(r, \mathbf{x}, \mathbf{y})$ depend on the location of all BSs $(\mathbf{x}, \mathbf{y})$ because the shape of a cell is determined using Voronoi tessellation as described in Sect. 10.2.1. Using (10.2) and (10.6), the coverage probability of farthest UE from the BS in $i$th cell is expressed as:

$$P^{\text{far}}_{\text{cov},i} = \int_r \exp\left(-\frac{\mathcal{T}\sigma^2 r^\alpha}{P_t}\right) f^i_{\text{far}}(r, \mathbf{x}, \mathbf{y})\, dr \tag{10.7}$$

### *10.3.3 Optimization Formulation*

In this chapter, our aim is to minimize the total power dissipation over the network while satisfying the given coverage constraint. The corresponding optimization problem (P0) can be formulated as:

$$\begin{aligned}
&\text{(P0)}: \underset{N_b, P_t, \mathbf{x}, \mathbf{y}}{\text{minimize}}\ \ N_b[\epsilon_1 P_t + \epsilon_2] \\
&\text{subject to}\quad C1: P^{\text{far}}_{\text{cov},i} \geq 1 - \epsilon,\ \ \forall i \in \{1, 2, \ldots, N_b\}, \\
&\qquad\qquad C2: 0 \leq N_b \leq N_{\max}, \\
&\qquad\qquad C3: 0 \leq P_t \leq P_{\max}, \\
&\qquad\qquad C4: 0 \leq x_i \leq x_{\max},\ \ \forall i \in \{1, 2, \ldots, N_b\}, \\
&\qquad\qquad C5: 0 \leq y_i \leq y_{\max},\ \ \forall i \in \{1, 2, \ldots, N_b\},
\end{aligned}$$

where $1 - \epsilon$ (for $0 \leq \epsilon \leq 1$) is the threshold coverage probability of the farthest UE that is satisfied under the constraint $C1$. $C2 - C5$ are convex constraints that represent the bounds on $N_b$, $P_t$, $\mathbf{x}$, and $\mathbf{y}$, respectively. As the objective function of

(P0) has integer variable $N_\mathrm{b}$, the problem is nonconvex in the underlying variables. To obtain its optimal solution, we split the combinatorial problem into individual optimization of underlying variables $P_\mathrm{t}$, $(\mathbf{x}, \mathbf{y})$, and $N_\mathrm{b}$ in the following discussions.

### *10.3.4 Optimization of Transmit Power*

The constraint $C1$ is not tractable in the present form. To investigate, we simplify it using a tight approximation as follows. For satisfying the constraint above 90% $\left(P^{\mathrm{far}}_{\mathrm{cov},i} \geq 0.9,\ \forall i\right)$, the argument $\frac{\mathcal{T}\sigma^2 r^\alpha}{P_\mathrm{t}}$ of the exponential term should be less than 0.1 for a given PDF $f^i_{\mathrm{far}}(r, \mathbf{x}, \mathbf{y})$. Therefore, $\exp\left(-\frac{\mathcal{T}\sigma^2 r^\alpha}{P_\mathrm{t}}\right) \approx 1 - \frac{\mathcal{T}\sigma^2 r^\alpha}{P_\mathrm{t}}$ for $\frac{\mathcal{T}\sigma^2 r^\alpha}{P_\mathrm{t}} \leq 0.1$ with percentage error less than 0.05%. After applying this approximation in the constraint $C1$, we get

$$P_\mathrm{t} \geq \frac{\mathcal{T}\sigma^2}{\epsilon} \int\limits_r r^\alpha f^i_{\mathrm{far}}(r, \mathbf{x}, \mathbf{y})\mathrm{d}r \tag{10.8}$$

Though (10.8) gives the lower bound for the transmit power $P_\mathrm{t}$, using the coverage probability in the $i$th cell, the optimal transmit power $P^*_\mathrm{t}$ over the homogeneous network can be obtained by taking the maximum of the lower bound computed over different cells. The optimal transmit power $P^*_\mathrm{t}$ can be mathematically expressed as:

$$P^*_\mathrm{t} = \max_i \left\{ \frac{\mathcal{T}\sigma^2}{\epsilon} \int\limits_r r^\alpha f^i_{\mathrm{far}}(r, \mathbf{x}, \mathbf{y})\mathrm{d}r \right\};\ \text{for } i \in \{1, 2, \ldots, N_\mathrm{b}\} \tag{10.9}$$

After substituting $P^*_\mathrm{t}$ into (P0), the problem (P1) can be formulated as:

$$\text{(P1): } \underset{N_\mathrm{b}, \mathbf{x}, \mathbf{y}}{\text{minimize}}\ \mathrm{N_b} \left[ c_\mathrm{b} \cdot \max_i \left\{ \int\limits_r r^\alpha f^i_{\mathrm{far}}(r, \mathbf{x}, \mathbf{y})\mathrm{d}r \right\} + \epsilon_2 \right]$$
$$\text{subject to } C2, C4, C5,$$

where $c_\mathrm{b} = \frac{\mathcal{T}\sigma^2\epsilon_1}{\epsilon}$. Next, using (P1), we optimize the location of the BSs for their given number $N_\mathrm{b}$ over the specified field.

## 10.4 Deployment Strategy for a Single BS

To analyze the deployment of a single BS over the field, the optimization problem (P1) is equivalently expressed as:

$$\text{(P2): } \underset{x,y}{\text{minimize}} \int_{r} r^{\alpha} f_{\text{far}}(r,x,y)\mathrm{d}r$$
$$\text{subject to } \widehat{C4}: \ 0 \le x \le x_{\max}, \ \widehat{C5}: \ 0 \le y \le y_{\max}$$

Using (P2), we now describe the optimal deployment strategy of a single BS when the number of UEs is (i) large or (ii) moderate over the field.

### *10.4.1 For Large Number of UEs*

In the scenario, when the number of UEs is large, i.e., $N_{\text{u}} \to \infty$, the optimal location of the BS can be determined using Lemma 1 as follows.

**Lemma 1** *When $N_u \to \infty$, then the PDF of distance of farthest UE from the BS is given as:*

$$\lim_{N_u \to \infty} f_{far}(r,x,y) = \delta(r - r_f(x,y)), \tag{10.10}$$

*where $r_f(x,y)$ is the farthest Euclidean distance from the BS located at $(x,y)$ and $\delta(\cdot)$ is a Dirac delta function.*

***Proof*** If we consider Fig. 10.3, the probability of lying of farthest UE near the point $\mathcal{P}$ can be computed using the distribution given by (10.4). Probability of lying of farthest UE near the point $\mathcal{P}$ at $r_f$ from the BS for $N_{\text{u}} \to \infty$ is given as:

$$\lim_{\substack{\Delta \to 0^+ \\ N_{\text{u}} \to \infty}} \Pr(r_f - \Delta < r \le r_f) = \lim_{\substack{\Delta \to 0^+ \\ N_{\text{u}} \to \infty}} [F_{\text{far}}(r_f,x,y) - F_{\text{far}}(r_f - \Delta,x,y)]$$
$$= \lim_{N_{\text{u}} \to \infty} [F(r_f,x,y)]^{N_{\text{u}}} - \lim_{\substack{\Delta \to 0^+ \\ N_{\text{u}} \to \infty}} [F(r_f - \Delta,x,y)]^{N_{\text{u}}} = 1 - 0 = 1 \tag{10.11}$$

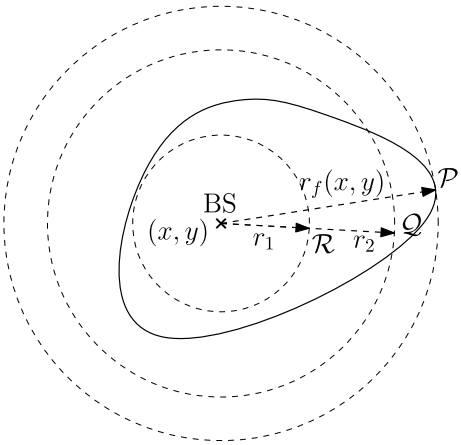



**Fig. 10.3** Farthest Euclidean distance $r_f(x,y)$ from the BS located at a point $(x,y)$

Here, note that $F(r, x, y) = 1$ and $< 1$ for $r = r_f$ and $< r_f$, respectively. Therefore, in (10.11), $\lim_{N_u \to \infty}[F(r_f, x, y)]^{N_u} = 1$ and $\lim_{\substack{\Delta \to 0^+ \\ N_u \to \infty}}[F(r_f - \Delta, x, y)]^{N_u} = 0$. Now, we find the probability of lying of farthest UE between the intermediate points $\mathcal{Q}$ at $r_2$ and $\mathcal{R}$ at $r_1$ as shown in Fig. 10.3. It can be expressed as:

$$\lim_{N_u \to \infty} \Pr(r_1 < r \leq r_2) = \lim_{N_u \to \infty} [F(r_2, x, y)]^{N_u} - \lim_{N_u \to \infty} [F(r_1, x, y)]^{N_u} = 0. \quad (10.12)$$

Therefore, from (10.11) and (10.12), for $N_u \to \infty$, the farthest UE always lies at the farthest Euclidean distance from the BS. Hence, from (10.11), $\lim_{N_u \to \infty} \int_{r_f - \Delta}^{r_f} f_{\text{far}}(r, x, y)\mathrm{d}r = \int_{r_f - \Delta}^{r_f} \delta(r - r_f)\mathrm{d}r$ that gives the PDF as expressed in (10.10).

For $N_u \to \infty$, if we substitute the obtained PDF in (10.10) into the objective function of (P2), the problem equivalently changed to the minimization of $\{r_f(x, y)\}^{\alpha}$ or $r_f(x, y)$ under the constraints $\widehat{C4}$ and $\widehat{C5}$. *Based on it, it can be easily shown that the optimal location of a BS in a circular or in a regular polygon is at the center from which the farthest euclidean distance is minimum.*

### 10.4.2 For Moderate Number of UEs

To obtain the optimal solution of (P2) for a moderate number of UEs present over the field, first, we investigate it for a square field as shown in Fig. 10.4. Here, a BS is located at distance $d$ (leftward) from the center of the square field. The expression of PDF $f(r, d)$ of distance $r$ of a UE from the BS is derived in the appendix at the end of the chapter. The variation of $f(r, d)$ with $r$ for a different location $d$ of the BS is numerically plotted in Fig. 10.5. It shows that the peak of the PDF is highest when BS is located at $d = 0$ (center of the square field), and it gradually decreases with $d$, and in contrary, the farthest Euclidean distance $r_f$ from the BS increases with $d$. From (10.3), PDF $f_{\text{far}}(r, d)$ of farthest UE distance also has the same variation, and based on it, we find the optimal deployment strategy of a single BS in the square field using Lemma 2 when $N_u$ is moderate.

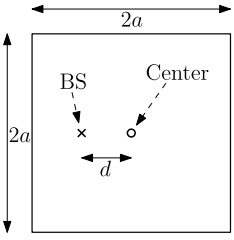



**Fig. 10.4** Optimal deployment strategy of a single BS in a square field when number of distributed UEs is moderate

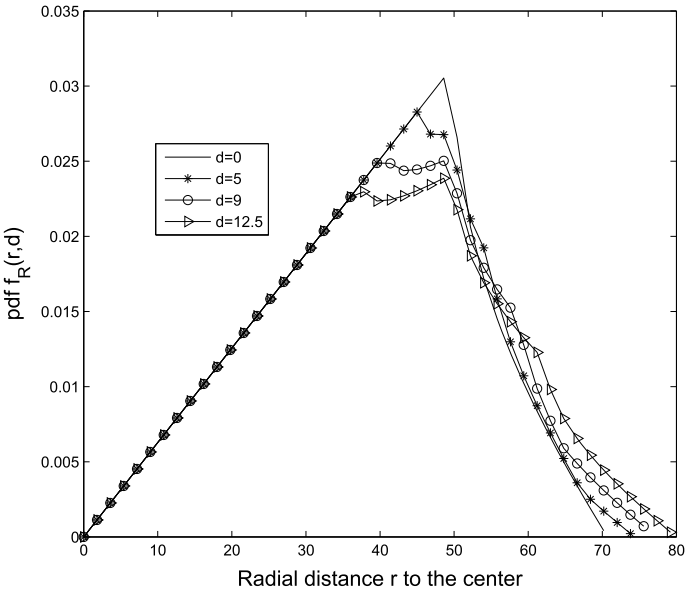



**Fig. 10.5** PDF $f(r, d)$ of distance $r$ of an UE from the BS located at distance $d$ from the center of the square field (cf. Fig. 10.4)

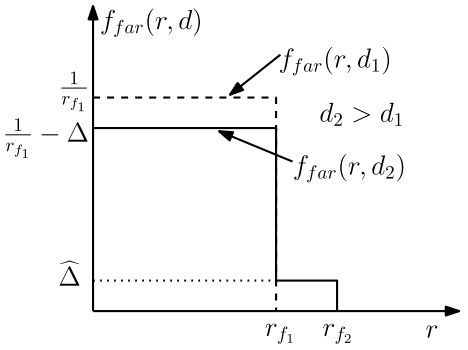



**Fig. 10.6** Diagram for describing the variation of peak and farthest Euclidean distance of $f(r, d)$ (cf. Fig. 10.5) for $d = d_1$ and $d_2$, where $d_1 < d_2$

**Lemma 2** *If peak of PDF $f_{far}(r, d_1)$ is greater than the peak of $f_{far}(r, d_2)$, whereas the corresponding farthest Euclidean distance $r_{f_1} < r_{f_2}$ for $d_1 < d_2$ as shown in Fig. 10.6, then the Nth moment of the farthest UE distance under the two distribution satisfies:*

$$\int_0^{r_{f_1}} r^N f_{far}(r, d_1)\mathrm{d}r < \int_0^{r_{f_2}} r^N f_{far}(r, d_2)\mathrm{d}r. \tag{10.13}$$

***Proof*** From Fig. 10.6, $\left(\frac{1}{r_{f_1}} - \Delta\right) r_{f_1} + \widehat{\Delta}(r_{f_2} - r_{f_1}) = 1$ that gives $\widehat{\Delta} = \frac{\Delta\, r_{f_1}}{r_{f_2} - r_{f_1}}$ for $d_1 < d_2$. Using it, the result in (10.13) can be proved as:

$$\begin{aligned}
\int_0^{r_{f_2}} r^N f_{\text{far}}(r, d_2)\mathrm{d}r &= \int_0^{r_{f_1}} r^N \left(\frac{1}{r_{f_1}} - 1\right) \mathrm{d}r + \int_{r_{f_1}}^{r_{f_2}} r^N \widehat{\Delta}\mathrm{d}r \\
&= \left(\frac{1}{r_{f_1}} - \Delta\right) \frac{r_{f_1}^{N+1}}{N+1} + \frac{\Delta\, r_{f_1}}{r_{f_2} - r_{f_1}} \frac{r_{f_2}^{N+1} - r_{f_1}^{N+1}}{N+1} \\
&= \int_0^{r_{f_1}} r^N f_{\text{far}}(r, d_1)\mathrm{d}r + \frac{\Delta r_{f_1}}{N+1}(r_{f_2}^N + r_{f_2}^{N-1} r_{f_1} + \cdots + r_{f_2} r_{f_1}^{N-1}) \\
&> \int_0^{r_{f_1}} r^N f_{\text{far}}(r, d_1)\mathrm{d}r \qquad (10.14)
\end{aligned}$$

Therefore, the $N$th moment has the minimum value at $d = 0$ (center) of the square field.

Thus, from (10.14), the objective function of (P2) achieves the minimum value when the BS is located at the center of the square field. *Similarly, it can be shown that for moderate* $N_{\text{u}}$, *the center of any regular polygon or circular field is the optimal location for the deployment of a single BS.* Besides, at the optimal location, the $N$th moment can be reduced by minimizing the farthest Euclidean distance $r_f$ of the fields the same as the case of large $N_{\text{u}}$.

## 10.5 Deployment Strategy for Multiple BSs

In order to deploy multiple BSs over the square and circular field, we need to find the optimum shape of the generated cells with respect to coverage of UEs and coverage holes. To compare different shapes of the cells, we assume that the UEs are distributed using HPPP over a large field. Note that although our analysis is based on BPP, it converges into HPPP when the size of the field becomes large. If the cells as shown in Fig. 10.7 have same farthest Euclidean distance $r_f$ from their associated BSs, they can be compared in coverage using Lemma 3.

**Lemma 3** *If the number of UEs over a field is deployed using HPPP with density* $\lambda$, *then the coverage to average number of UEs in square, hexagonal, and circular cells is* 53.96, 100, *and* 141.84% *more than triangular cell of same farthest Euclidean distance.*

***Proof*** As the area of a equilateral triangular cell with the farthest Euclidean distance $r_f$ is $\frac{3\sqrt{3}r_f^2}{4}$, average number of UEs $N_{u,\mathcal{T}} = \frac{3\sqrt{3}r_f^2}{4}\lambda$. Likewise, the average number

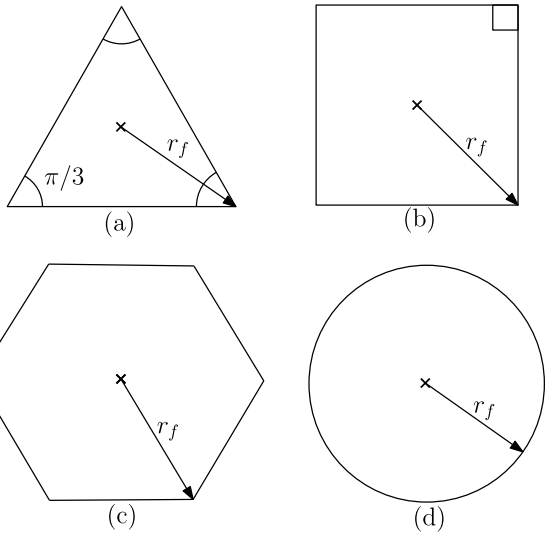



**Fig. 10.7** Comparison of different shapes of the cells having same farthest Euclidean distance with respect to coverage and coverage hole

of UEs $N_{u,\mathcal{S}}$, $N_{u,\mathcal{H}}$, and $N_{u,\mathcal{C}}$ in square, hexagonal, and circular cells is $2r_f^2\lambda$, $\frac{3\sqrt{3}r_f^2}{2}\lambda$, and $\pi r_f^2\lambda$, respectively. Percentage coverage improvement in square field compared to triangular cell is $\Delta_{u,\mathcal{T}\to\mathcal{S}} = \frac{N_{u,\mathcal{S}}-N_{u,\mathcal{T}}}{N_{u,\mathcal{T}}} \times 100 = 53.96\%$. Similarly, the improvement in hexagonal and circular cells against the triangular is $\Delta_{u,\mathcal{T}\to\mathcal{H}} = 100\%$ and $\Delta_{u,\mathcal{T}\to\mathcal{C}} = 141.84\%$, respectively.

From Lemma 3, the circular cell has the best coverage capability, but it creates the coverage holes over the field Rappaport (1996). Therefore, hexagonal cell better in coverage of UEs as well as in coverage hole performances in a large field. However, in a finite square or circular field, hexagonal cells create holes at the boundaries; therefore, we take square or rectangular cells in a square field and arc or triangular cells in a circular field to avoid coverage holes as shown in Figs. 10.8 and 10.9. Now, the optimal deployment of multiple BSs over the fields can be determined using the optimization problem (P1) when $N_{\mathrm{u}}$ is large or moderate.

### *10.5.1 For Large $N_u$*

For a large number of UEs ($N_{\mathrm{u}} \to \infty$) in a finite field, the number of UEs in each cell of the field is also large. Therefore, from Lemma 1, for a large number of UEs in $i$th cell ($N_{u,i} \to \infty$), the corresponding PDF of the distance of farthest UE from the associated BS can be expressed as:

$$f_{\mathrm{far}}^{i}(r, \mathbf{x}, \mathbf{y}) = \delta(r - r_{f,i}(\mathbf{x}, \mathbf{y})), \tag{10.15}$$

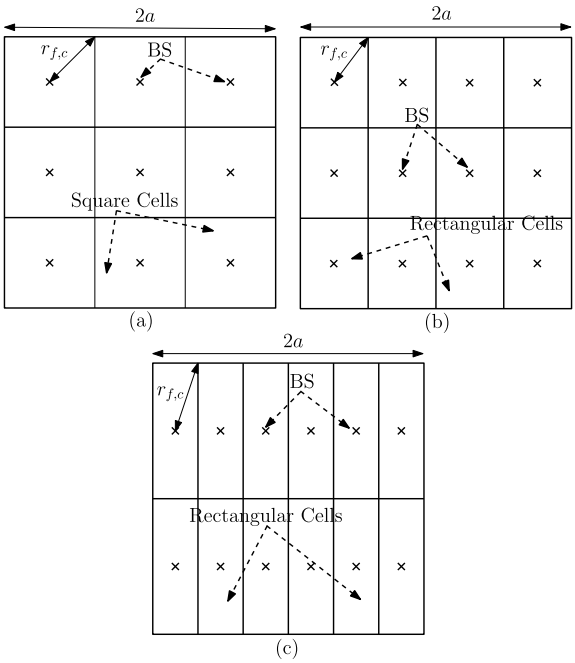



**Fig. 10.8** Square or rectangular cells in a square field due to different locations of the BSs. **a** $N_\mathrm{b} = 9$, $m = 3$, and $n = 3$. **b** $N_\mathrm{b} = 12$, $m = 3$, and $n = 4$. **c** $N_\mathrm{b} = 12$, $m = 2$, and $n = 6$

where $r_{f,i}$ is farthest Euclidean in the $i$th cell which is function of location $(\mathbf{x}, \mathbf{y})$ of all BSs in the field. After substituting (10.15) into the objective function of (P1), we get the optimization problem:

$$\text{(P3)}: \underset{N_\mathrm{b},\mathbf{x},\mathbf{y}}{\text{minimize}}\ \mathrm{N_b}\left[c_\mathrm{b} \cdot \max_i \left\{[r_{u,i}(\mathbf{x},\mathbf{y})]^\alpha\right\} + \epsilon_2\right]$$
$$\text{subject to } C2, C4, C5,$$

Now, using (P3), we find the optimal deployment of $N_\mathrm{b}$ BSs over the square field. For a given $N_\mathrm{b}$, BSs' locations can be optimized by minimizing $\max_i \left\{r_{f,i}(\mathbf{x},\mathbf{y})\right\}$ under the constraints $C4$ and $C5$. As described above, the optimal shape of the cells with respect to coverage of UEs and coverage hole is square or rectangular with equal farthest Euclidean distance ($r_{f,i} = r_{f,c}\ \forall i$) from their BSs optimally located at the center of the cells as shown in Fig. 10.8.

As a result, minimization of $\max_i \left\{r_{f,i}(\mathbf{x},\mathbf{y})\right\}$ is reduced to minimizing $r_{f,c}$ using a different arrangement of the cells along the length and the width of the field. The optimal arrangement can be determined using Lemma 4 as follows.

**Lemma 4** *For $N_b = m \times n$ number of BSs are deployed in their respective cells over a square field, the minimum value of farthest Euclidean distance $r_{f,c}$ is achieved at $m = n = \sqrt{N_b}$ or minimum $|m - n|$ when $\sqrt{N_b}$ is an integer or not an integer, respectively.*

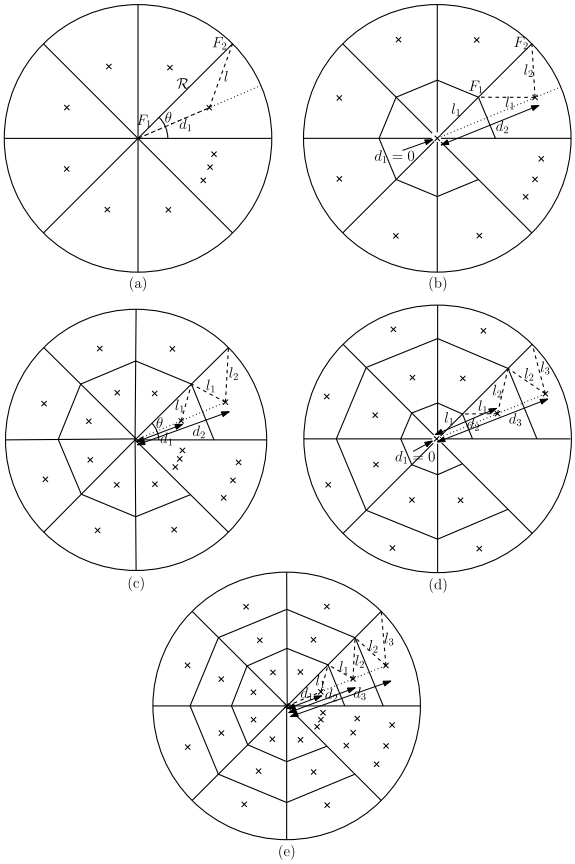



**Fig. 10.9** Arc, rhombus, and triangular cells in a circular field due to different locations of the BSs. **a** $N_\mathrm{b} = q$, **b** $N_\mathrm{b} = q + 1$, **c** $N_\mathrm{b} = 2q$, **d** $N_\mathrm{b} = 2q + 1$, and **e** $N_\mathrm{b} = 3q$

***Proof*** If we relax the integer value of number of rows $m$ and number of columns $n$, then we assume that $n = m - \omega$, where $\omega \geq 0$. The farthest Euclidean distance $r_{f,c} = \sqrt{\left(\frac{a}{m}\right)^2 + \left(\frac{a}{m-\omega}\right)^2}$, where $2a$ is the side length of the square field as in Fig. 10.8. Here, $r_{f,c}$ is equivalently minimized by minimizing $\mathcal{D} \triangleq \frac{1}{m^2} + \frac{1}{(m-\omega)^2}$. As $N_\mathrm{b} = m(m - \omega)$, we get $m = \frac{\omega + \sqrt{\omega^2 + 4N_\mathrm{b}}}{2}$. After substitution of $m$, $\mathcal{D}$ can be further expressed as $\mathcal{D} = \frac{\omega^2 + 2N_\mathrm{b}}{N_\mathrm{b}^2}$. As, $\frac{\partial \mathcal{D}}{\partial \omega} = \frac{2\omega}{N_\mathrm{b}^2}$ and $\frac{\partial^2 \mathcal{D}}{\partial \omega^2} = \frac{2}{N_\mathrm{b}^2} > 0$, $\mathcal{D}$ is convex. Thus, $\mathcal{D}$ and $r_{f,c}$ achieve its minimum value at $\omega = m - n = 0$ that gives $m = n = \sqrt{N_\mathrm{b}}$ when $\sqrt{N_\mathrm{b}}$ is an integer. Otherwise, factorize $N_\mathrm{b}$ into $m$ and $n$ such that $|\omega| = |m - n|$ is minimum.

For example, in Figs. 10.8b, c, through the total number of BSs $N_\mathrm{b} = 12$, but in Fig. 10.8b, the number of rows denoted as $m$ is 3 and number of columns denoted as $n$ is 4. Whereas in Fig. 10.8c, $m = 2$ and $n = 6$. As $\sqrt{N_\mathrm{b}} = \sqrt{12} = 3.46$ is not an integer, we factorize $N_\mathrm{b} = 12$ into $m$ and $n$ such that the difference $|m - n|$ is minimum (cf. Lemma 4). The arrangement in Fig. 10.8b is optimum because $|m -$

$n| = |3 - 4| = 1$ is minimum possible value as compared to the arrangement in Fig. 10.8c where $|m - n| = |2 - 6| = 4$.

In case of circular field, the BSs are deployed in radial and angular directions as shown in Fig. 10.9. Based on it, we define the arrangement as $pq + t$, where $t = 0$ and $t = 1$ are for the absence and presence of a BS at the center of the field, $q$ is the number of arcs in angular direction, and $p$ is the number of cells in each arc along the radial direction. The deployment of BSs in each arc is same; therefore, we study the optimal deployment of BSs in an arc of the field. Note that in each arc, the BSs are located at the symmetric line as shown in Fig. 10.9. As $N_u \to \infty$, the optimal deployment is determined using farthest Euclidean distance of the cells. For a given $N_b$, the optimal deployment can be determined by $\min \max_i \{r_{f,i}\}$ for $i \in \{1 - t, 2 - t, \cdots, p\}$. As described in (Prasad et al. 2018, Sect. 4), using the trigonometric relationship in the particular arc of Figs. 10.9a–e, we can find the optimal arrangement (type), $\min \max_i \{r_{f,i}\}$, optimal location of BSs for a given $N_b$ which are listed in Tables 10.1 and 10.2.

**Table 10.1** For a given number of BSs $N_b$, optimal arrangement (type) of BSs and corresponding optimal location in an arc of the circular field

| Given $N_b$ | Optimal Type | Optimum location of the BSs in an arc |
|---|---|---|
| $N_b = 3$ | $q$ | $d_1^* = \mathcal{R}\cos\left(\frac{\pi}{N_b}\right)$ |
| $N_b \in \{4, 5, 6\}$ | | $d_1^* = \frac{\mathcal{R}}{2\cos\left(\frac{\pi}{N_b}\right)}$ |
| $N_b \in \{7, 8, \ldots, 17\} \cup \{19\}$ | $q + 1$ | $d_0^* = 0, d_1^* = \frac{2\mathcal{R}\cos\left(\frac{\pi}{N_b - 1}\right)}{4\cos^2\left(\frac{\pi}{N_b - 1}\right) - 1}$ |
| $N_b \in \{18, 20, \ldots, 44\}$ | $2q$ | $d_1^* = \frac{\mathcal{R}}{4\cos\left(\frac{2\pi}{N_b}\right)\cos\left(\frac{4\pi}{N_b}\right)}$, $d_2^* = \frac{\mathcal{R}\left(1+\cos\left(\frac{4\pi}{N_b}\right)\right)}{4\cos\left(\frac{2\pi}{N_b}\right)\cos\left(\frac{4\pi}{N_b}\right)}$ |
| $N_b \in \{21, 23, \ldots, 45\}$ | $2q + 1$ | $d_0^* = 0, d_1^* = \frac{2\mathcal{R}\left(1+2\cos\left(\frac{4\pi}{N_b - 1}\right)\right)\cos\left(\frac{2\pi}{N_b - 1}\right)}{16\cos^2\left(\frac{2\pi}{N_b - 1}\right)\cos^2\left(\frac{4\pi}{N_b - 1}\right) - 1}$, $d_2^* = \frac{4\mathcal{R}\left(1+2\cos\left(\frac{4\pi}{N_b - 1}\right)\right)\cos\left(\frac{4\pi}{N_b - 1}\right)\cos\left(\frac{2\pi}{N_b - 1}\right)}{16\cos^2\left(\frac{2\pi}{N_b - 1}\right)\cos^2\left(\frac{4\pi}{N_b - 1}\right) - 1}$ |
| $N_b \in \{48, 51, \ldots\}$ | $3q$ | $d_1^* = \frac{\mathcal{R}\cos\left(\frac{3\pi}{N_b}\right)}{\left(2\cos\left(\frac{6\pi}{N_b}\right)+1\right)\left(\cos\left(\frac{12\pi}{N_b}\right)+\cos\left(\frac{6\pi}{N_b}\right)\right)}$, $d_2^* = \frac{\mathcal{R}\cos\left(\frac{3\pi}{N_b}\right)\left(1+2\cos\left(\frac{6\pi}{N_b}\right)\right)}{\left(2\cos\left(\frac{6\pi}{N_b}\right)+1\right)\left(\cos\left(\frac{12\pi}{N_b}\right)+\cos\left(\frac{6\pi}{N_b}\right)\right)}$, $d_3^* = \frac{\mathcal{R}\cos\left(\frac{3\pi}{N_b}\right)\left(1+2\cos\left(\frac{6\pi}{N_b}\right)+2\cos\left(\frac{9\pi}{N_b}\right)\right)}{\left(2\cos\left(\frac{6\pi}{N_b}\right)+1\right)\left(\cos\left(\frac{12\pi}{N_b}\right)+\cos\left(\frac{6\pi}{N_b}\right)\right)}$ |

**Table 10.2** For a given $N_b$, optimal arrangement (type) of BSs and corresponding minimum of maximum of farthest Euclidean distance over the cells in an arc of the circular field

| Given $N_b$ | Optimal Type | Minimum of $\max_i \{r_{f,i}\}$ |
|---|---|---|
| $N_b = 3$ | $q$ | $\mathcal{R}\sin\left(\frac{\pi}{N_b}\right)$ |
| $N_b \in \{4, 5, 6\}$ | | $\frac{\mathcal{R}}{2\cos\left(\frac{\pi}{N_b}\right)}$ |
| $N_b \in \{7, 8, \ldots, 17\} \cup \{19\}$ | $q+1$ | $\frac{\mathcal{R}}{4\cos^2\left(\frac{\pi}{N_b-1}\right)-1}$ |
| $N_b \in \{18, 20, \ldots, 44\}$ | $2q$ | $\frac{\mathcal{R}}{4\cos\left(\frac{2\pi}{N_b}\right)\cos(\frac{4\pi}{N_b})}$ |
| $N_b \in \{21, 23, \ldots, 45\}$ | $2q+1$ | $\frac{\mathcal{R}\left(1+2\cos\left(\frac{4\pi}{N_b-1}\right)\right)}{16\cos^2\left(\frac{2\pi}{N_b-1}\right)\cos^2\left(\frac{4\pi}{N_b-1}\right)-1}$ |
| $N_b \in \{48, 51, \ldots\}$ | $3q$ | $\frac{\mathcal{R}\cos\left(\frac{3\pi}{N_b}\right)}{\left(2\cos\left(\frac{6\pi}{N_b}\right)+1\right)\left(\cos\left(\frac{12\pi}{N_b}\right)+\cos\left(\frac{6\pi}{N_b}\right)\right)}$ |

### *10.5.2 For Moderate $N_u$*

When $N_u$ is moderate, the number of UEs occurring in a cell depends on the occurrence of UEs in other cells. As described above, square or rectangular cells are optimal in the square field and the optimal location of the BSs is at the center of the cells even for moderate $N_u$. Therefore, for a given $N_b$, the optimal deployment of BSs is based on minimization of farthest Euclidean distance over cells the same as the case of large $N_u$. Also, it has been shown in Prasad et al. (2018) that deployment based on the minimization of the farthest euclidean distance over the cells is acceptable in the circular field within negligible root-mean-square error (RMSE). Thus, for $N_u \to \infty$, the optimal deployment strategy of given BSs over the square and circular fields is the same as described in Sect. 10.5.1.

## 10.6 Joint Optimization

In this section, we study about the joint optimization of number of BSs $N_b$ and their deployment (arrangement of cells) in both square and circular field.

### *10.6.1 Optimization over the Square Field*

To jointly optimize the number of BSs $N_b$ and their location over the square field for $N_u \to \infty$, (P3) can be simplified as:

$$\text{(P4)}: \underset{N_\text{b},m,n}{\text{minimize}}\ \text{N}_\text{b}\left[\widehat{c}_\text{b}\cdot\left(\frac{1}{m^2}+\frac{1}{n^2}\right)^{\frac{\alpha}{2}}+\epsilon_2\right],$$
$$\text{subject to}\ \ C2,\ \ \overline{C4}: 1\le m\le N_\text{b},\ \ \overline{C5}: 1\le n\le N_\text{b},$$

where $\widehat{c}_\text{b}=a^\alpha c_\text{b}$ and $2a$ is side length of the square field. Iteratively, the optimal value of $N_\text{b}$, $m$, and $n$ can be obtained as follows. We start from $N_\text{b}=1$ and for each $N_\text{b}\in\{1,2,\cdots,N_\text{max}\}$, optimal value of $m$ and $n$ is calculated using Lemma 4; thereafter, we find the corresponding total power consumption using the objective function of (P4). The obtained total power consumption for different $N_\text{b}$ is compared and that $N_\text{b}$ is set to optimal number of BSs $N_\text{b}^*$ which gives minimum value of total power consumption, and corresponding $m$ and $n$ are set as optimal number of rows $m^*$ and optimal number of columns $n^*$, respectively.

For moderate $N_\text{u}$, the joint optimization can be obtained by solving (P1) which can be expressed for the square field as:

$$\text{(P5)}: \underset{N_\text{b},m,n}{\text{minimize}}\ \text{N}_\text{b}\left[c_\text{b}\cdot\int_0^{\sqrt{\left(\frac{a}{m}\right)^2+\left(\frac{a}{n}\right)^2}} r^\alpha f_\text{far}^c(r,d=0)\text{d}r+\epsilon_2\right]$$
$$\text{subject to}\ \ C2,\overline{C4},\overline{C5},$$

where $d=0$ denotes that BS in each cell is located at the center and $f_\text{far}^c(r,d=0)$ is obtained using (10.6), (10.17), and (10.18). Here, superscript $c$ in the distribution denotes that the distribution in each cell is same because the cells are identical in the square field. Again, for the optimal solution, we start from $N_\text{b}=1$, and for each $N_\text{b}$, optimal value of $m$ and $n$ is obtained using Lemma 4. From it, the total power consumption is computed using the objective function of (P5) for each $N_\text{b}$. The $N_\text{b}$ which gives the minimum value of power consumption is set as $N_\text{b}^*$, and corresponding $m$ and $n$ are set as $m^*$ and $n^*$, respectively.

### *10.6.2 Optimization over the Circular Field*

In case of joint optimization over the circular field, we find the optimal arrangement (type) and corresponding optimal location and min $\max_i\{r_{f,i}\}$ using Tables 10.1 and 10.2 for each $N_\text{b}$. For $N_\text{u}\to\infty$ and moderate $N_\text{u}$, we find the total power consumption using the objective function of (P3) and (P1), respectively. The $N_\text{b}$ which gives the minimum power consumption is set to $N_\text{b}^*$, and corresponding arrangement (type) and locations are set as the optimal value.

## 10.7 Numerical Results and Discussions

In this section, we describe the obtained numerical results for the discoursed analysis where the default value of the used system parameters for the results is listed in Table 10.3. Using it, we find the design insights on the transmit power $P_t$, number of BSs $N_b$, and total power consumption with variation of system parameters.

Figures 10.10 and 10.11 describe the variation of optimal transmit power $P_t^*$ with noise power $\sigma^2$ for different values of path loss exponent $\alpha$ and acceptable tolerance $\epsilon$. It can be observed that to increase the coverage from 90 to 99%, we require to increase $P_t^*$ on average by 10 dB Watt over both the fields. On the other hand, if $\alpha$ is changed from 3 to 3.3, $P_t^*$ is increased by 6 dB Watt over the circular field, whereas

**Table 10.3** List of system parameters with their default values

| System parameter | Value |
|---|---|
| Side length of the square field, $2a$ | 1000 m |
| Radius of the circular field, $R$ | 500 m |
| AWGN noise power, $\sigma^2$ | $-70$ dBm |
| Mean of exponential distribution, $\mu$ | 1 |
| Threshold SNR, $\mathcal{T}$ | $-10$ dB |
| Path loss exponent, $\alpha$ | 4 |
| Scaling parameter, $\epsilon_1$ | 5.5 |
| Electronics, processing, and battery backup power losses, $\epsilon_2$ | 32 W |
| Maximum transmit power, $P_{max}$ | 5 W |
| Maximum number of BSs, $N_{max}$ | 35 |

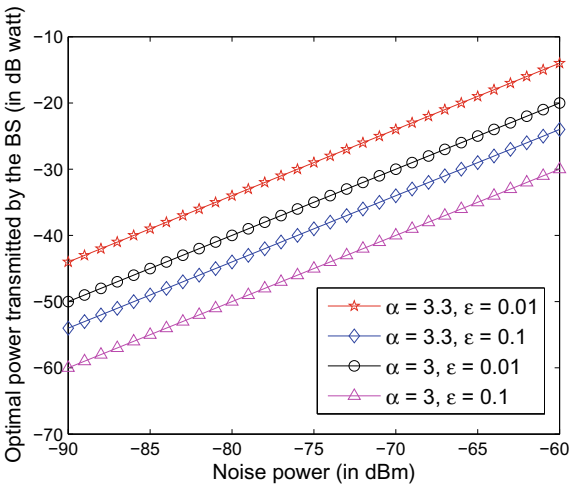



**Fig. 10.10** Optimal transmit power $P_t^*$ with noise power for different values of $\alpha$ and $\epsilon$ over the circular field

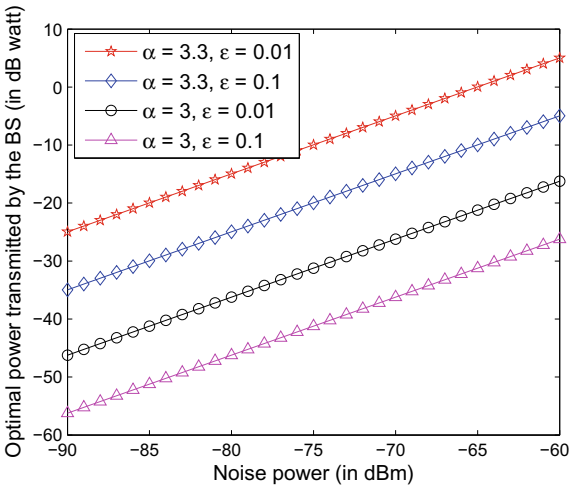



**Fig. 10.11** Optimal transmit power $P_t^*$ with noise power for different values of $\alpha$ and $\epsilon$ over the square field

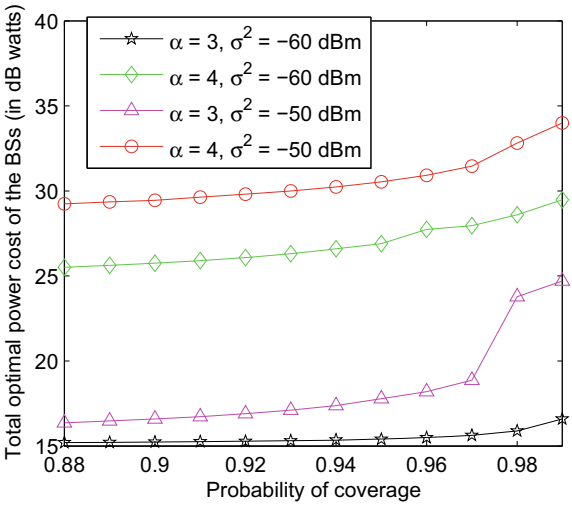



**Fig. 10.12** Minimized total power consumption with coverage probability for different values of $\alpha$ and $\sigma^2$ over the circular field

it is increased by 20 dB Watt over the square field. Therefore, the circular field is more energy-efficient than the square field.

In Figs. 10.12 and 10.13, the obtained results depict the total power consumption with coverage probability over the circular and square field, respectively. Here, in the circular field, the increment in the total power consumption at $\alpha = 4$ is more than $\alpha = 3$ when $\sigma^2$ changes from $-60$ dB to $-50$ dB. In contrast, over the square field, higher increment takes place at $\alpha = 3$ for the same change in $\sigma^2$. Therefore, the energy saving in circular field is highly sensitive with noise for higher path loss exponent, but in square field, the sensitivity is high for its low value.

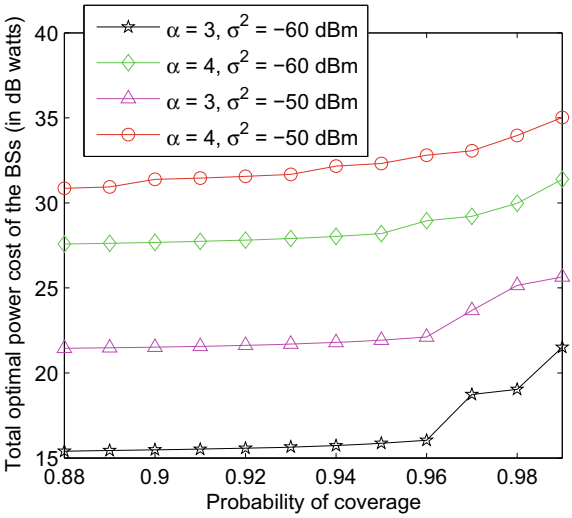



**Fig. 10.13** Minimized total power consumption with coverage probability for different values of $\alpha$ and $\sigma^2$ over the square field

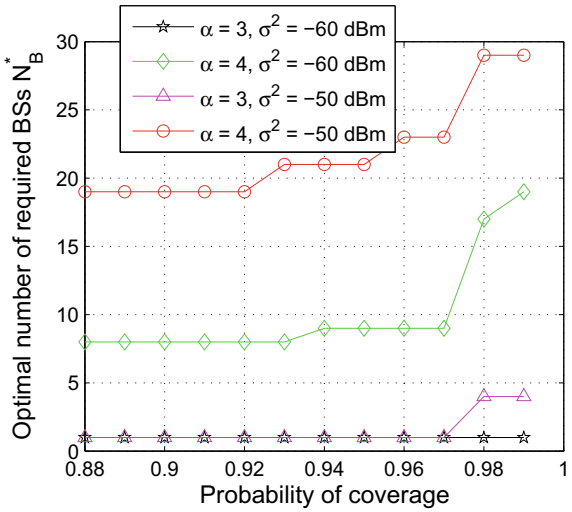



**Fig. 10.14** Optimal number of BSs $N_b^*$ with coverage probability for different values of $\alpha$ and $\sigma^2$ over the circular field

If we examine Figs. 10.14, 10.15, 10.16, and 10.17, Figs. 10.14 and 10.15 describe the optimal number of BSs $N_b^*$ and Figs. 10.16 and 10.17 depict the optimal transmit power $P_t^*$ with coverage probability over the two fields. In Fig. 10.14, at $\alpha = 3$, for the change of $\sigma^2$ from $-60$ dBm to $-50$ dBm, $N_b^*$ is same up to the 97% coverage over the circular field. It can be explained using Fig. 10.16 where the optimal transmit power $P_t^*$ is higher for $\sigma^2 = -50$ dBm which helps in compensation of the higher

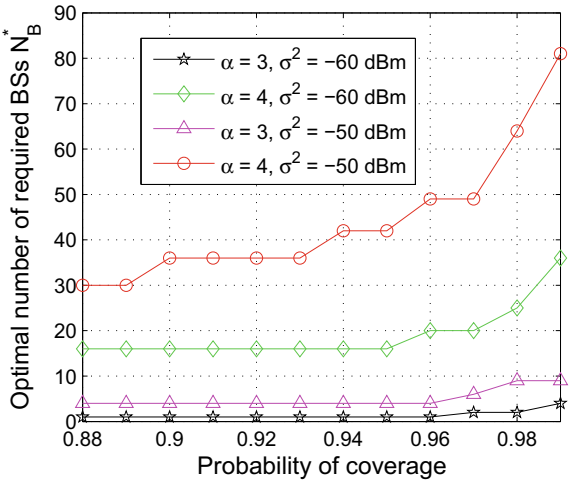



**Fig. 10.15** Optimal number of BSs $N_b^*$ with coverage probability for different values of $\alpha$ and $\sigma^2$ over the square field

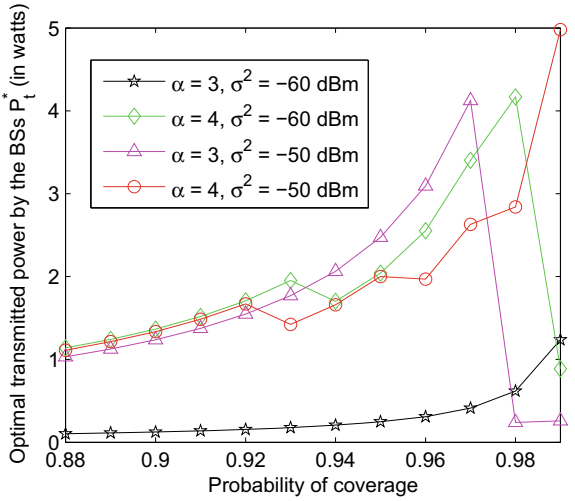



**Fig. 10.16** Optimal transmit power $P_t^*$ with coverage probability for different values of $\alpha$ and $\sigma^2$ over the circular field

noise present in the channel. But, in case of the square field, at $\sigma^2 = -50$ dBm, $N_b^*$ is increased (cf. Fig. 10.15) while keeping almost same transmit power (cf. Fig. 10.17) that compensate the noise by decreasing the cell size over the square field. We can also find that at around 97% coverage, $N_b^*$ is increased abruptly but $P_t^*$ is decreased in large amount. Therefore, the abrupt changes do not take place in the total power consumption as shown in Figs. 10.12 and 10.13.

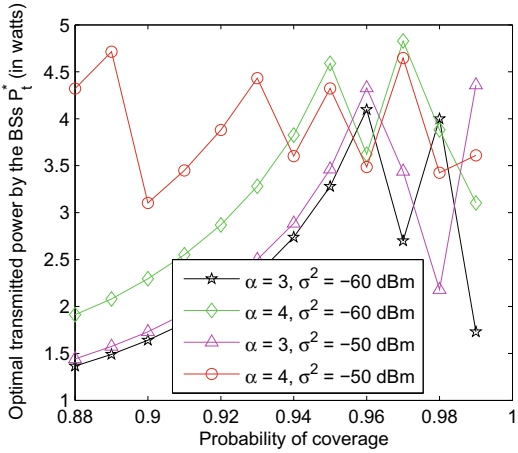



**Fig. 10.17** Optimal transmit power $P_t^*$ with coverage probability for different values of $\alpha$ and $\sigma^2$ over the square field

## 10.8 Summary

This chapter describes the strategy for minimization of total power consumption while satisfying the desired coverage of the UEs to provide the minimum throughput over a wireless network. In order to achieve the goal, the deployment of BSs, their number, and transmit power are optimized in two scenarios: (i) when large number of UEs are present and (ii) when moderate UEs are distributed over a square or circular field. From the obtained numerical results, we find that the circular field is more energy-efficient than the square field in achieving the goal.

## Appendix

Here, we derive the distribution of distance of a UE which is located leftward at distance $d$ from the center of the square field as shown in Fig. 10.18. Probability that a UE lies at $\leq r$ distance from the BSs can be given by:

$$\Pr(R \leq r, d) = F(r, d) = \frac{Ar(C(r, d) \cap \square)}{Ar(\square)}, \tag{10.16}$$

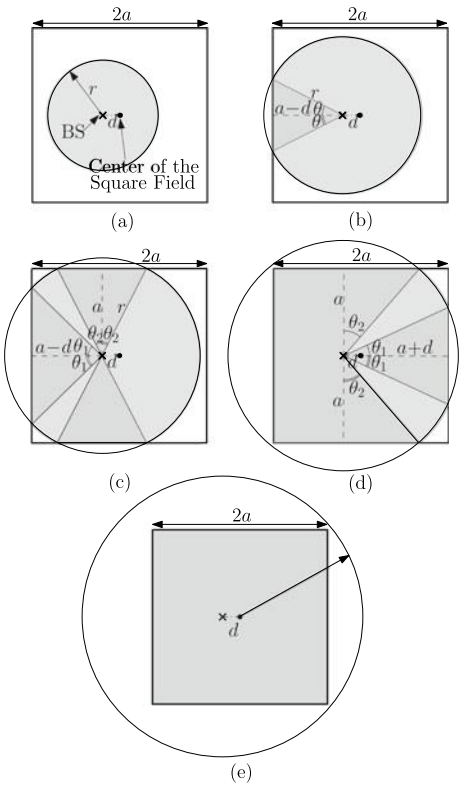



**Fig. 10.18** Distribution of distance of an UE from the BS located at distance $d$ from the center of the square field

where $F(r, d)$ is the CDF of distance $R$ of a UE from the BS located at $d$ from the center of the square field, $C(r, d)$ is the circular field centered at the BS with radius $r$, $Ar(C(r, d) \cap \square)$ is the area of intersection of the circular field and the square field, and $Ar(\square)$ is the area of the square field. From Fig. 10.18, it can be observed that the intersection of the circular with square field changes with $r$. Moreover, the intersection also changes with $d$ from the center of the square field. Based on it, the CDF $F(r, d)$ for $d \in [0, \frac{a}{4}]$ can be determined as:

$$F(r,d) = \begin{cases} F_{11} = \frac{\pi r^2}{4a^2}, & \text{a}: \ 0 \le r < a-d \\ F_{12} = \frac{1}{4a^2}\big[(a-d)\sqrt{r^2-(a-d)^2} \\ +\big(\pi - \cos^{-1}\frac{(a-d)}{r}\big)\frac{r^2}{2}\big], & \text{b}: \ a-d \le r < a \\ F_{13} = \frac{1}{4a^2}\big[(a-d)\sqrt{r^2-(a-d)^2} \\ +2a\sqrt{r^2-a^2} \\ +\big(\pi - \cos^{-1}\frac{(a-d)}{r} \\ -2\cos^{-1}\frac{a}{r}\big)r^2\big], & \text{c}: \ a \le r < a+d \\ F_{14} = \frac{1}{4a^2}\big[(a-d)\sqrt{r^2-(a-d)^2} \\ +2a\sqrt{r^2-a^2} \\ +(a+d)\sqrt{r^2-(a+d)^2} \\ +\big(\pi - \cos^{-1}\frac{(a-d)}{r} - \\ \cos^{-1}\frac{(a+d)}{r} - 2\cos^{-1}\frac{a}{r}\big)r^2\big], & \text{d}: \ a+d \le r \\ & \quad < \sqrt{(a-d)^2+a^2} \\ F_{15} = \frac{1}{4a^2}\big[2a(a-d) + a\sqrt{r^2-a^2} \\ +(a+d)\sqrt{r^2-(a+d)^2} \\ +\big(\pi/2 - \cos^{-1}\frac{a}{r} \\ -\cos^{-1}\frac{(a+d)}{r}\big)r^2\big], & \text{e}: \ \sqrt{(a-d)^2+a^2} \le r \\ & \quad < \sqrt{(a+d)^2+a^2} \\ F_{16} = 1, & \text{f}: \ r \ge \sqrt{(a+d)^2+a^2}. \end{cases} \tag{10.17}$$

$$f(r,d) = \begin{cases} f_{11} = \frac{\pi r}{2a^2}, & \text{a}: \ 0 \le r < a-d \\ f_{12} = \frac{1}{2a^2}\big(\pi - \cos^{-1}\frac{(a-d)}{r}\big)r, & \text{b}: \ a-d \le r < a \\ f_{13} = \frac{1}{2a^2}\big(\pi - \cos^{-1}\frac{(a-d)}{r} \\ -2\cos^{-1}\frac{a}{r}\big)r, & \text{c}: \ a \le r < a+d \\ f_{14} = \frac{1}{2a^2}\big(\pi - \cos^{-1}\frac{(a-d)}{r} - \\ \cos^{-1}\frac{(a+d)}{r} - 2\cos^{-1}\frac{a}{r}\big)r, & \text{d}: \ a+d \le r \\ & \quad < \sqrt{(a-d)^2+a^2} \\ f_{15} = \frac{1}{2a^2}\big(\pi/2 - \cos^{-1}\frac{a}{r} \\ -\cos^{-1}\frac{(a+d)}{r}\big)r, & \text{e}: \ \sqrt{(a-d)^2+a^2} \le r \\ & \quad < \sqrt{(a+d)^2+a^2} \\ f_{16} = 0, & \text{f}: \ r \ge \sqrt{(a+d)^2+a^2}. \end{cases} \tag{10.18}$$

Similarly, we can determine the CDF $F(r,d)$ and PDF $f(r,d)$ for other ranges of $d$. Now, using the same procedure, we determine the CDF $F(r,d)$ and PDF $f(r,d)$ of distance of a UE from the BS located at $d$ from the peak of the triangular field as shown in Fig. 10.19. For $d \in [0, \frac{a}{2}]$, the distribution can be determined as:

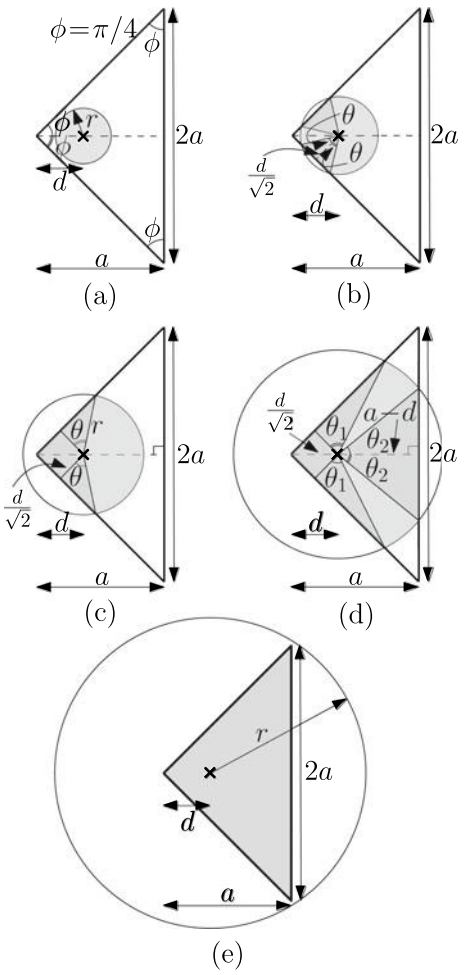



**Fig. 10.19** Distribution of distance of an UE from the BS located at distance $d$ from the peak of the triangular field

$$F(r,d) = \begin{cases} F_{21} = \frac{\pi r^2}{a^2}, & \text{p} : 0 \le r < \frac{d}{\sqrt{2}} \\ F_{22} = \frac{1}{a^2}\Big[\big(\sqrt{2}d\sqrt{r^2 - \frac{d^2}{2}} + (\pi & \\ -2\cos^{-1}\frac{d}{r\sqrt{2}}\big)r^2\Big], & \text{q} : \frac{d}{\sqrt{2}} \le r < d \\ F_{23} = \frac{1}{a^2}\Big[\frac{d^2}{2} + \frac{d}{\sqrt{2}}\sqrt{r^2 - \frac{d^2}{2}} & \\ +\big(\frac{3\pi}{4} - \cos^{-1}\frac{d}{r\sqrt{2}}\big)r^2\Big], & \text{r} : d \le r < a - d \\ F_{24} = \frac{1}{a^2}\Big[\frac{d^2}{2} + \frac{d}{\sqrt{2}}\sqrt{r^2 - \frac{d^2}{2}} & \\ +(a-d)\sqrt{r^2 - (a-d)^2} & \\ +\big(\frac{3\pi}{4} - \cos^{-1}\frac{d}{r\sqrt{2}} & \\ -\cos^{-1}\frac{(a-d)}{r}\big)r^2\Big], & \text{s} : a - d \le r \\ & < \sqrt{(a-d)^2 + a^2} \\ F_{25} = 1, & \text{t} : r \ge \sqrt{(a-d)^2 + a^2}. \end{cases} \tag{10.19}$$

$$f(r,d) = \begin{cases} f_{21} = \frac{2\pi r}{a^2}, & \text{p} : 0 \le r < \frac{d}{\sqrt{2}} \\ f_{22} = \frac{2}{a^2}\big(\pi - 2\cos^{-1}\frac{d}{r\sqrt{2}}\big)r, & \text{q} : \frac{d}{\sqrt{2}} \le r < d \\ f_{23} = \frac{2}{a^2}\big(\frac{3\pi}{4} - \cos^{-1}\frac{d}{r\sqrt{2}}\big)r, & \text{r} : d \le r < a - d \\ f_{24} = \frac{2}{a^2}\big(\frac{3\pi}{4} - \cos^{-1}\frac{d}{r\sqrt{2}} & \\ -\cos^{-1}\frac{(a-d)}{r}\big)r, & \text{s} : a - d \le r \\ & < \sqrt{(a-d)^2 + a^2} \\ f_{25} = 0, & \text{t} : r \ge \sqrt{(a-d)^2 + a^2}. \end{cases} \tag{10.20}$$

Similarly, we can compute the distribution for other ranges of $d$.

## References

Andrews JG, Baccelli F, Ganti RK (2011) A tractable approach to coverage and rate in cellular networks. IEEE Trans Commun 59(11):3122–3134

Elsawy H, Hossain E, Haenggi M (2013) Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey. IEEE Commun Surv Tutor 15(3):996–1019

Fettweis G, Zimmermann E (2008) ICT energy consumption-trends and challenges. In: Proceedings of international symposium on wireless personal multimedia communications, vol 2, no 4, pp 1–6

González-Brevis P et al (2011) Base station location optimization for minimal energy consumption in wireless networks. IEEE Veh Technol Conf 1–5

Jie G, Sheng Z, Zhisheng N (2012) A dynamic programming approach for base station sleeping in cellular networks. IEICE Trans Commun 95(2):551–562

Marsan MA, Chiaraviglio L, Ciullo D, Meo M (2009) Optimal energy savings in cellular access networks. In: Proceedings of IEEE ICC workshops, pp 1–5

Mhiri F, Sethom K, Bouallegue R (2013) A survey on interference management techniques in femtocell self-organizing networks. J Netw Comput Appl 36(1):58–65
Oh E, Son K, Krishnamachari B (2013) Dynamic base station switching-on/off strategies for green cellular networks. IEEE Trans Wirel Commun 12(5):2126–2136
Oh E, Krishnamachari B (2010) Energy savings through dynamic base station switching in cellular wireless access networks. Proceedings of IEEE GLOBECOM, pp 1–5
Peng J, Tang H, Hong P, Xue K (2013) Stochastic geometry analysis of energy efficiency in heterogeneous network with sleep control. IEEE Wirel Commun Lett 2(6):615–618
Perabathini B, Kountouris M, Debbah M, Conte A (2014) Optimal area power efficiency in cellular networks. In: Proceedings of IEEE Globecom pp 1157–1161
Prasad G, Mishra D, Hossain A (2018) Joint optimization framework for operational cost minimization in green coverage-constrained wireless networks. IEEE Trans Green Commun Netw 2(3):693–706
Rappaport TS (1996) Wireless communications: principles and practice. Prentice Hall
Sarkar S, Ganti RK, Haenggi M (2014) Optimal base station density for power efficiency in cellular networks. In: Proceedings of IEEE ICC, pp 4054–4059
Srinivasa S, Haenggi M (2010) Distance distributions in finite uniformly random networks: theory and applications. IEEE Trans Veh Technol 59(2):940–949
Stojmenovic I, Ruhil AP, Lobiyal D (2006) Voronoi diagram and convex hull based geocasting and routing in wireless networks. Wirel Commun Mob Comput 6(2):247–258
Thompson HR (1956) Distribution of distance to nth neighbour in a population of randomly distributed individuals. Ecology 37(2):391–394
Verenzuela D, Björnson E, Sanguinetti L (2016) Optimal design of wireless networks for broadband access with minimum power consumption. Proceedings of IEEE ICC, pp 1–6
Wu J, Zhou S, Niu Z (2013) Traffic-aware base station sleeping control and power matching for energy-delay tradeoffs in green Cellular networks. IEEE Trans Wirel Commun 12(8):4196–4209
Wu J, Zhang Y, Zukerman M, Yung EKN (2015) Energy-efficient base-stations sleep-mode techniques in green cellular networks: a survey. IEEE Commun Surv Tutor 17(2):803–826
Zhou S, Gong J, Yang Z, Niu Z, Yang P (2009) Green mobile access network with dynamic base station energy saving. In: Proceedins of ACM MobiCom, pp 10–12

# Chapter 11
# Spectrum Sensing for Cognitive Radio Networks


**Abhijeet Bishnu**

**Abstract** There is a massive increase in the number of global users for wireless services such as broadband, cellular, and television. Thus, there is a requirement of uninterrupted connectivity with higher data rates. However, the current static frequency allocation method is unable to serve such a massive increase in users and their desired services. As a result, innovative methods should be used or developed, which can efficiently exploit the available spectrum with the minimum requirement of extra resources. Cognitive radio is one such attractive solution that exploits the spectrum efficiently by making opportunistic usage of the frequency bands that are not being used by licensed or primary users. Cognitive radio has the capability to measure, learn, and sense the parameters related to the radio environment, available spectrum, and power in the spectrum. In this chapter, first, we discuss the methods of spectrum sharing between the primary and secondary users. Next, we introduce different types of spectrum sensing schemes, which are classified under blind and non-blind techniques. We also examine different noise models, including Gaussian and color noise. In addition, cooperative spectrum sensing is also discussed, followed by the implementation of various algorithms on software-defined radio. At last, cognitive radio with full-duplex communication is discussed, which is considered in the context of 5G and beyond communication.

**Keywords** Cognitive radio · Spectrum sensing · Frequency sharing · Blind detection · Non-blind detection

## 11.1 Introduction

Each country has its own telecommunication regulation authority which allocates frequency bands to the users which are licensed to use the allotted band on a long-term

A. Bishnu (✉)
Institute for Digital Communications, School of Engineering,
The University of Edinburgh, Scotland EH9 3FG, UK
e-mail: abhijeet.bishnu87@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_11

basis. The licensed users are also allotted the topographical area for the use of allotted frequency bands. However, in the study conducted by different researchers worldwide, it has been found that a large amount of allotted spectrum across the globe remains underutilized in time, frequency, and space (Mishra and Johnson 2014). This inefficient usage of the precious limited spectrum motivates for the evolution of advanced spectrum access/sharing methods. In such advanced accessing schemes, the unlicensed users which are also known as secondary users are allowed to access the unused frequency bands temporarily. This technology of allowing secondary user (SU) to use spectrum of primary user (PU) in a more flexible and comprehensive way is called as cognitive radio (CR) technology (Wang and Liu 2011). CR is a promising solution which provides an effective and efficient usage of spectrum through dynamic spectrum allocation. CR has the capability to adjust the transmitter parameters in accordance with the operating wireless environment. Unlike conventional radio devices, CR-enabled devices provide cognitive capability and reconfigurability to the users. Cognitive capability means the ability to sense and collect information such as transmission frequency, power, and modulation from the nearby environment (Wang and Liu 2011). Reconfigurability means the ability to quickly change the operational parameters according to the gathered information to achieve the optimal usage of the resources thereby providing better performance (Wang and Liu 2011). By making opportunistic use of spectrum, CR enables SUs to observe the vacant sections of the spectrum, chose the best-unoccupied channel, coordinate spectrum sharing with secondary users (SUs), and free the channel usage whenever a PU desired to use the same spectrum. Based on dynamic spectrum sharing in CR, the SUs opportunistically use the spectrum without interfering or with an acceptable interference with the PUs communication using three techniques: underlay, overlay, and interweave techniques (Haykin and Setoodeh 2015).

In underlay CR technique, the interference caused by the unlicensed transmitter to the licensed receiver must be maintained under a certain threshold limit. Whereas in overlay CR technique, the SUs must have information about both PUs and SUs and the wireless channel between them which is a challenging task. In the interweave CR technique, SUs must not interfere the operation of PUs which means that as soon as the SUs detect the presence of active PUs in the same frequency band, they must shift to different frequency bands. The first wireless standard known as IEEE 802.22 wireless regional area networks (WRAN) is defined based on the interweave CR technique where digital TV (DTV) and wireless microphone are the PUs, and SUs are customer premises equipment (CPE) (Bishnu and Bhatia 2015). According to the federal communication commission (FCC), in in-band spectrum sensing (SS), the detection of active PUs should be detected within 2 s, and the missed detection and false alarm probabilities to be less than or equal to 0.1 for maximum protection of active PUs (Kim and Shin 2010). To meet such stringent conditions, in-band sensing must be done once in every two seconds, and also, a detection algorithm should be selected to yield optimal performance. Since SS is done in the course of quiet periods within which SUs communication is restricted. Thus, the effect of SS on SUs quality-of-service must also be evaluated.

SS in CR can be of two types; (a) blind SS (BSS) techniques, and (b) non-blind SS (NBSS) techniques. In BSS techniques, such as covariance (Cov), energy detection (ED), eigenvalue (Egnval), principal component analysis (PCA); the a priori information of PUs signal attributes is not required. Whereas in NBSS techniques such as power spectral density (PSD), cyclic prefix (CP), pilot tones (PT), and cyclostationary techniques (CST), SUs must have some attributes of the PUs. However, in the colored Gaussian noise, the signal detection performance of most of the algorithms deteriorates. On the other hand, in the literature, there are few algorithms which perform sensing in colored noise scenario. These are based on the correlation (Corr) coefficient of the noise samples. There are other issues like noise uncertainty, fading, and shadowing which degrades the performance of the algorithms. To circumvent the issues cooperative, SS is proposed in the literature. Cooperative sensing mitigates the probabilities of missed-detection, and also, the false alarm.

In this chapter, first, we discuss the methods of spectrum sharing between the primary and secondary user. Next, we introduce different types of SS schemes which are classified under BSS and NBSS, and also, in both Gaussian and colored noise scenarios. In addition, cooperative spectrum sensing is also discussed followed by the implementation of various algorithms on software-defined radio. At last, cognitive radio with full-duplex communication is discussed which is considered as 5G and beyond communication.

## 11.2 Mathematical Model

Let us consider that the symbol $x(t)$ represent continuous-time (CT) signal received at the SU and is sampled with period $T_s$, denoted is as $x[n] = x(nT_s)$. The signal detection at PU can mathematically be modeled using binary hypothesis (Wang and Liu 2011) as:

$$H_0 : x[n] = w[n] \tag{11.1}$$

$$H_1 : x[n] = s[n] + w[n] \tag{11.2}$$

where $w[n]$ and $s[n]$ are the additive white Gaussian noise (AWGN) and the transmitted signal of PU, respectively, in discrete-time (DT) domain. Also, we assume that the signals $s[n]$ and $w[n]$ are uncorrelated. In null hypothesis $H_0$, the PU's signal is absent, and it is present under the alternative hypothesis, $H_1$.

The problem of sensing is based on the statistics of sensing segment which is having $M$ sensing vectors and $L$ (smoothing factor) subsequent output sample vectors in DT domain as:

$$\mathrm{x}_q = [x[q], x[q+1], \ldots, x[q-L+1]]^{\mathrm{T}} \tag{11.3}$$

where the symbol $(.)^{\mathrm{T}}$ is the transpose operation of (.) vector/matrix.

For large $M$, the Cov matrix of received sample can be approximated as:

$$\mathbf{C}_{xx} = \frac{1}{M} \sum_{q=1}^{M} \mathrm{x}_q \mathrm{x}_q^{\mathrm{T}} \tag{11.4}$$

Therefore, the Cov matrix of received sample of $i$th sub-segment can be obtained as:

$$\mathbf{C}_{xx,i} = \frac{1}{M} \sum_{q=(i-1)M+1}^{(i-1)M+M} \mathrm{x}_q \mathrm{x}_q^{\mathrm{T}} \tag{11.5}$$

Similarly, the Cov matrix of sample transmitted is given as:

$$\mathbf{C}_{\mathrm{ss}} = \frac{1}{M} \sum_{q=1}^{M} \mathrm{s}_q \mathrm{s}_q^{\mathrm{T}} \tag{11.6}$$

and, the Cov matrix of noise sample is given as:

$$\mathbf{C}_{\mathrm{ww}} = \frac{1}{M} \sum_{q=1}^{M} \mathrm{w}_q \mathrm{w}_q^{\mathrm{T}} \tag{11.7}$$

To evaluate the performance of SS, two probabilities are of interest as detection probability, $P_{\mathrm{d}} = P(H_1|x[n] = s[n] + w[n])$, and the probability of false alarm $P_{\mathrm{f}} = P(H_1|x[n] = w[n])$. In CR, the $P_{\mathrm{f}}$ associates with the spectral utilization, and $P_{\mathrm{d}}$ associates with the interference to PU (Hou and Qiu 2014).

## 11.3 Spectrum Sensing Techniques for White Gaussian Noise

SS algorithms for white Gaussian noise (WGN) can be broadly categorized into two types, i.e., (a) blind, and (b) non-blind.

### *11.3.1 Blind Sensing Algorithms*

In BSS, the apriori information of some attributes of signal of PU is not required. BSS can further be categorized into Cov matrix, ED, and Egnval-based algorithms. From the aforementioned techniques, the Egnval-based algorithm depends on the Cov matrix of the samples received. Various BSS techniques are discussed below.

#### 11.3.1.1 Energy Detection

ED is one the most simple techniques for detection of the presence of PU's signal. However, at low signal-to-noise ratio (SNR), performance of ED deteriorates. The key idea behind this technique is that it computes the energy of the received $N_s$ samples as:

$$Z_{ED} = \frac{1}{N_s} \sum_{q=1}^{N_s} |x[i]|^2 \tag{11.8}$$

The probability density function (PDF) of the computed energy $Z_{ED}$ is known to follow Chi-squared ($\chi^2$) distribution with $N_s$ degree of freedom. Further, it can be approximated by the Gaussian distribution with $\mu_{ED}$ mean and $\sigma^2_{ED}$ variance. Threshold is calculated under $H_0$ on the basis of PDF of $Z_{ED}$ and $P_f$ as (Sobron et al. 2015):

$$Z_{ED} = Q^{-1}(P_f)\sigma_{ED} + \mu_{ED} \tag{11.9}$$

where $Q^{-1}(\cdot)$ is the inverse $Q$ function and $\lambda_{ED}$ is the detection threshold. PU's signal exits if $Z_{ED} > \lambda_{ED}$.

#### 11.3.1.2 Three-Event Energy Detection

This detection technique is used for PUs with duty cycle. Duty cycle is defined as the time in which an active PU is present to the total sensing time of detection of PUs. In this technique, three consecutive sensing slots energy, $E_{q-1}$, $E_q$, $E_{q+1}$ are estimated. If the energy of the first slot, i.e., $E_{q-1}$, is found greater than some predefined threshold then the PU is considered to be active. If the energy of the first slot is less than the threshold, then estimate the energy of the second slot ($E_q$). If the energy of the second slot is greater than the threshold, then PU exists; otherwise, estimate the energy in the third slot. If energy of all the three slots is less than the threshold, then the sensing band is available for SUs. The computational complexity of this method is greater than the ED.

Threshold of three event energy detection (3EED) under the hypothesis $H_0$ on the basis of $P_f$ and pdf is given as (Vladeanu et al. 2016):

$$\lambda_{3EED} = [Q^{-1}(1 + \sqrt[3]{P_f - 1})^2\sqrt{2N_s} + N_s]\sigma_w^2 \tag{11.10}$$

where $\lambda_{3EED}$ is detection threshold and $\sigma_w^2$ is the noise variance.

#### 11.3.1.3 Covariance Absolute Value

In this technique, the statistical Cov matrix of received DT samples is obtained. Next, from the Cov matrix, the two test statistics are evaluated as (Zeng and Liang 2009):

$$Z_{\text{CAV1}} = \sum_{x=1}^{M} \sum_{y=1}^{M} |r_{xy}| \tag{11.11}$$

$$Z_{\text{CAV2}} = \sum_{x=1}^{M} |r_{xx}| \tag{11.12}$$

where $r_{xy}$ is the element of $\mathbf{C}_{xx}$. The detection threshold $\lambda_{\text{CAV}}$ under $H_0$ is given as:

$$\lambda_{\text{CAV}} = \frac{1 + (L-1)\sqrt{\frac{2}{N_s \pi}}}{1 - Q^{-1}(P_f)\sqrt{\frac{2}{N_s}}} \tag{11.13}$$

It is observed from the Eq. (11.13) that the detection threshold for CAV method is independent of noise variance. The PUs signal exists if $Z_{\text{CAV1}}/Z_{\text{CAV2}} > \lambda_{\text{CAV}}$.

#### 11.3.1.4 Maximum Eigenvalue Detection

In maximum ED (MED) signal detection technique, the maximum Egnval of Cov matrix of the received sample is used for the detection of active PU's signal. Basically, to analysis of the method and calculation of detection threshold is performed by random matrix theory. The test statistics can be calculated as (Zeng et al. 2008):

$$Z_{\text{ME}} = \gamma_{\max} \tag{11.14}$$

where $\gamma_{\max}$ is the maximum eigenvalue of $\mathbf{C}_{xx}$. The detection threshold $\lambda_{\text{ME}}$, under $H_0$ is calculated as:

$$\lambda_{\text{ME}} = \frac{(\sqrt{N_s} + \sqrt{L})^2}{N_s}\left(1 + \frac{(\sqrt{N_s} + \sqrt{L})^{-2/3}}{(N_s L)^{1/6}} F_1^{-1}(1 - P_f)\right) \tag{11.15}$$

where $F_1$ is the cumulative distribution function (CDF) of order-1 Tracy-Widom distribution. The maximum eigenvalue of Cov matrix follows order 1 and order 2 Tracy-Widom distribution for real and complex data, respectively. The PU's signal exists if $Z_{\text{ME}} > \lambda_{\text{ME}} \sigma_w^2$.

#### 11.3.1.5 Max. to Min. Eigenvalue Detection

In max. to min. Egnval (MME)-based technique, the detection is performed by finding the ratio of maximum Egnval to minimum Egnval of Cov matrix of the received sample. In this technique, the test statistics is calculated as (Zeng and Liang 2009):

$$Z_{\mathrm{MME}} = \frac{\gamma_{\max}}{\gamma_{\min}} \tag{11.16}$$

where $\gamma_{\min}$ is the minimum eigenvalue of $\mathbf{C}_{xx}$. The detection threshold $\lambda_{\mathrm{MME}}$, under $H_0$ is calculated as:

$$\lambda_{\mathrm{MME}} = \frac{(\sqrt{N_{\mathrm{s}}} + \sqrt{L})^2}{(\sqrt{N_{\mathrm{s}}} - \sqrt{L})^2}\left(1 + \frac{(\sqrt{N_{\mathrm{s}}} + \sqrt{L})^{-2/3}}{(N_{\mathrm{sL}})^{1/6}} F_1^{-1}(1 - P_{\mathrm{f}})\right) \tag{11.17}$$

It is observed from the Eq. (11.17) that the detection threshold for MME method is independent of noise variance, and therefore, it is more robust than ED- and ME-based signal detection. The PU's signal exists if $Z_{\mathrm{MME}} > \lambda_{\mathrm{MME}}$.

#### 11.3.1.6 Energy with MED

The ratio of the energy of received samples to the minimum Egnval of Cov matrix of received sample is used for signal detection in energy with minimum Egnval (EME). In this technique, the test statistics can be calculated as (Zeng and Liang 2009):

$$Z_{\mathrm{EME}} = \frac{Z_{\mathrm{ED}}}{\gamma_{\min}} \tag{11.18}$$

The detection threshold $\lambda_{\mathrm{EME}}$, under $H_0$ is calculated as:

$$\lambda_{\mathrm{EME}} = \left(\sqrt{\frac{2}{N_{\mathrm{s}}}} Q^{-1}(P_{\mathrm{f}}) + 1\right)\frac{N_{\mathrm{s}}}{(\sqrt{N_{\mathrm{s}}} - \sqrt{L})^2} \tag{11.19}$$

Again, it is observed from the Eq. (11.19) that the detection threshold for EME method is independent of noise variance, and therefore, it is more robust than ED- and ME-based signal detection under noise uncertainty. The PU's signal exists if $Z_{\mathrm{EME}} > \lambda_{\mathrm{EME}}$.

#### 11.3.1.7 Arithmetic to Geometric Mean Detector

In this type of signal detection, a likelihood ratio is used to find an unstructured estimate of $\mathbf{C}_{xx}$. In particular, this detector computes the arithmetic-to-geometric (AGM) of eigenvalues of covariance matrix of received sample which is used for signal detection. The test statistics is calculated as (Font-Segura and Wang 2010):

$$Z_{\mathrm{AGM}} = \frac{\frac{1}{L}\sum_{l=1}^{L} \gamma_l}{(\prod_{l=1}^{L} \gamma_l)^{1/L}} \tag{11.20}$$

where $\gamma_l$ are the eigenvalues of $\mathbf{C}_{xx}$. The threshold $\lambda_{\text{AGM}}$ is obtained with the help of PDF of $Z_{\text{AGM}}$ and given $P_{\text{f}}$ under $H_0$. The PU's signal exists if $Z_{\text{AGM}} > \lambda_{\text{AGM}}$. Further, the detection threshold of AGM is independent of noise variance.

#### 11.3.1.8 Scaled Largest Eigenvalue (SLE) Detection

SLE signal detection is based on the computation of the scaled version of the maximum eigenvalue of covariance matrix of received sample. The detection performance of this scheme is better than that of the MME and EME as shown in (Wei and Tirkkonen 2011). The test statistics is given as (Wei and Tirkkonen 2011):

$$Z_{\text{SLE}} = \frac{\gamma_{\max}}{\text{Tr}(\mathbf{C}_{xx})} \tag{11.21}$$

where $\text{Tr}(\cdot)$ is trace, i.e., sum of all the diagonals of $(\cdot)$. The detection threshold $\lambda_{\text{SLE}}$ is calculated under $H_0$ for the specified $P_{\text{f}}$ as:

$$1 - H(u) = P_{\text{f}} \tag{11.22}$$

where $H(u)$ is the CDF of $Z_{\text{SLE}}$ with a polynomial in the variable $u$. The PU's signal exists if $Z_{\text{SLE}} > \lambda_{\text{SLE}}$.

#### 11.3.1.9 Function of Matrix Based Detection (FMD)

In FMD, monotonically increasing property of trace operation on matrix is used for signal detection. The test statistics is calculated as (Lin et al. 2012):

$$Z_{\text{FMD}} = \text{Tr}\Big(\frac{1}{K}\sum_{q=1}^{K}\mathbf{C}_{xx,q}\Big) \tag{11.23}$$

The detection threshold $\lambda_{\text{FMD}}$, under $H_0$ for the specified $P_{\text{f}}$ is calculated as:

$$\lambda_{\text{FMD}} = \Big(1 + Q^{-1}(P_{\text{f}})\sqrt{\frac{2}{N_{\text{sK}}}}\Big)L\sigma_w^2 \tag{11.24}$$

The PU's signal exists if $Z_{\text{FMD}} > \lambda_{\text{FMD}}$.

#### 11.3.1.10 Feature Template Matching

The signal detection, in this, is performed by using the dominant Eigenvector of the Cov matrix of both the transmitted and the received samples. Under this scheme, the

test statistics is calculated as (Zhang and Qiu 2013):

$$Z_{\text{FTM}} = |\varphi_{ss,1}^{T} \varphi_{xx,1}| \tag{11.25}$$

where $|t|$ represents the absolute value of $t$, and $\varphi_{ss,1}$ and $\varphi_{xx,1}$ are the dominant Eigenvector of $\mathbf{C}_{\text{ss}}$ and $\mathbf{C}_{xx}$, respectively. The detection threshold $\lambda_{\text{FTM}}$, with the help of given $P_{\text{f}}$ and PDF of $Z_{\text{FTM}}$ under $H_0$ is calculated. The PU's signal exists if $Z_{\text{FTM}} > \lambda_{\text{FTM}}$.

#### 11.3.1.11 Principal Component Analysis

It PCA, the dimensionality of a data set is reduced which consists of a large number of variables, while preserving the variation present in the data set as much as possible (Jolliffe 1986). It is attained by transformation to a new set of variables which is known as principal components (PCs) that are not correlated and are ordered such that the initial few PCs preserve significant amount of the data variation as compared with the rest of the PCs (Bhatti et al. 2012). In this scheme, the test statistics consists of the sum of the $P$ largest Egnvals of $\mathbf{C}_{xx}$ as (Bhatti et al. 2012):

$$Z_{\text{PCA}} = \frac{1}{P} \sum_{p=1}^{P} \gamma_p \tag{11.26}$$

where $\gamma_1 \geq \gamma_2 \ldots \geq \gamma_p$. The threshold $\lambda_{\text{PCA}}$, with the help of given $P_{\text{f}}$ and PDF of $Z_{\text{PCA}}$ is calculated under $H_0$. The PU's signal exists if $Z_{\text{PCA}} > \lambda_{\text{PCA}}$.

#### 11.3.1.12 Grassmannian Manifold-Based Spectrum Sensing

This method is based on the canonical correlation analysis between the two data sets. These data sets are formed with the help of the principal Eigenspace of $\mathbf{C}_{\text{ss}}$ and $\mathbf{C}_{xx}$ (Bishnu and Bhatia 2018). This principal Eigenspace is also known as Grassmannian covariance matrix (GCM). With the help of GCMs, a test statistics is defined which is used for signal detection as:

$$Z_{\text{GCM}} = \left( \prod_{i=1}^{r} \cos^2 \theta_i \right)^{1/2} \tag{11.27}$$

where $\theta_i$ are the principal angles between two GCMs and $r$ is the first $r$ column of GCMs. The threshold $\lambda_{\text{GCM}}$, with the help of given $P_{\text{f}}$ and PDF of $Z_{\text{GCM}}$ is calculated under $H_0$ as:

$$\lambda_{\text{GCM}} = Q^{-1}(P_{\text{f}})\sigma + \mu \tag{11.28}$$

where $\sigma^2$ and $\mu$ are the variance and mean of $Z_{\text{GCM}}$ respectively. The PU's signal exists if $Z_{\text{GCM}} > \lambda_{\text{GCM}}$.

### *11.3.2 Non-Blind Spectrum Sensing Algorithms*

In NBSS method, receiver has knowledge of the signal transmitted from the users or its attribute in advanced. NBSS algorithms comprise of pilot signal, PSD, CP, and cyclo-stationary-based technique for orthogonal frequency division multiplexing (OFDM) system. The NBSS algorithms are given as:

#### 11.3.2.1 Pilot Tones

This SS algorithm is build on the time-domain symbol cross-correlation (TDSC) of two OFDM symbols. The TDSC algorithm uses the property that there is a non-zero constant term inserted in the TDSC if the two OFDM symbols have the same frequency-domain pilot symbols (Chen et al. 2009). Thus, the proposed SS algorithm can be applicable to any OFDM system which allocates pilot symbols in frequency domain. The TDSC algorithm consists of two methods as:

- **Neyman-Pearson Method** In this method, the signal detection test statistics is calculated as (Chen et al. 2009):

$$Z_{\text{NPPT}} = |C(u)| \tag{11.29}$$

  where $u$ is the symbol index difference between two OFDM symbols and $C(u)$ is the accumulated TDSC function. $C(u)$ is represented as (Kim et al. 2007):

$$H_0 : C(u) = \xi(u) \tag{11.30}$$

$$H_1 : C(u) = \theta(u)\Lambda + \xi(u) \tag{11.31}$$

  where $\theta(u)$ represents a phase rotation of carrier frequency offset, $\Lambda$ is the average pilot tone received signal power divided by $N^2$ ($N$ is the FFT size of OFDM system), and $\xi(u)$ follows the standard normal distribution.

- **Maximal Ratio Combining (MRC) Method** The signal detection test statistics of MRC method is calculated as (Chen et al. 2009):

$$Z_{\text{MRCPT}} = \left| \sum_{u} S(u)S(u+d)R(u, u+d) \right| \tag{11.32}$$

where $S(u)$ represents the number of TDSC function of two OFDM symbols that are accumulated and added, and $R(u, u+d)$ represents the conjugate product of these accumulated functions. The detection threshold $\lambda_{\text{MRCPT}}$, under $H_0$ for a particular $P_\text{f}$ is calculated as:

$$\lambda_{\text{MRCPT}} = \sqrt{-\sigma_{H_0}^2 \ln(P_\text{f})} \tag{11.33}$$

The PU's signal exists if $Z_{\text{MRCPT}} > \lambda_{\text{MRCPT}}$.

#### 11.3.2.2 Cyclic Prefix

In CP signal detection, the correlation of the two $N$ samples apart pieces of the received samples is used for signal detection. The test statistics for this method is calculated as (Lei and Chin 2010):

$$H_0 : Z_{\text{CP}} = \sum_{d=1}^{W} \frac{w[d]w^*[N+d]}{|x[d]|^2} \tag{11.34}$$

$$H_1 : Z_{\text{CP}} = \sum_{d=1}^{W} \frac{x[d]x^*[N+d]}{|x[d]|^2} \tag{11.35}$$

where $(\cdot)^*$ represents the conjugate of $(\cdot)$ and $W$ represents the observation window. The detection threshold $\lambda_{\text{CP}}$, under $H_0$ for a particular $P_\text{f}$ can be calculated as:

$$\lambda_{\text{CP}} = \frac{W}{2} P^{-1}_{k=2,\lambda_0=\frac{2c^2}{W}}(1 - P_\text{f}) \tag{11.36}$$

where $P^{-1}_{k,\lambda_0}(t)$ denotes the inverse of the non-central chi-square CDF with $k$-degrees of freedom and non-centrality parameter $\lambda_0$ at a particular $P_\text{f}$. The PU's signal exists if $Z_{\text{CP}} > \lambda_{\text{CP}}$.

#### 11.3.2.3 Joint Sensing Method

The detection in this method is based on CP, and PT inserted in OFDM signals (Chen et al. 2012). The test statistic for this method is calculated as (Chen et al. 2012):

$$Z_{\text{CPPT}} = \frac{\mu_{1,R}}{\sigma^2_{0,R}} \Re(R) + \frac{\mu_{1,C(u)}}{\sigma^2_{0,C(u)}} \Re(C(u)) \tag{11.37}$$

where $\mu_{1,R}$ represents the mean value of cross-correlation $(R)$ of received samples under $H_1$, $\sigma^2_{0,R}$ is the variance of $R$ under $H_0$, $\mu_{1,C(u)}$ is the mean of $C(u)$ under $H_1$,

$\sigma^2_{0,C(u)}$ is the variance of $C(u)$ under $H_0$ ($C(u)$ is test statistics for pilot tones bases SS), and $\Re(t)$ is the real part of $(t)$. The detection threshold $\lambda_{\text{CPPT}}$, under $H_0$ for a particular $P_\text{f}$ is calculated as:

$$\lambda_{\text{CPPT}} = Q^{-1}(P_\text{f})\sqrt{\sigma^2_{0,Z_{\text{CPPT}}}} \tag{11.38}$$

where $\sigma^2_{0,Z_{\text{CPPT}}}$ represents the variance of $Z_{\text{CPPT}}$ under $H_0$. The PU's signal exists if $Z_{\text{CPPT}} > \lambda_{\text{CPPT}}$.

#### 11.3.2.4 Periodogram Based Method

In this method, correlation of received signal's periodogram with the a priori spectral attributes of the PU's signal is used for signal detection. The test statistic for this method is calculated as (Quan et al. 2009) as:

$$Z_{\text{PG}} = \frac{1}{N}\sum_{k=1}^{N} S_x^N(k) S_\text{s}^N(k) \tag{11.39}$$

where $S_x^N(k)$ and $S_\text{s}^N(k)$ are the power spectral density of the received signal and transmitted PU's signal, respectively. Next, the detection threshold $\lambda_{\text{PG}}$ is based on the PDF of $Z_{\text{PG}}$ at a particular $P_\text{f}$ under $H_0$. The PU's signal exists if $Z_{\text{PG}} > \lambda_{\text{PG}}$.

#### 11.3.2.5 Cyclo-Stationary

Cyclo-stationary is the property of signal which varies periodically with time.

- **Spectral Correlation (SC)**
  In SC method, the correlation of the received signal is obtained by searching over the complete bandwidth which is given as (Han et al. 2006):

$$S(f)_{\Delta t} = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} \frac{1}{T} X_T\left(t, f + \frac{\alpha}{2}\right) X_T^*\left(t, f - \frac{\alpha}{2}\right) \text{d}t \tag{11.40}$$

  where $T$ is the cyclo-stationary period, $\alpha$ is the cyclic frequency parameter, $\Delta t$ represents the interval time, and $X_T(t_1, t_2)$ is calculated as (Lunden et al. 2007):

$$X_T(t_1, t_2) = \int_{t_1 - T/2}^{t_1 + T/2} x(u) \text{e}^{-j2\pi u t_2} du \tag{11.41}$$

The above correlation function is used to search for the unique cyclic frequencies. The PU's signal exists if unique cyclic frequencies are detected.

- **Multiple Cyclic Frequency** The key idea behind this method is that a signal that is cyclo-stationary at symbol frequency is also (typically) cyclo-stationary at all the harmonics of symbol frequency. The test statistic is calculated as (Lunden et al. 2007):

$$Z_{\text{MCF}} = \max_{\alpha \in G} r_{xx}(\alpha) \Sigma_{xx}^{-1}(\alpha) r_{xx}^{T}(\alpha) \tag{11.42}$$

where $G$ represents the set of cyclic frequency, $r_{xx}(\alpha)$ is calculated as:

$$r_{xx}(\alpha) = [\Re\{R_{xx}(\alpha, \tau_1)\}, \ldots, \Re\{R_{xx}(\alpha, \tau_N)\}, \Im\{R_{xx}(\alpha, \tau_1)\}, \ldots, \Im\{R_{xx}(\alpha, \tau_N)\}] \tag{11.43}$$

$R_{xx}(\alpha, \tau)$ is given as:

$$R_{xx}(\alpha, \tau) = \frac{1}{M} \sum_{t=1}^{M} x(t) x^*(t + \tau) \mathrm{e}^{-j2\pi\alpha t} \tag{11.44}$$

$\Sigma_{xx}(\alpha)$ is given as:

$$\Sigma_{xx}(\alpha) = \begin{bmatrix} \Re\left\{\frac{Q+Q^*}{2}\right\} & \Im\left\{\frac{Q-Q^*}{2}\right\} \\ \Im\left\{\frac{Q+Q^*}{2}\right\} & \Re\left\{\frac{Q-Q^*}{2}\right\} \end{bmatrix} \tag{11.45}$$

where the $(u, v)$th entry of Cov matrix is given as:

$$Q(u, v) = S_{f\tau_u f\tau_v}(2\alpha, \alpha) \tag{11.46}$$

$$Q^*(u, v) = S^*_{f\tau_u f\tau_v}(0, -\alpha) \tag{11.47}$$

where $S_{f\tau_u f\tau_v}$ denotes the unconjugated and $S^*_{f\tau_u f\tau_v}$ denotes the conjugated cyclic spectra of $f(t, \tau) = x(t)x^*(t + \tau)$. The threshold $\lambda_{\text{MCF}}$, under $H_0$ is obtained using a specific value of $P_{\text{f}}$ and the PDF of $Z_{\text{MCF}}$. The PU's signal exists if $Z_{\text{MCF}} > \lambda_{\text{MCF}}$.

- **Cycle Frequency Domain Profile** The signal detection and pre-processing, in this method, for signal detection is performed by using the cycle frequency domain profile (CFDP) (Kim et al. 2007). Threshold-test method is used for extracting the signal features from CFDP. Hidden Markov model is used to process extracted signal attribute for classification. The test statistics is calculated as (Kim et al. 2007):

$$Z_{\text{CFDP}} = \frac{\max(I(\alpha))}{\sqrt{\frac{\sum_{\alpha=0}^{M-1} I^2(\alpha)}{M}}} \tag{11.48}$$

where $I(\alpha) \triangleq \max_f |C_x^\alpha(f)|$ and $C_x^\alpha(f) = \frac{S_x^\alpha(f)}{\sqrt{S(f+\frac{\alpha}{2})S(f-\frac{\alpha}{2})}}$ and $S_x^\alpha(f)$ represents the spectral coherence and the spectral correlation functions, respectively. The threshold $\lambda_{\text{CFDP}}$, under $H_0$ is obtained with the help of a particular $P_\text{f}$ and PDF of $Z_{\text{CFDP}}$. The PU's signal exists if $Z_{\text{CFDP}} > \lambda_{\text{CFDP}}$.

## 11.4 Spectrum Sensing Algorithms for Colored Gaussian Noise

The performance of the algorithms discussed in Sect. 11.3 is good in AWGN scenario. However, signal detection of most of the techniques deteriorates for the CGN scenario. The techniques such as CP, PT, and cyclo-stationary which are typically based on the cross-correlation function perform good at low SNR for the WGN but the performance of the NBSS algorithms deteriorates at low SNR for the correlated samples due to the colored noise. Similarly, performance of eigenvalue of Cov matrix of the received sample-based techniques is good at low SNR for uncorrelated samples or AWGN. However, the performance of Egnval-based techniques deteriorates and sometimes fails to detect the presence of PU's signal for correlated signals because of the swelling effect of Cov matrix (Horev et al. 2017. Due to filtering and oversampling, the received samples are correlated (Sharma et al. 2013). It is worth noting that passing the received signal at the input of the receiver through a filter leads to correlated samples of noise due to filter's transfer function. Similarly, analog-to-digital converter (ADC) at receiving side samples received CT signal at the rate higher than the Nyquist rate which leads to correlation in the output (Sharma et al. 2013). There are few methods based on Cov matrix of the received signal for the correlated samples as follows:

### *11.4.1 Standard Condition Number (SCN)*

In this method, SCN is used for signal detection. SCN of a matrix is defined as the ratio of maximum Egnval to minimum Egnval. The test statistic is calculated as (Sharma et al. 2013):

$$Z_{\text{SCN}} = \frac{\tilde{b}}{\tilde{a}} \tag{11.49}$$

where

$$\tilde{b} = 1 + \beta + 2\alpha\beta + 2\sqrt{\beta}\sqrt{(1+\alpha)(1+\alpha\beta)} \tag{11.50}$$

and

$$\tilde{a} = 1 + \beta + 2\alpha\beta - 2\sqrt{\beta}\sqrt{(1+\alpha)(1+\alpha\beta)} \tag{11.51}$$

$\beta < 1, \alpha = (\sqrt{\sigma_2} - \sqrt{\sigma_1})^2/4\sigma_1\sigma_2$, $\sigma_1 = (1-\rho)/(1+\rho)$, $\sigma_2 = 1/\sigma_1$, and $\rho(< 1)$ is the correlation coefficient of colored noise. The threshold $\lambda_{\text{SCN}}$, under $H_0$ is obtained with the help of particular $P_{\text{f}}$ and PDF of $Z_{\text{SCN}}$. The PU's signal exists if $Z_{\text{SCN}} > \lambda_{\text{SCN}}$.

### 11.4.2 Maximum Eigenvalue

In this detection, maximum Egnval of Cov matrix of the received sample in the presence of colored noise is used for signal detection. The test statistic is given by Sharma et al. (2014):

$$Z_{\text{MECN}} = \gamma_{\max}^{c} \tag{11.52}$$

where $\gamma_{\max}^{c}$ is the maximum eigenvalue of $\mathbf{C}_{xx}$ in the presence of colored noise. The threshold $\lambda_{\text{MECN}}$, under $H_0$ is obtained with the help of particular $P_{\text{f}}$ as:

$$\lambda_{\text{MECN}} = \frac{\sigma_c F_1^{-1}(1-P_{\text{f}}) + N_{\text{s}}^{2/3}\nu_c}{N_{\text{s}}^{2/3}} \tag{11.53}$$

where

$$\nu_c = \frac{1}{c}\left(1 + \frac{c}{2\pi\mu\delta}\int_{\sigma_1}^{\sigma_2} \frac{1}{(1-\gamma c)\gamma}\sqrt{-\gamma^2 - 1 + 2\gamma\mu_1}d\gamma\right) \tag{11.54}$$

and

$$\sigma_c^3 = \frac{1}{c^3}\left(1 + \frac{c^3}{2\pi\mu\delta}\int_{\sigma_1}^{\sigma_2} \frac{1}{(1-\gamma c)^3}\sqrt{-\gamma^2 - 1 + 2\gamma\mu_1}d\gamma\right) \tag{11.55}$$

here, $c \in [0, 1/\gamma_{\max}^{c}]$, $\delta > 1$, $\sigma_1 \leq \gamma \leq \sigma_2$, and $\mu_1 = (1+\rho^2)/(1-\rho^2)$. The PU's signal exists if $Z_{\text{MECN}} > \lambda_{\text{MECN}}$.

### 11.4.3 LogDet Covariance

In the presence of colored noise, the Cov matrix of received signal may have swelling effect due to spurious variation to the signal, and hence Euclidean geometry is unsuitable for signal detection (Bishnu and Bhatia 2018). Two algorithms based on LogDet of Cov matrix of received samples are proposed in (Bishnu and Bhatia 2018).

#### 11.4.3.1 LogDet of Received Samples Covariance Matrix

The test statistics of LogDet of Cov matrix of received samples (LDRSCM) for signal detection is calculated as:

$$Z_{\text{LDRSCM}} = \text{logdet}\Big(\frac{1}{K}\sum_{q=1}^{K}\mathbf{C}_{xx,q}\Big) \tag{11.56}$$

The detection threshold $\lambda_{\text{LDRSCM}}$, under $H_0$ is obtained with the help of particular $P_{\text{f}}$ as:

$$\lambda_{\text{LDRSCM}} = F_1^{-1}(1-P_{\text{f}})\frac{\rho^2|\log(|\mu_{\text{TW1}}|)|}{\sqrt{N_{\text{s}}}+\sqrt{L}}\Big(\frac{1}{\sqrt{N_{\text{s}}}}+\frac{1}{\sqrt{L}}\Big)^{1/3}+\rho L\log(\sigma_w^2) \tag{11.57}$$

where $\mu_{\text{TW1}}$ is the mean of Tracy-Widom distribution of order 1. The PU's signal exists if $Z_{\text{LDRSCM}} > \lambda_{\text{LDRSCM}}$.

#### 11.4.3.2 Energy with LogDet of Received Samples Covariance Matrix

The detection threshold of LDRSCM depends on the noise variance. Hence, Energy with LogDet of received samples covariance matrix (ELDRSCM) is proposed whose test statistics is defined for signal detection as:

$$Z_{\text{ELDRSCM}} = \frac{\log(Z_{\text{ED}})}{Z_{\text{LDRSCM}}} \tag{11.58}$$

The detection threshold $\lambda_{\text{ELDRSCM}}$, under $H_0$ is obtained with the help of particular $P_{\text{f}}$ as:

$$\lambda_{\text{ELDRSCM}} = \frac{\rho\sqrt{N_{\text{sL}}}+Q^{-1}(P_{\text{f}})}{\rho^2 L\sqrt{N_{\text{sL}}}} \tag{11.59}$$

The PU's signal exists if $Z_{\text{ELDRSCM}} > \lambda_{\text{ELDRSCM}}$.

## 11.5 Cooperative Sensing

SS is highly vulnerable to noise variance, fading, and shadowing. To overcome this problem, cooperative SS (CSS) is proposed in the literature. The CSS reduces the missed detection and false alarm probabilities substantially (Yucek and Arslan 2009). Moreover, the cooperation also solves the hidden user (primary) problem and reduces sensing time which result in the complexity reduction (Ganesan and Li 2005). Developing efficient and effective information exchange algorithms and reduction in

the complexity are the key challenges of SS (Guo et al. 2006). Cooperative SS is most efficient when independent fading or shadowing influence each cooperative CR. Liu and Sai (2006). In (Peh and Liang 2007), it is shown that coordination with all the users in the network does not achieve the optimal signal detection and cognitive users having strong PU's SNR are chosen for cooperation. CSS can be accomplished in three strategies as; (a) centralized sensing (Cen-S), (b) distributed sensing (Dis-S), and (c) external sensing (Ext-S).

### *11.5.1 Centralized Sensing*

In Cen-S, a central unit, also known as access point or controller, collects the sensing information from cognitive or sensing devices, identifies the available spectrum based on the information, and relays this information to all cognitive devices. There are various methods which are applied by the central unit for signal detection by using collected information as:

#### 11.5.1.1 Signal-to-Noise Ratio

In SNR method (Visotsky et al. 2005), the presence of active PU's signal is detection by using the observed SNR value at each node $U_i$, for $i = 1, 2, \ldots, n$, where n represents the number of sensing devices. Here, the SS problem can be formulated as:

$$H_0 : \text{SNR}_U \sim \mathcal{N}(\mu_0 \times \mathbf{1}, \sigma_w^2 \Sigma) \tag{11.60}$$

$$H_1 : \text{SNR}_U \sim \mathcal{N}(\mu_1 \times \mathbf{1}, \sigma_w^2 \Sigma) \tag{11.61}$$

where $\text{SNR}_U$ is the vector of $\text{SNR}_{U_i}$ having dimension $n \times 1$, $\mathbf{1}$ is an $n \times 1$ vector with all ones, $\Sigma = \overline{\Sigma}/\sigma_{\mathbf{w}}^{\mathbf{2}}$ (where $\overline{\Sigma}$ is the Cov matrix of the received samples), $\mu_0 = \mu(R)$ and $\mu_1 = \mu(R + \delta)$ with $\delta > 0$, where $R$ represents the distance from the PU transmitter where there is no interference by SU transmitter to PU receivers). The test statistics for the method is given as:

$$Z_{\text{SNR}} = \sum_{i=1}^{n} B_i \tag{11.62}$$

if $i$th SU detects the PU;s signal then $B_i = 1$, otherwise 0. The detection threshold $\lambda_{\text{SNR}}$, under $H_0$ is calculated with the help of given $P_{\text{f}}$. The information of each $B_i$ is sent to central unit for deciding the presence of PU's signal by central unit.

#### 11.5.1.2 Cluster Based Detection

In CSS, each sensing device sends its received information to the control unit over the fading channel. This deteriorates the performance of signal detection. Cluster-based SS (Clu-ss) enhances the signal detection performance by choosing all the sensing devices into a few clusters and chose the most favorable sensing device in each cluster to communicate to the control unit (Sun et al. 2007). The Clu-SS is implemented as follows (Sun et al. 2007):

- First, very sensor $m$ in cluster $n$ collects energy $E_{n,m}$ and sends observation $O_{n,m}$ to the cluster head, where $O_{n,m}$ is some function $\Omega$ of $E_{n,m}$.
- Next, each cluster head receives information and makes a decision $C_n$ on the basis fusion function $\Phi$.
- Finally, the cluster decision $C_n$ is communicated to the control unit, and last decision of whether the PU's signal is present or not is made on the basis of some fusion function $\varphi$.

### *11.5.2 Distributed Sensing*

In Dis-S, sensing devices exchange their received informations with each other devices. Dis-S does not require the backbone structure, and it also yields the reduced cost, hence it is more advantageous than centralized sensing (Yucek and Arslan 2009). A sensing device receives information from rest of the devices or SU's and decides the presence of PU's signal if any of the devices detect the active PUs. This fusion rule is known as OR-rule (logical OR operator) (Ghasemi and Sousa 2005).

### *11.5.3 External Sensing*

In Ext-S, the information of availability of spectrum is collected by an external agent or device and relay this information to the other sensing devices. Ext-S is used to solve hidden PUs problem and the uncertainty of signal detection due to shadowing and fading. In addition, as SU devices achieves savings in time, and hence, leads to increased in spectrum efficiency.

## 11.6 Results and Discussion

From all the above various techniques for SS that have been discussed so far, the most popular SS are Cov and Egnval based SS. In this section, the performance of the techniques based on Cov and Egnval of Cov matrix of the received sample is analyzed

.PU's signal is real DTV signal (Tawil 2006) captured in Washington, DC, USA. For all results, $L = 10$, $K = 10$, $M = 500$, $N_s = 50000$ are considered. Both WGN and CGN are added to attain various SNR levels to the DTV signal (Shellhammer et al. 2006).

Figure 11.1 shows the $P_d$ versus SNR of blind techniques for WGN at $P_f = 0.1$. It is perceived from Fig. 11.1 that the FMD-based detection has best performance and ED has worst performance at low SNR in terms of PU's signal detection. PCA- and ME-based SS achieves also good performance with marginal performance deterioration (3–4 dB) as compared to FMD for $P_d = 0.9$. GCM-based SS has improved performance as compared to the other of the algorithms except FMD, PCA, and ME. The signal detection of other algorithms except that the FTM and EME are same with high performance deterioration (9 dB) as compared to FMD. The performance degradation of FTM and EME is same which is 11 dB as compared to FMD. Figure 11.2 revels the $P_d$ versus $P_f$ of BSS techniques for WGN at -20 dB SNR. Also, it is perceived from Fig. 11.2 that FMD, PCA, ME, and GCM achieve approximately similar performance and better ROC than the other SS techniques. It is because they are able to detect signal at very low $P_f$.

Figure 11.3 revels the $P_d$ versus SNR of BSS techniques for CGN at $P_f = 0.1$. It is perceived from Fig. 11.3 that the signal detection performance of LDRSCM-based detection is much better than rest of the techniques. There are various techniques including CAV, ME, MME, and SLE which are not able to detect the presence of PU's signal. LDRSCM provides 8 dB gain as compared to FMD for $P_d = 0.9$. For the colored noise, MECN and SCN have much better performance as compared with

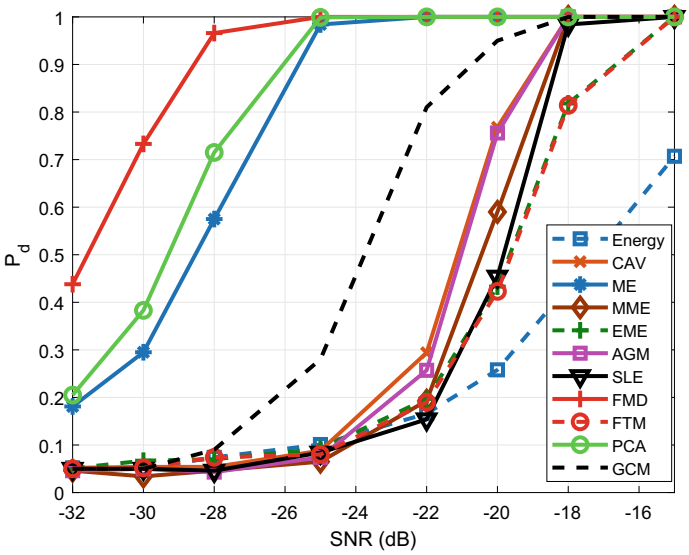



**Fig. 11.1** $P_d$ versus SNR for BSS technique for WGN at $P_f = 0.1$

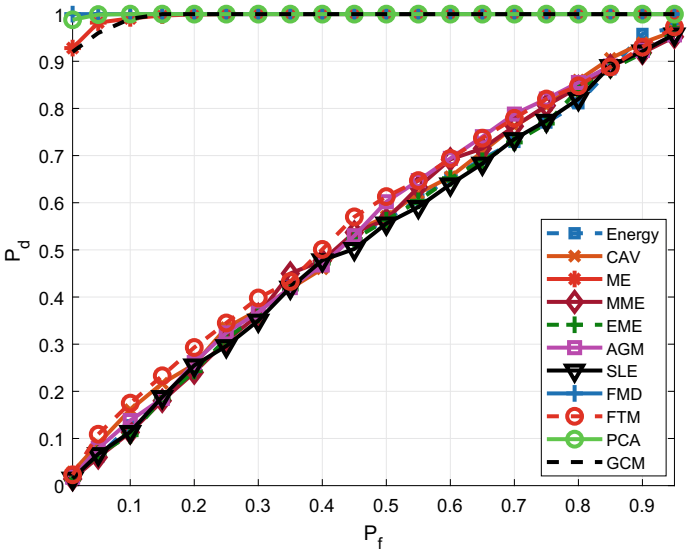



**Fig. 11.2** $P_d$ versus $P_f$ of BSS for WGN at −20 dB SNR

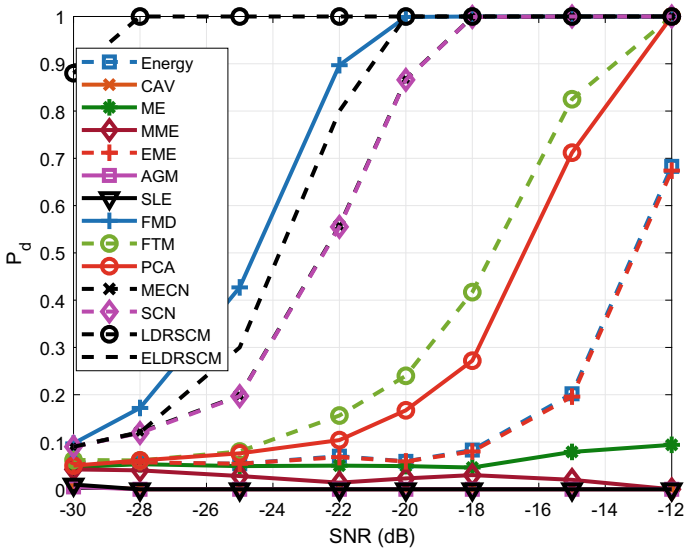



**Fig. 11.3** $P_d$ versus SNR for BSS techniques for CGN at $P_f = 0.1$

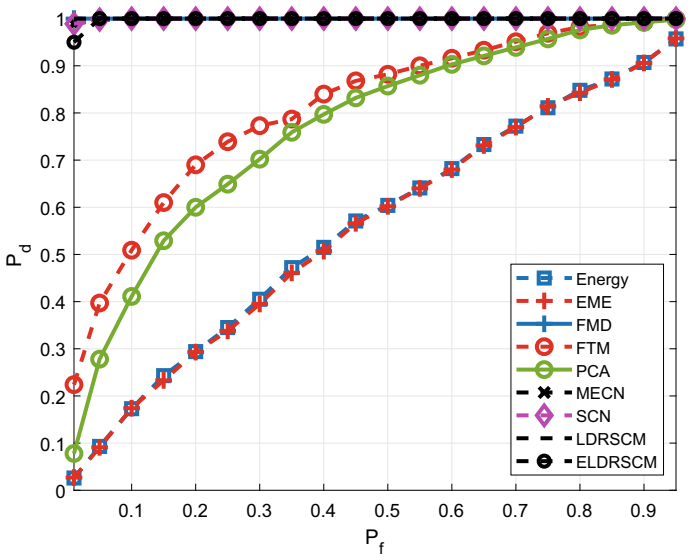



**Fig. 11.4** $P_d$ versus $P_f$ of blind techniques for colored Gaussian noise at −20 dB SNR

ME and MME for WGN. ELDRSCM has 1 dB performance degradation over than FMD. MECN and SCN have 2 dB deterioration as compare to the FMD, whereas PCA and FTM have 8 dB deterioration as compared to the FMD. It is also observed Fig. 11.3 that there is more SNR requirement for Energy and EME as compare to that of the FMD. Figure 11.4 revels the ROC of blind techniques for CGN at −20 dB of SNR. It is perceived from Fig. 11.4 that the signal detection performance of LDRSCM, FMD, ELDRSCM, MECN, and SCN are approximately same and have better performance over the other techniques as they are able to detect the signal at very low $P_f$ and low SNR. The PCA and FTM have superior ROC performance over the EME and EBD.

## 11.7 Experimental Results

In this section, we present the analysis on the results of experimental setup for some of the BSS algorithms which are discussed in previous section. The received noise samples are correlated due to the filter in universal software radio peripheral and correlation in noise samples (decaying exponentially) depends on the sampling frequency. Figures 11.5 and 11.6 revel the $P_d$ versus average received signal strength (RSS) at $P_f = 0.1$ for uncorrelated and correlated PU's signal, respectively. It is perceived from both the Figs. that FMD has better signal detection than the other blind SS methods. However, in the presence of correlated PU's signal the performance

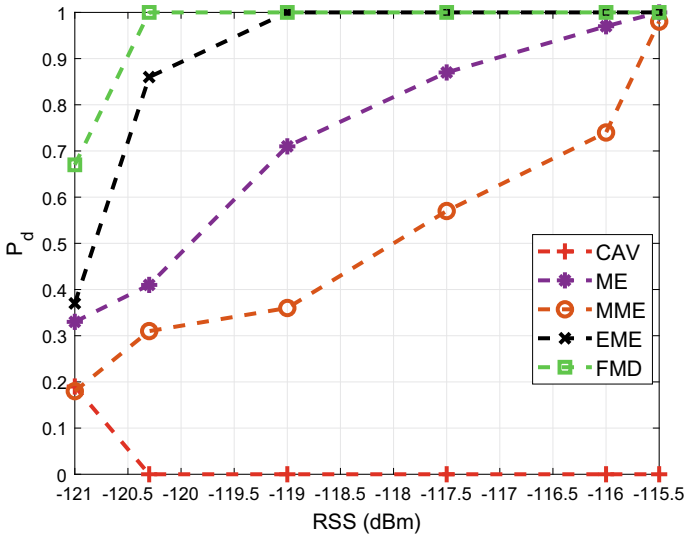



**Fig. 11.5** $P_d$ versus RSS at $P_f = 0.1$ for uncorrelated PU's signal

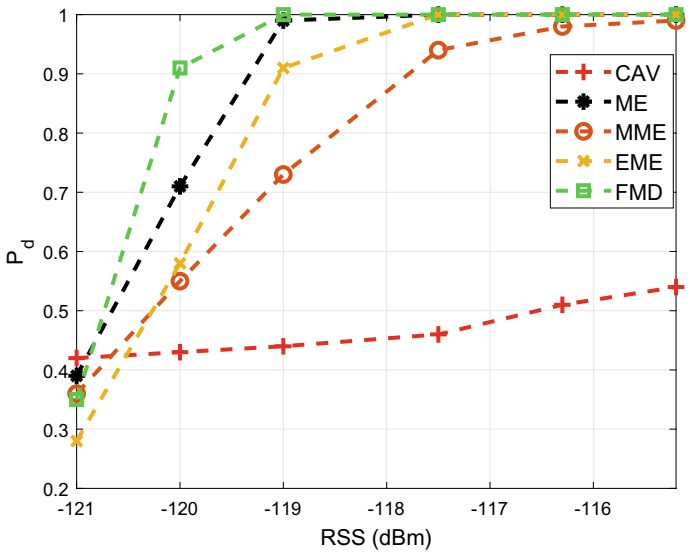



**Fig. 11.6** $P_d$ versus RSS at $P_f = 0.1$ for correlated PU's signal

of FMD degrades as compared to uncorrelated PU's signal. CAV detector is unable to detect the presence of PU's signal for uncorrelated signal, while it requires high RSS to achieve signal detection for correlated PU's signal as compared to other SS methods. Similar to FMD, the signal detection performance of EME deteriorates for correlated samples as compared to uncorrelated samples. However, the signal detection performance of MME and ME deteriorates for uncorrelated samples as compared to correlated samples.

## 11.8 Cognitive Radio with Full-Duplex

It is revealed from the above section that SU's base stations sense the presence of PU's signal during quiet period. It is a period in which there is no communication between SUs. In order to protect the PUs communication, BS has to sense the channel in every 2 s. This frequent quiet period leads to drastically reduces the spectral efficiency of SUs. To increase the spectral efficiency of SUs, CR with full-duplex is proposed in the literature. In CR with full-duplex, SUs base station employed full-duplex in which base station of SUs simultaneously communicate with CPEs and also detect the presence of PU's signal. However, the major problem is the self-interference created by the SUs base station to the PU's signal and hence the base station of SUs should employed some mechanism to cancel this self-interference for easily detecting the PU's signal.

## 11.9 Conclusion

In this chapter, various blind and non-blind spectrum sensing methods for both white and colored Gaussian noise is discussed. The cooperative spectrum sensing whose performance is much better than local spectrum sensing is also discussed. By analyzing the simulation results of different spectrum sensing techniques, the maximum to minimum eigenvalue technique is found to be suitable under white Gaussian noise, while for fine sensing in the IEEE 802.22 standard, energy with LogDet of received sample covariance matrix based sensing is found to be suitable under colored Gaussian noise. In addition, performance of some blind algorithms in real environment using software-defined radio is also analyzed. Further, method of increasing the spectral efficiency of secondary users by employing full-duplex is also presented.

## References

Bhatti FA, Rowe GB, Sowerby KW (2012) Spectrum sensing using principal component analysis. In: Proceedings of IEEE wireless communications and networking conferences. Paris, France, pp 725–730

Bishnu A, Bhatia V (2018) Grassmann manifold-based spectrum sensing for TV white spaces. IEEE Trans Cogn Commun Netw 4(3):462–472

Bishnu A, Bhatia V (2018) LogDet covariance based spectrum sensing under colored noise. IEEE Trans Veh Technol 67(7):6716–6720

Bishnu A, Bhatia V (2015) On performance analysis of IEEE 802.22 (PHY) for COST-207 channel models. In: Proceedings of IEEE conference on standards for communications and networking. Tokyo, Japan, pp 229–234

Chen H-S, Gao W, Daut DG (2009) Spectrum sensing for OFDM systems employing pilot tones. IEEE Trans Wirel Commun 8(12):5862–5870

Chen Z, Luan T, Zhang X-D (2012) Sensing orthogonal frequency division multiplexing systems for cognitive radio with cyclic prefix and pilot tones. IET Commun 6(1):97–106

Font-Segura J, Wang X (2010) GLRT-based spectrum sensing for cognitive radio with prior information. IEEE Trans Commun 58(7):2137–2146

Ganesan G, Li Y (2005) Cooperative spectrum sensing in cognitive radio networks. In: Proceedings of IEEE symposium on new frontiers dynamic spectrum access networks. Baltimore, Maryland, USA, pp 137–143

Ghasemi A, Sousa ES (2005) Collaborative spectrum sensing for opportunistic access in fading environments. In: Proceedings of IEEE symposium on new frontiers dynamic spectrum access networks. Baltimore, Maryland, USA, pp 131–136

Guo C, Zhang T, Zeng Z, Feng C (2006) Investigation on spectrum sharing technology based on cognitive radio. In: Proceedings of IEEE international conference on communications and networking. Beijing, China, pp 1–5

Han N, Shon S, Chung JH, Kim JM (2006) Spectral correlation based signal detection method for spectrum sensing in IEEE 802.22 WRAN systems. In: Proceedings of IEEE international conference on advanced communication technology, vol 3. Gangwon-Do, South Korea

Haykin S, Setoodeh P (2015) Cognitive radio networks: the spectrum supply chain paradigm. IEEE Trans Cogn Commun Netw 1(1):3–28

Horev I, Yger F, Sugiyama M (2017) Geometry-aware principal component analysis for symmetric positive definite matrices. Spring Mach Learn 106(4):93–522

Hou S, Qiu RC (2014) Kernel feature template matching for spectrum sensing. IEEE Trans Veh Technol 63(5):2258–2271

Jolliffe IT (1986) Principal component analysis. Springer, New York

Kim K, Akbar IA, Bae KK, Um J-S, Spooner C-M, Reed JH (2007) Cyclostationary approaches to signal detection and classification in cognitive radio. In: Proceedings of IEEE symposium on new frontiers dynamic spectrum access networks. Dublin, Ireland, pp 212–215

Kim H, Shin KG (2010) In-band spectrum sensing in IEEE 802.22 WRANs for incumbent protection. IEEE Trans Mob Comput 9(12):1766–1779

Lei Z, Chin FPS (2010) Sensing OFDM systems under frequency-selective fading channels. IEEE Trans Veh Technol 59(4):1960–1968

Lin F, Qiu RC, Hu Z, Hou S, Browning JP, Wicks MC (2012) Generalized FMD detection for spectrum sensing under low signal-to-noise ratio. IEEE Commun Lett 16(5):604–607

Liu X, Sai Shankar N (2006) Sensing-based opportunistic channel access. Mob Netw Appl 11(4):577–591

Lunden J, Koivunen V, Huttunen A, Poor H V (2007) Spectrum sensing in cognitive radios based on multiple cyclic frequencies. In: Proceedings of IEEE international conference on cognitive radio oriented wireless network and communications. Orlando, USA, pp 37–43

Mishra AK, Johnson DL (2014) White space communications: advances, developments and engineering challenges. Springer, New York

Peh E, Liang Y-C (2007) Optimization for cooperative sensing in cognitive radio networks. In: Proceedings of IEEE wireless communications and networking conference. Paris, France, pp 27–32

Quan Z, Shellhammer ST, Zhang W, Sayed AH (2009) Spectrum sensing by cognitive radios at very low SNR. In: Proceedings of IEEE global communications. Honolulu, Hawaii, USA, pp 1–6

Sharma SK, Chatzinotas S, Ottersten B (2014) Maximum Eigenvalue detection for spectrum sensing under correlated noise. In: Proceedings of IEEE international conference on acoustic, speech, and signal processing. Florence, Italy, pp 7268–7272

Sharma SK, Chatzinotas S, Ottersten B (2013) Eigenvalue-based sensing and SNR estimation for cognitive radio in presence of noise correlation. IEEE Trans Veh Technol 62(8):3671–3683

Shellhammer S, Tawil V, Chouinard G, Muterspaugh M, Ghosh M (2006) Spectrum sensing simulation model, IEEE 802.22-06/0028r10, 2006

Sobron I, Diniz PSR, Martins WA, Velez M (2015) Energy detection technique for adaptive spectrum sensing. IEEE Trans Commun 63(3):617–627

Stevenson CR, Cordeiro C, Sofer E, Chouinard G (2006) Functional requirements for the 802.22 WRAN Standard, IEEE 802.22-05/0007r47, Jan 2006

Sun C, Zhang W, Letaief KB (2007) Cluster-based cooperative spectrum sensing in cognitive radio systems. In: Proceedings of IEEE international conference on communications. Glasgow, Scotland, pp 2511–2515

Tawil V (2006) 51 captured DTV signal, 2006. (Online). Available: http://grouper.ieee.org/groups/802/22/Meetingsdocuments/2006May

Visotsky E, Kuffner S, Peterson R (2005) On collaborative detection of TV transmissions in support of dynamic spectrum sharing. In: Proceedings of IEEE symposium on new frontiers dynamic spectrum access networks. Baltimore, Maryland, USA, pp 338–345

Vladeanu C, Nastase C-V, Martian A (2016) Energy detection algorithm for spectrum sensing using three consecutive sensing events. IEEE Wirel Commun Lett 5(3):284–287

Wang B, Liu KJR (2011) Advances in cognitive radio networks: a survey. IEEE J Select Top Sign Process 5(1):5–23

Wei L, Tirkkonen O (2011) Analysis of scaled largest Eigenvalue based detection for spectrum sensing. In: Proceedings of IEEE international conference on communications. Kyoto, Japan, pp 1–5

Yucek T, Arslan H (2009) A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun Surv Tutor 11(1):116–130

Zeng Y, Liang Y-C (2009) Eigenvalue-based spectrum sensing algorithms for cognitive radio. IEEE Trans Commun 57(6):1784–1793

Zeng Y, Liang Y-C (2009) Spectrum-sensing algorithms for cognitive radio based on statistical covariances. IEEE Trans Veh Technol 58(4):1804–1815

Zeng Y, Koh CL, Liang YC (2008) Maximum eigenvalue detection: theory and application. In: Proceedings of IEEE international conference on communications. Beijing, China, pp 4160–4164

Zhang P, Qiu R (2013) GLRT-based spectrum sensing with blindly learned feature under rank-1 assumption. IEEE Trans Commun 61(1):87–96

# Chapter 12
# Cooperative Spectrum Sensing in Energy Harvesting Cognitive Radio Networks Under Diverse Distribution Models


**Banani Talukdar, Deepak Kumar, Shanidul Hoque, and Wasim Arif**

**Abstract** To address spectrum underutilization and energy constraint in wireless communication technologies, cognitive radio networks (CRN) incorporated with energy-harvesting (EH) ability is an evergreen solution. The dynamic behavior of the primary user (PU) activity is a primary component that affects the performance of a CRN. In this chapter, we investigate the impact of distribution functions on the performance of an energy harvesting enabled cooperative CRN. We consider the two most relevant distribution functions, namely Weibull and Erlang distributions, to characterize the PU behavior in a prediction-based sensing under a cooperative EH-CRN. Cooperative CRN offers better reliability of event detection, which results in efficient spectrum utilization. In this chapter, we consider a centralized cooperative EH-CRN whereby each cognitive radio (CR) node has the capability of scavenging energy from radio frequency (RF) or non-RF sources depending on a combined decision taken by the fusion center (FC). We use conventional and estimation-based energy detection schemes in our analysis. Analytical formulae for the detection probability, harvested energy, normalized throughput, and energy penalty are established, employing OR fusion rule. The impact of prediction error, number of cooperative CR nodes, number of frames, and collision constraint on energy harvesting and normalized throughput is also studied. Simulations are performed, and a thorough,

B. Talukdar · W. Arif (✉)
Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India
e-mail: arif.ece.nits@gmail.com

B. Talukdar
e-mail: talukdarbanani.smit@gmail.com

D. Kumar
Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, India
e-mail: dpk.arrah@gmail.com

S. Hoque
Department of Electronics and Communication Engineering, Madanapalle Institute of Technology & Science, Madanapalle, India
e-mail: shanidulhoque257@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_12

comprehensive comparison of the results is presented. A detailed comparative analysis for both Weibull and Erlang distributions is also presented. The results show that both distributions perform better than the conventional exponential distribution in a centralized cooperative EH-CRN and signify the usability of the model in designing practical systems.

**Keywords** Cognitive radio · Cooperative network · Energy harvesting · OR fusion rule · Spectrum sensing · Weibull distribution

## 12.1 Introduction

Radio electromagnetic spectrum is perhaps the most precious radio resource and is finite in nature. However, numerous studies have revealed that a greater portion of these licensed radio bands remains unutilized, i.e., the spectrum occupancy is less than <35% to be precise. This leads to an inept use of the radio spectrum (Federal Communications Commission (FCC) 2003; Shared Spectrum Company). The Federal Communications Commission (FCC) authorizes the secondary users to access the licensed spectrum ingeniously (Mitola and Maguire 1999). Cognitive radio is a proficient technology to address the spectrum crunch in wireless networks given the rapid rise in wireless technology and applications. CR is a smart system in the wireless communication domain that is receptive toward its surroundings, uses cognition to learn from its surroundings, and adapts its parametric variations accordingly (Haykin 2005; Jondral 2005; Tandra and Sahai 2008; Akyildiz et al. 2011). The two main purposes of CR are highly authentic and smooth communication and effective usage of the radio resource. CR technology provides an efficient spectrum sharing policy amongst the licensed users (i.e., PU) and the unlicensed users (i.e., secondary user or CR node) when the spectrum remains unused by the PU (Federal Communications Commission (FCC) 2003). Spectrum sensing is a crucial step in a CRN as it encounters the presence of PUs correctly and promptly when the secondary users are unaware of the PU signals (Akyildiz et al. 2006, 2009; Ma et al. 2009; Yucek and Arslan 2009). There exist various spectrum-sensing approaches, viz energy detection (Digham et al. 2007), matched filter-based detection (Yucek and Arslan 2009) and cyclostationary spectrum detection (Gardner 1988, 1994), multi-taper method, wavelet detection, etc. (Ma et al. 2009; Yucek and Arslan 2009). On the other side, energy harvesting (EH) finds enormous application to meet the upcoming explosion of wirelessly connected devices and associated network technologies. An energy harvesting enabled cognitive radio network (EH-CRN) addresses both the matters of spectrum underutilization and energy-efficient communication. This chapter investigates the performance of estimation-based detection under a centralized cooperative EH-CRN under the constraint of primary user activity model. We introduce Weibull and Erlang distribution function to characterize the PU activity model and develop the analytical expressions for various performance measuring metrics of the system model.

### *12.1.1 Literature Review*

The ultimate motive of the spectrum-sensing approaches is to efficiently and swiftly identify the available spectrum bands so that a secondary user can utilize the unused spectrum band ingeniously without compromising the quality of service (QoS) of the licensed user. The effect of the sensing period on the performance of a CRN is investigated in Liang et al. (2008). In Liang et al. (2008), the authors studied that there subsists a trade-off amongst the sensing duration as well as the throughput. Literature (Xing et al. 2012) studied the outcome of the temporal behavior of PU activity over sensing. Prediction-based sensing improves the quality of service (QoS) of PU and spectral use altogether (Mariani et al. 2015). In Mariani et al. (2015), authors investigated different types of prediction mechanisms (e.g., prediction based on a hidden Markov model, prediction done employing a multilayer perceptron neural network, etc.). Prediction and sensing mechanisms in a CRN contribute significantly to identifying the present status of the PU in a radio spectrum. In Bhowmick et al. (2017), the authors investigated the performance of a prediction-based sensing cooperative CRN.

Awareness over energy conservation in the case of 5G wireless communication networks is on the rise because of the excessive energy expenditure. 5G networks are envisaged for its large-scale deployment in 2020 enabling extensive connectivity and high data rate speeds. The remarkable lowering in communication delay, as well as the support for practical multimedia applications, will rise. However, fulfilling the strict energy usage as well as computation costs in a reasonable and viable way is challenging. ICT alone leads to 5% of the total $CO_2$ emanation as recorded in the year 2015. This is expected to rise, as the number of mobile devices will soar in the coming years. Extensive forecasts reveal that the total global electricity demand ICT will escalate in 2020 and the data centers will consume the bigger slice, i.e., 20.9% of the projected electricity demand. That means communication networks will eat up around 300–9000 TWh (Barani 2020). To tackle these issues, it is anticipated that 5G networks will disperse low power communication nodes. Energy harvesting is a novel approach of powering wireless devices by converting the energy present in the environment into useful electrical energy. Energy harvesting is a very promising technology and possibly a greener solution to environmental degradation (Shafie et al. 2015; Shafie and Khattab 2014). In Shafie et al. (2015) and Shafie and Khattab (2014), the authors studied that a CR node harnesses the energy, either from RF sources (e.g., downlink RF signal from the base station) or from ambient sources (e.g., temperature, solar, wind, etc.). Energy harvesting will lead to truly self-sustaining and perpetual nodes. Spectral as well as energy efficiency improved by energy-harvesting CRN is investigated in Park and Hong (2013). The distinct features of RF energy-harvesting wireless networks such as low power and long distance wireless transfer clearly indicate that it is a suitable method concerning energy efficiency in 5G technology. It is anticipated that beyond 5G technologies will incorporate the modern wireless technologies into an IP-based network that will provide high performing network services universally (Akhtar

et al. 2018). Lately, studies have been conducted over the dense distribution of small cells upon a greater section of low radio frequency signals as this is a prospective solution to spectrum sharing and energy efficiency in 5G systems. The deployment of such a diverse environment accompanied by the concomitance of different classes of cells and user equipment builds a heterogeneous network (HetNet) architecture that have cracked huge prospects to meet the 5G standards (Bhandari and Joshi 2018). A HetNet deployment assists in spectral efficiency by means of greater frequency reuse and a lower transmit power. A certain other way of integrating CR principles with 5G standards is by allowing the coexistence of two or several HetNets spectrally in various dimensions like time, spatial, frequency, and polarization by adopting various approaches like carrier aggregation, cognitive beamforming, cognitive interference alignment, etc. (Sharma and Rawat 2016).

A unique spectrum-sensing policy is proposed, and the average throughput of an EH-CRN over energy causality with collision constraints has been investigated in Park and Hong (2014). In Shafie (2014), the author studied a space-time coding scheme to obtain maximum throughput in an EH cooperative CRN scenario. In Mao et al. (2012), the author investigated an optimal energy controlling strategy that maximizes the throughput of an EH sensor node. A novel approach of coexistence is projected in Lee et al. (2013) where a SU can scavenge RF signals from the neighboring PU transmitters in an opportunistic manner or transmit when the PU transmitters are distant. In Zhang and Ho (2013), the authors mentioned that simultaneous sensing, as well as energy harvesting, is possible employing a power-splitting device. In Ahmed et al. (2017), with the purpose of maximizing the amount of data transmission of the SUs, an optimal detection threshold centered over the harvested energy is obtained for protecting the PUs from a collision. To magnify the energy efficacy of the CR networks, Yuan et al. (2018) put forward a spectrum sensing as well as data transmission paradigm facilitated by energy harvesting. In Lee et al. (2018), an optimal spectrum-sensing scheme for opportunistic spectrum utilization is suggested. Besides, an energy-harvesting scheme taking into account the PU collision as well as the energy causality constraint is proposed. In Kim (2019), a novel method for concurrent spectrum-sensing technique and energy-harvesting paradigm is recommended. Here, the entire received RF power is employed for energy harvesting. Subsequently, depending on a portion of the harvested power, spectrum sensing is carried out. Kumar et al. (2019a) analyze the performance of the hard fusion rules, in particular, OR, AND, and majority rules in an EH-CRN in terms of throughput and energy harvesting.

The PU activity contributes significantly to analyzing the performance of an EH-CRN. Tele-traffic has always been characterized using an exponential distribution. The user communication pattern has changed from voice-driven to content-driven. Therefore, the exponential model used to define the PU activity is found to be inappropriate in such scenarios. Also, the assumption of exponential distribution in case of an ideal and busy period of a PU is not valid for all frequency bands, instead, distribution with heavier tails is more appropriate for such characterization. The throughput performance also drops significantly when considering an exponential distribution. This is the primary motivation behind carrying out this work to analyze

the performance of some other widely used distribution functions in the domain of signal processing and communication in EH-enabled CRN. Weibull and Erlang distributions are extensively used in reliability engineering, as well as other applications such as survival analysis, information retrieval, model fading channels in wireless communications because of its simplicity and adaptability (Hoque et al. 2018). Also, these two functions are more generic with respect to exponential function and therefore are very useful in analyzing the performance of any system. The flexibility of Weibull and Erlang distribution is a prime reason for incorporating them in analyzing the performance of EH-CRN. The detection probability, false alarm probability, spectrum reuse, harvested energy, normalized throughput, and energy penalty of a central cooperative EH-CRN by employing OR fusion rule are developed. In Arif et al. (2015), the authors studied the impact of SU mobility by considering intra-cell spectrum handoff mechanism, modeling, and characterizing the PU activity using distinct distribution functions in a CR network is shown. Hoque and Arif (2017) present an elaborate mathematical analysis of spectrum handoff considering general distributions characterizing the length of the SU's call. For analysis of conventional data and telecommunication network, the call holding and channel holding intervals are usually considered to obey exponential distribution for accommodating the analytical results. It is stated that modeling the call and channel holding time using exponential distribution by no means are appropriate for the investigation of any tele-traffic network. In Kumar et al. (2019b), the authors studied the impact of the Weibull distributed PU activity model on an EH-CRN. The performance of such a network for ideal energy detector and estimated noise power-based energy detection spectrum-sensing strategies is studied. In this chapter, we consider both energy detection and estimated noise power-based energy detection methods and analyze the system performance.

### *12.1.2 Organization of the Chapter*

In Sect. 12.2, we present the system model, activity model of PU, global prediction, global sensing, and combined decision, and we obtain the analytical equations for determining the probability of detection, probability of false alarm, and improvement in spectrum reuse. We formulate the analytical expressions for energy harvesting, normalized throughput, and energy penalty in Sect. 12.3. Section 12.4 exhibits the mathematical results and discussions and lastly, and in Sect. 12.5, we put forward the conclusion to our work.

## 12.2 System Model

In this chapter, we discuss about a centralized cooperative CRN with $N$ number of CR nodes connected to a fusion center (FC), and each node is capable to harvest

energy. We use prediction-based sensing for local as well as global decisions in such a framework. A CR node predicts the existence of a PU and sends the prediction outcome in the form of hard binary [0 1] to the FC where '0' represents the absence and '1' represents the presence of PU at any instant. We use OR fusion rule at FC to combine the cooperative data to acquire the global prediction toward the existence of a PU. The local sensing data (hard decision: {0 or 1}) from CR is also fused at FC for global sensing decisions based on OR fusion rule. There exists various rules for issuing a final decision concerning the existence of PU. Cooperation amongst the existing nodes is incorporated to strengthen the reliability of event detection in CRN. The method of fusing the local decisions at the FC is termed as decision fusion that could be a hard decision fusion or soft decision fusion. In a hard decision fusion, the local decisions taken by the nodes are combined to reach to a global decision, whereas in soft decision, a weighted value of test statistics of individual nodes is combined to obtain a global decision. Here, we use OR fusion rule which is a simple decision-making rule and is categorized as a hard decision rule used in centralized or decentralized cooperative CRN. The OR rule is formulated as $H_1 : \Lambda = \sum_{i=1}^{N} I_i \geq 1$; $H_0$ : Otherwise. According to this rule, if any of the outcomes of the sensing decisions stands in favor of existence of PU, then the overall decision stands in support of the existence of a PU. For making global decision, binary hypothesis is used where hypothesis $H_0$ states the inexistence of PU and $H_1$ states that the PU exists based on the decision rule with respect to the test statistics of the received signal. In the presented model, both the global prediction and global sensing decisions are used to frame a combined decision, which determines the overall transmission and energy-harvesting schedule of CR nodes in the network. Here, FC creates a four-level quantized decision (00 or 01 or 10 or 11) depending on the mutual decision of global prediction as well as global sensing where the start bit signifies global prediction outcome and the next bit signifies global sensing outcome. In the proposed model, we consider that a CR node scavenges energy over the prediction and the sensing durations from both the RF sources and the non-RF sources. The energy harvesting as well as the transmission of the CR nodes all through the transmission interval is centered on the combined decision of FC as shown in Table 12.1. The CR node begins transmitting after the prediction time and the sensing time if and only if the combined decision declares that PU is absent (i.e., '00') and harvests energy from RF sources if and only if combined decision declares that PU

**Table 12.1** CR energy harvesting and transmission decisions (Kumar et al. 2019b)

| Collective prediction decision | Collective sensing decision | Combined global decision | CR harvesting source | CR transmission |
|---|---|---|---|---|
| $H_0$ | $H_0$ | 00 | Non-RF source | Yes |
| $H_0$ | $H_1$ | 01 | Non-RF source | No |
| $H_1$ | $H_0$ | 10 | Non-RF source | No |
| $H_1$ | $H_1$ | 11 | RF source | No |

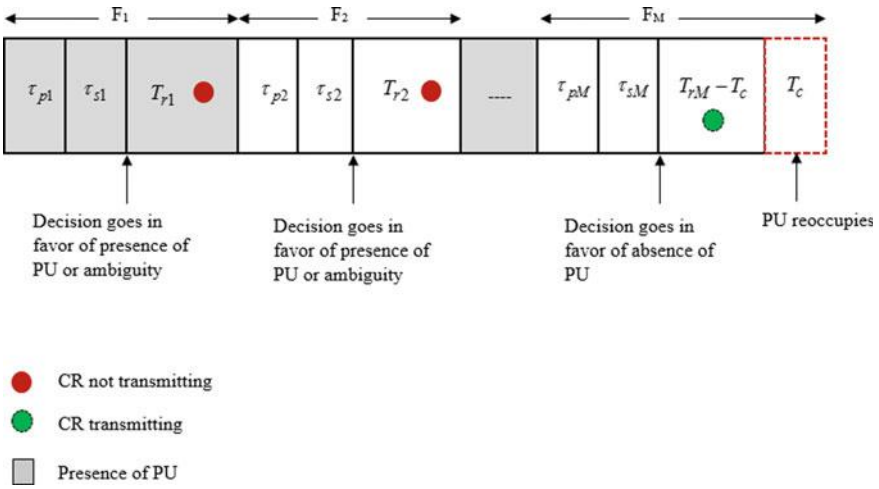



**Fig. 12.1** Frame organization of a multiple detection cycle

is present (i.e., '11'). We consider that the energy needed to run the harvester per unit time is less than the harvested energy out of RF and non-RF sources per unit time. We also assume that a CR switch ON each of the two harvesters (i.e., RF as well as non-RF) during the prediction and the sensing time given the total time for both predictions as well as sensing is very short, in comparison with the transmission interval of the frame. Alternatively, a CR turns ON just one harvester during the transmission time of the frame because of energy constraints. The prediction-sensing circuit receives $1 - \eta$ part, and the energy-harvesting circuit obtains $\eta$ a fraction of the reserved energy instantaneously using an energy-splitting device.

The projected system model is portrayed in Fig. 12.1. The detection frame having a length $T$ is a combination of prediction interval $\tau_{\mathrm{p}}$, sensing interval $\tau_{\mathrm{s}}$, and transmission interval $T_{\mathrm{r}}$ for a single frame. In case considering multiple frames, $\tau_{\mathrm{p}} = \tau_{\mathrm{p},i}$, $\tau_{\mathrm{s}} = \tau_{\mathrm{s},i}$ and $T_{\mathrm{r}} = T_{\mathrm{r},i}$ where $i = 1, 2, 3, \ldots M, M$ represents the number of detection frames. Thereafter, we consider that the PU behavior is random in nature and can take up the channel again during transmission. In this scenario, the transmission time comprises useful transmission interval $T_{\mathrm{r,use}}$ and dwelling instant (or the PU reoccupation interval), i.e., $T_{\mathrm{r}} = T_{\mathrm{r,use}} + T_{\mathrm{c}}$. The analytical framework for a single frame and multiple frames for the projected system model is established in the subsequent subsections.

### *12.2.1 Activity Model of PU*

To define the states of a PU, namely *idle* and *busy*, a Markov model (Shi et al. 2013) having two states is employed under the hypotheses $H_0$ and $H_1$ where hypothesis

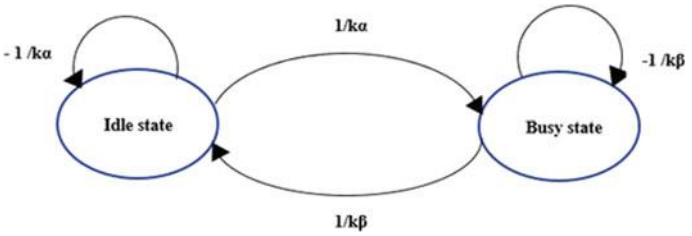



**Fig. 12.2** Primary user activity model

$H_0$ states the inexistence of PU and $H_1$ states the existence of PU, respectively. In this chapter, we introduce Weibull distribution and Erlang distribution functions to describe the primary user behavior for evaluating the overall performance of the different performance measuring parameters for a centralized EH-enabled cooperative CRN. In Fig. 12.2, a two-state Markov model is shown having two states where the PU activity model follows an Erlang distribution. The PU changes its state from idle to busy and back with rate $1/k\alpha$ and $1/k\beta$ correspondingly. The PU remains in idle state and busy state with $-1/k\alpha$ and $-1/k\beta$, respectively.

#### 12.2.1.1 Weibull Distribution

Weibull distribution is widely employed in reliability engineering additionally with other applications because of its effortlessness and adaptability (Hoque et al. 2018; Hoque and Arif 2018). Weibull distribution has a broad application area such as in communication engineering. To model a fading channel in wireless communication, the Weibull distribution model appears to display a good fit to experimental fading channel measurements. This distribution is influenced by its shape and scale parameters. The flexibility of this distribution motivates us to employ this in modeling the PU activity. The *idle* state and the *busy* state have a probability density function, which is expressed as Eqs. (12.1–12.2) (Kumar et al. 2019b):

$$P_{\mathrm{I}}(t) = m\alpha^{-1}\left(t\alpha^{-1}\right)^{m-1}\exp\left(-t\alpha^{-1}\right)^{m} \tag{12.1}$$

$$P_{\mathrm{B}}(t) = m\beta^{-1}\left(t\beta^{-1}\right)^{m-1}\exp\left(-t\beta^{-1}\right)^{m} \tag{12.2}$$

where $\alpha, \beta, m$ and $P(.)$ indicate the rate parameter of the idle state, rate parameter of busy state, shape parameter of the Weibull distribution, and probability, respectively. The mean value of the idle and the busy states is $\alpha\Gamma(1+1/m)$ and $\beta\Gamma(1+1/m)$ where $\Gamma(\cdot)$ characterizes the Gamma function and is specified by $\Gamma(1+m) = \int_0^{\infty} x^m \mathrm{e}^{-x}\mathrm{d}x$. The stationary probabilities of the *idle* state and the *busy* state are given as $P(H_0) = \alpha/(\alpha+\beta)$ and $P(H_1) = \beta/(\alpha+\beta)$. The Weibull distribution is associated to several other probability distribution functions. For instance,

Weibull distribution interpolates between exponential distribution when $m = 1$ and Rayleigh distribution when $m = 2$ and $\alpha = \sqrt{2}\sigma^2$.

#### 12.2.1.2 Erlang Distribution

Being a continuous probability distribution, the Erlang distribution has a wide applicability because of its affinity to the exponential distribution (Arif et al. 2015; Hoque and Arif 2017). The *idle* and the *busy* states have the probability density function given as Eqs. (12.3–12.4):

$$P_{\mathrm{I}}(t) = a^k t^{k-1} \exp(-ta) \Big/ (k-1)! \tag{12.3}$$

$$P_{\mathrm{B}}(t) = b^k t^{k-1} \exp(-tb) \Big/ (k-1)! \tag{12.4}$$

where $a$, $b$ and $k$ indicate the rate parameter of the idle state, rate parameter of busy state, and shape parameter of the Erlang distribution, respectively. The mean value of the idle and the busy states is $k/a$ and $k/b$ accordingly. The *idle* state has $P(H_0) = a \Big/ (a+b)$ as its stationary probability, and the *busy* state has $P(H_1) = b \Big/ (a+b)$ as its stationary probability, respectively. In the established system model, we feel that a PU can reoccupy the channel at any instant of time, and therefore, there are four possible hypotheses ($H_{0,0}$, $H_{0,1}$, $H_{1,0}$, and $H_{1,1}$) about PU occupancy status in the channel as follows:

$H_{0,0}$: PU stays idle for the complete transmission duration, and the equivalent probability is called $P_{idle}$.
$H_{0,1}$: Primarily, the PU is in idle state, then after a certain period, the PU reoccupies the channel, and the equivalent probability is called $P_{re}$.
$H_{1,0}$: At first, the PU occupies the channel for a definite interval of time and then abandons the channel. The equivalent probability is called $P_{dis}$.
$H_{1,1}$: PU stays busy for the complete transmission duration, and the respective probability is called $P_{busy}$.

Correspondingly, the probabilities of all these four hypotheses can be written as Eqs. (12.5–12.8) (Bhowmick et al. 2017; Kumar et al. 2019b):

$$P\left(H_{0,0}\right) = P_{\mathrm{idle}} = \int_0^{\tau} P_{\mathrm{I}}(t)\mathrm{d}t \tag{12.5}$$

$$P\left(H_{0,1}\right) = P_{\mathrm{re}} = \int_0^{T_{\mathrm{r}}} P_{\mathrm{I}}(t)\mathrm{d}t \tag{12.6}$$

$$P\left(H_{1,0}\right) = P_{\text{dis}} = \int_0^{T_r} P_{\text{B}}(t)\mathrm{d}t \tag{12.7}$$

$$P\left(H_{1,1}\right) = P_{\text{busy}} = \int_0^{\tau} P_{\text{B}}(t)\mathrm{d}t \tag{12.8}$$

## 12.3 Mathematical Modeling and Analysis of the Proposed System

For ensuring the quality of service in both the networks, both the probability of detection and the false alarm have to be improved, and the energy-harvesting ability should be maximally utilized. We use prediction-based sensing which offers an improved probability of detection and enhances the overall harvested energy in the network.

### *12.3.1 Global Prediction*

All CR nodes predict the present status of the PU from the past data of the PU activity. Many methods are there for prediction. They are the hidden Markov model (HMM), neural networks, and so forth (Xing et al. 2012; Mariani et al. 2015; Borgne et al. 2007). However, in this work, we have no interest in the prediction approaches rather we aim to study the implications of prediction on the performance metrics about the detection and the false alarm probability, system throughput, and harvested energy. We use an HMM model for prediction of the PU status. Let the estimated final prediction regarding the existence of PU on the channel be represented by $\hat{X}$. It is a function of the distinct prediction outcomes ($\hat{x}$) of all the CRs present in the network. Here, $i = 1, 2, 3, \ldots, N$, $N$ being the various CR nodes operating in cooperation. Therefore,

$$\hat{X} = f\left(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\right) \tag{12.9}$$

Here, $f(\cdot)$ gives the fusion function incorporated in the FC and represents the prediction outcome of the individual CR node. To merge, the calculated predictions cooperatively received from the CR nodes. The OR fusion scheme is used for its superiority in terms of detection probability and throughout when channel utilization is a major concern. We assume the different prediction probabilities like $\hat{P}_{d=1,i}$, $\hat{P}_{d=0,i}$, $\hat{P}_{f,i}$ and $\hat{P}_{m,i}$ to depict the results of prediction. They are given as Eqs. (12.10–12.13) (Bhowmick et al. 2017; Kumar et al. 2019b):

$$\hat{P}_{d=1,i} = P(H_1)(1 - P_e) \tag{12.10}$$

$$\hat{P}_{d=0,i} = P(H_0)(1 - P_e) \tag{12.11}$$

$$\hat{P}_{f,i} = P(H_0)(P_e) \tag{12.12}$$

$$\hat{P}_{m,i} = P(H_1)(P_e) \tag{12.13}$$

where $P_e$ signifies the prediction error. We take into consideration that for all the above circumstances, the prediction error remains the same. Therefore, the global prediction probability using OR fusion rule at the FC can be obtained as Eqs. (12.14–12.17) (Kumar et al. 2019b):

$$P\left(\hat{X} = H_1 | H_1\right) = 1 - \prod_{i=1}^{N} (1 - \hat{P}_{d=1,i}) \tag{12.14}$$

$$P\left(\hat{X} = H_0 | H_0\right) = \prod_{i=1}^{N} \hat{P}_{d=0,i} \tag{12.15}$$

$$P\left(\hat{X} = H_1 | H_0\right) = 1 - \prod_{i=1}^{N} (1 - \hat{P}_{f,i}) \tag{12.16}$$

$$P\left(\hat{X} = H_0 | H_1\right) = \prod_{i=1}^{N} \hat{P}_{m,i} \tag{12.17}$$

where $P(\cdot|\cdot)$ represents the global prediction decision.

### 12.3.2 Global Sensing

We consider that every node of the CRN begins the detection process with a residual $E_R$ kept in its battery. The signal obtained at the CR on sensing the channel is given as Eq. (12.18) (Bhowmick et al. 2017)

$$Y_{CR_i}(n) = \begin{cases} w(n), & H_0 : \text{PU absent} \\ hs(n) + w(n), & H_1 : \text{PU present} \end{cases} \tag{12.18}$$

Here, additive white Gaussian noise (AWGN) is characterized by $w(n)$, PU signal is represented by $s(n)$, $h$ gives the channel fading coefficient of the dedicated channel between the PU and the CR node. AWGN is independent and identically distributed with a zero mean and variance of $\sigma_w^2$.

#### 12.3.2.1 Conventional (Ideal) Energy Detector

We assume that $f_s$ is the signal sampling frequency at the CR. $N_s$ gives the number of samples such that $N_s = \tau_s f_s$. In an ideal ED, the test statistics $T_{CR_i}$ can be written as Eq. (12.19)

$$T_{CR_i} = \frac{1}{N_s} \sum_{n=1}^{N_s} \left| Y_{CR_i}(n) \right|^2 \tag{12.19}$$

For an increased set of samples, the test statistics $T_{CR_i}$ could be a Gaussian distribution (mean, $\mu_0 = \sigma_w^2$ and variance, $\sigma_0^2 = \frac{1}{N_s}\sigma_w^4$) for $H_0$ and (mean, $\mu_1 = (1+\gamma_r)\sigma_w^2$ and variance, $\sigma_1^2 = \frac{1}{N_s}(1+\gamma_r)\sigma_w^4$) for $H_1$ (Liang et al. 2008; Singh et al. 2016). Here, $\gamma_r$ can be expressed as $\gamma_r = |h|^2\gamma_s$; $\gamma_r$ being the SNR of the fading channel. Similarly, $\gamma_s$ (SNR of the Gaussian sensing channel) may be given as $\gamma_s = \sigma_s^2/\sigma_w^2$; where $\sigma_s^2$ signifies signal variance (Singh et al. 2016; Zhang et al. 2014). All the CRs in the network match their test statistics with a specific threshold value, '$\lambda_{CR}$.' Hence, the distinct sensing probabilities for the $i$th CR considering the ideal energy detector (ED) can be written as Eqs. (12.20–12.22) (Bhowmick et al. 2017; Kumar et al. 2019b):

$$P_{d,ced}(\tau_s) = Q\left(\left(\frac{\lambda_{CR}}{\sigma_w^2} - \gamma_r - 1\right)\sqrt{\frac{\tau_s f_s}{2\gamma_r + 1}}\right) \tag{12.20}$$

$$P_{f,ced}(\tau_s) = Q\left(\left(\frac{\lambda_{CR}}{\sigma_w^2} - 1\right)\sqrt{\tau_s f_s}\right) \tag{12.21}$$

$$P_{m,ced}(\tau_s) = 1 - P_{d,ced}(\tau_s) \tag{12.22}$$

The $Q$-function is given as $Q(x) = \left(1/\sqrt{2\pi}\right)\int_x^\infty e^{-t^2/2}dt$.

#### 12.3.2.2 Estimated Noise Power-Based Energy Detector (ENP-Based ED)

The CR nodes are unaware of the noise that exists in the network; and because of this insufficiency of information about the noise power, it might restrict the sensing performance of an ideal ED. In most cases, this uncertainty may also lead to signal-to-noise ratio (SNR) penalties. Sensing performance could be improved by knowing the noise power. To boost the performance of the sensing mechanism, it is necessary to determine the noise (Mariani et al. 2011). Literature (Mariani et al. 2011) investigates the efficiency of maximum likelihood-based noise power determination over ED in the case of a cooperative CR network. The sensing probabilities related to the

estimated noise power energy detector can be given as Eqs. (12.23–12.25) (Kumar et al. 2019b):

$$P_{\mathrm{d,enp}}(\tau_{\mathrm{s}}) = Q\left(\left(\frac{\lambda_{\mathrm{CR}}}{\sigma_w^2} - \gamma_{\mathrm{r}} - 1\right)\sqrt{\frac{Z}{2\gamma_{\mathrm{r}}+1}}\right) \tag{12.23}$$

$$P_{\mathrm{f,enp}}(\tau_{\mathrm{s}}) = Q\left(\left(\frac{\lambda_{\mathrm{CR}}}{\sigma_w^2} - 1\right)\sqrt{Z}\right) \tag{12.24}$$

$$P_{\mathrm{m,enp}}(\tau_{\mathrm{s}}) = 1 - P_{\mathrm{d,enp}}(\tau_{\mathrm{s}}) \tag{12.25}$$

Now, $Z = \frac{N_{\mathrm{s}}N_{\mathrm{w}}}{N_{\mathrm{s}}+N_{\mathrm{w}}}$ where $N_{\mathrm{w}}$ is the noise samples employed in determining the noise power. The sensing probabilities can be written as Eqs. (12.26–12.28):

$$P_{\mathrm{d,CR}_i}(\tau_{\mathrm{s}}) = \begin{cases} P_{\mathrm{d,ced}}(\tau_{\mathrm{s}}), \text{ CED} \\ P_{\mathrm{d,enp}}(\tau_{\mathrm{s}}), \text{ ENP ED} \end{cases} \tag{12.26}$$

$$P_{\mathrm{f,CR}_i}(\tau_{\mathrm{s}}) = \begin{cases} P_{\mathrm{f,ced}}(\tau_{\mathrm{s}}), \text{ CED} \\ P_{\mathrm{f,enp}}(\tau_{\mathrm{s}}), \text{ ENP ED} \end{cases} \tag{12.27}$$

$$P_{\mathrm{m,CR}_i}(\tau_{\mathrm{s}}) = \begin{cases} P_{\mathrm{m,ced}}(\tau_{\mathrm{s}}), \text{ CED} \\ P_{\mathrm{m,enp}}(\tau_{\mathrm{s}}), \text{ ENP ED} \end{cases} \tag{12.28}$$

The total sensing probabilities at the FC by applying OR fusion rule is characterized by the following Eq. (12.29) (Kumar et al. 2019a, b):

$$P(X = H_1|H_1) = 1 - \prod_{i=1}^{N}\left(1 - P_{\mathrm{d,CR}_i}(\tau_{\mathrm{s}})\right) \tag{12.29}$$

$$P(X = H_0|H_0) = 1 - P(X = H_1|H_0) \tag{12.30}$$

$$P(X = H_1|H_0) = 1 - \prod_{i=1}^{N}\left(1 - P_{\mathrm{f,CR}_i}(\tau_{\mathrm{s}})\right) \tag{12.31}$$

$$P(X = H_0|H_1) = 1 - P(X = H_1|H_1) \tag{12.32}$$

where $P(X = \cdot|\cdot)$ represents global sensing decision.

### 12.3.3 Combined Final Decision About PU Status

In the projected technique, the global prediction as well as the global sensing decisions is merged, and a combined decision is made at the FC. Since the prediction and sensing probabilities are independent of one another, the combined probability of detection at the individual CR is given as per the following Eq. (12.33) (Bhowmick et al. 2017; Kumar et al. 2019a, b):

$$P_{\mathrm{d}} = P\left(\hat{X} = H_1|H_1\right)P(X = H_1|H_1) \tag{12.33}$$

$$P_{\mathrm{m}} = P\left(\hat{X} = H_0|H_1\right)P(X = H_0|H_1) \tag{12.34}$$

$$P_{\mathrm{f}} = P\left(\hat{X} = H_1|H_0\right)P(X = H_1|H_0) \tag{12.35}$$

$$P_{\mathrm{nf}} = P\left(\hat{X} = H_0|H_0\right)P(X = H_0|H_0) \tag{12.36}$$

where $P_{\mathrm{d}}$, $P_{\mathrm{m}}$, $P_{\mathrm{f}}$ and $P_{\mathrm{nf}}$ are detection probability, missed detection probability, false alarm probability, and no false alarm probability, respectively. All other possible probabilities in such circumstances are termed as ambiguous probabilities $P_{\mathrm{ambg}}$ and are given as Eqs. (12.37–12.40):

$$P_{\mathrm{ambg},1} = P\left(\hat{X} = H_1|H_1\right)P(X = H_0|H_1) \tag{12.37}$$

$$P_{\mathrm{ambg},2} = P\left(\hat{X} = H_1|H_0\right)P(X = H_0|H_0) \tag{12.38}$$

$$P_{\mathrm{ambg},3} = P\left(\hat{X} = H_0|H_1\right)P(X = H_1|H_1) \tag{12.39}$$

$$P_{\mathrm{ambg},4} = P\left(\hat{X} = H_0|H_0\right)P(X = H_1|H_0) \tag{12.40}$$

### 12.3.4 Enhancing Spectrum Reusability

In an EH-CRN with cooperative sensing, strengthening the efficiency of spectrum reusability largely influenced by the false alarm probability. Lesser the false alarm probability, the higher is the efficiency of spectrum reuse. Therefore, the enhanced efficiency of spectrum reuse is given by Eq. (12.41) (Bhowmick et al. 2017):

$$I_{\mathrm{use}} = \frac{\text{False alarm probability in conventional strategy}}{\text{False alarm probability in proposed strategy}}$$

$$I_{\text{use}} = \frac{1}{P\left(\hat{X} = H_1 | H_0\right)} \tag{12.41}$$

As $P\left(\hat{X} = H_1 | H_0\right) \leq 1$, $I_{\text{use}} \geq 1$.

**Proposition 1** *An increasing detection probability also ensures protection to the primary users against interference. By comparing the detection probability of the conventional approach to the detection probability of the projected approach, it is observed that* $P_{\text{d}} \leq P(X = H_1 | H_1)$. *Hence, it is necessary to adjust the threshold* $\lambda_{\text{CR}}$ *pertaining to each CR node to* $\bar{\lambda}_{\text{CR}}$. *As a result,* $P_{\text{d}} \geq P(X = H_1 | H_1)$. *Therefore, the QoS of the PUs can be preserved at a preferred level. To attain the preferred level of protection to the PU, it is essential to establish the combined probability of detection* $P_{\text{d}}$, *to a specified value of* $\overline{P}_{\text{d}}$. *The novel sensing threshold* $\bar{\lambda}_{\text{CR}}$ *is expressed as Eq. (12.42) (Kumar et al. 2019b):*

$$\bar{\lambda}_{\text{CR}} = \sigma_w^2 \left[ 1 + \gamma_{\text{r}} + \sqrt{\frac{2\gamma_{\text{r}} + 1}{\tau_{\text{s}} f_{\text{s}}}} Q^{-1} \left\{ 1 - \left( 1 - \frac{\overline{P}_{\text{d}}}{P\left(\hat{X} = H_1 | H_1\right)} \right)^{1/N} \right\} \right] \tag{12.42}$$

### 12.3.5 Energy Harvesting

In our proposed model, every CR node can harvest energy out of the RF signal of PU or conventional source according to the combined decision at the FC (Zhang et al. 2014). We consider that the harnessed energy from conventional methods follows a Poisson process having mean $E_{\text{nrf}}$. At the same time, the energy harvested per unit time from the PU signal is expressed as Eq. (12.43) (Bhowmick et al. 2016):

$$E_{\text{rf}} = P_{\text{pu}} / g_{\text{hm}} \tag{12.43}$$

where $P_{\text{pu}}$ represents PU transmission power, $g_{\text{hm}}$ is the mean value of channel gain. We consider that the channel gain obeys Weibull distribution when PU activity is modeled as Weibull distributed. However, the channel gain obeys Erlang distribution when PU behavior follows Erlang distribution.

The CR nodes are able to scavenge energy by two means, i.e., RF signal of PU and also ambient sources for the period of $(\tau_{\text{p}} + \tau_{\text{s}})$ time. However, throughout the transmission time $(T_{\text{r}})$, the CR selects any one of the harvesters to harvest energy based on the final decision at FC according to Table 12.1. Due to the lack of knowledge regarding the PU status at a CR, it holds both the harvesters ON during $(\tau_{\text{p}} + \tau_{\text{s}})$ time. Consequently, it is essential to determine the probability that the PU exists

all through the prediction-sensing duration. The probability that the PU is present throughout prediction-sensing time is written as Eq. (12.44):

$$P_{\mathrm{ac}} = \int_{0}^{\tau_{\mathrm{p}}+\tau_{\mathrm{s}}} P_{\mathrm{B}}(t)\mathrm{d}t \tag{12.44}$$

The total harvested energy for single and multiple cycles is written as Eq. (12.45):

$$E_{\mathrm{H}} = \begin{cases} E_{\mathrm{H,S}}; & M = 1 \\ E_{\mathrm{H,M}}; & M > 1 \end{cases} \tag{12.45}$$

where $E_{\mathrm{H,S}}$ gives the harvested energy in case of single cycle-based detection and $E_{\mathrm{H,M}}$ represents the harvested energy in case of multiple cycles-based detection. Harvested energy in case of multiple cycle-based detection is written as Eq. (12.46) (Bhowmick et al. 2017; Kumar et al. 2019a):

$$\begin{aligned} E_{\mathrm{H,M}} = M\big[((1-P_{\mathrm{ac}})E_{\mathrm{nrf}} + P_{\mathrm{ac}}E_{\mathrm{rf}})\big(\tau_{\mathrm{p}}+\tau_{\mathrm{s}}\big) + E_{\mathrm{a}} + E_{\mathrm{b}}\big] \\ + [X]^{(M-1)}[P(H_0)P_{\mathrm{f}}\{(T_{\mathrm{r}}-T_{\mathrm{c}})E_{\mathrm{nrf}} + T_{\mathrm{c}}E_{\mathrm{rf}}\}] \end{aligned} \tag{12.46}$$

where $E_{\mathrm{a}}$, $E_{\mathrm{b}}$, and $X$ are some of the essential mathematical terms and can be written as Eqs. (12.47–12.49) (Kumar et al. 2019a):

$$E_{\mathrm{a}} = P(H_1)P_{\mathrm{d}}(1-P_{\mathrm{dis}})T_{\mathrm{r}}E_{\mathrm{rf}} \tag{12.47}$$

$$E_{\mathrm{b}} = T_{\mathrm{r}}E_{\mathrm{nrf}}[P(H_1)P_{\mathrm{m}} + P(H_1)P_{\mathrm{AM1}} + P(H_0)P_{\mathrm{AM2}}] \tag{12.48}$$

$$X = [P(H_1)(P_{\mathrm{d}} + P_{\mathrm{AM1}}) + P(H_0)(P_{\mathrm{f}} + P_{\mathrm{AM2}})] \tag{12.49}$$

$T_{\mathrm{c}}$ is PU dwelling time after the reoccurrence of the PU in the channel and is given as Eq. (12.50) (Kim 2019):

$$\rho = \int_{0}^{T_{\mathrm{c}}} P_{\mathrm{B}}(t)\mathrm{d}t \tag{12.50}$$

where $\rho$ implies the collision constraint (Bhowmick et al. 2016; Yingxi and Ahmed 2014).

### 12.3.6 Cost of Energy Harvesting

The energy penalty in the course of harvesting is the quantity of energy required to operate the harvester. The energy penalty can be characterized as Eq. (12.51):

$$E_{\mathrm{h,pe}} = M\eta P_{\mathrm{s}}\left[\left(\tau_{\mathrm{p}} + \tau_{\mathrm{s}}\right) + P_{\mathrm{f}}T_{\mathrm{r}}\right] \tag{12.51}$$

where $P_{\mathrm{s}}$ is the power consumed, $\eta$ is the energy-splitting parameter, $\eta P_{\mathrm{s}} t_{\mathrm{sp}}$ is the energy required to run the harvester during prediction and sensing time and $\eta P_{\mathrm{s}} P_{\mathrm{f}} T_{\mathrm{r}}$ is the amount of energy needed to operate RF harvester during a $T_{\mathrm{r}}$ time while PU is not present.

### 12.3.7 Throughput of CRN

The CR node begins transmission only when the CR comes up with a combined final decision of $P\left(\hat{X} = H_0, X = H_0 | H_0\right)$. Therefore, the useful normalized throughput is obtained as Eq. (12.52) (Kumar et al. 2019a):

$$R_0 = \begin{cases} R_{0,\mathrm{S}}; & M = 1 \\ R_{0,\mathrm{M}}; & M > 1 \end{cases} \tag{12.52}$$

Here, $R_{0,\mathrm{S}}$ and $R_{0,\mathrm{M}}$ are the normalized throughput for single and multiple cycles-based detection, respectively. Normalized throughput for a single cycle and multiple cycle-based detections (Khalili et al. 2013) is written as Eqs. (12.53) and (12.54) (Bhowmick et al. 2017; Kumar et al. 2019a, b):

$$R_{0,\mathrm{S}} = \left[\frac{T - \tau_{\mathrm{p}} - \tau_{\mathrm{s}} - T_{\mathrm{c}}}{T}\right] P(H_0) P_{\mathrm{nf}} C_0 \tag{12.53}$$

$$\begin{aligned} R_{0,\mathrm{M}} = &\left[\frac{T - \tau_{\mathrm{p}} - \tau_{\mathrm{s}} - T_{\mathrm{c}}}{T}\right] P(H_0)[P(H_1)(P_{\mathrm{d}} + P_{\mathrm{AM1}}) \\ &+ P(H_0)(P_{\mathrm{f}} + P_{\mathrm{AM2}})]^{(M-1)} P_{\mathrm{nf}} C_0 \end{aligned} \tag{12.54}$$

where $C_0 = \left[\frac{T-\tau_{\mathrm{p}}-\tau_{\mathrm{s}}-T_{\mathrm{c}}}{T}\right] \log\left(1 + |h|^2 P_{\mathrm{tr}} / \sigma_w^2\right)$, $P_{\mathrm{tr}}$ is transmitted power. The transmitted power in the case of single cycle and multiple cycle-based detections is given as Eqs. (12.55–12.57) (Bhowmick et al. 2017):

$$P_{\mathrm{tr}} = \begin{cases} P_{\mathrm{tr,s}}; & M = 1 \\ P_{\mathrm{tr,M}}; & M > 1 \end{cases} \tag{12.55}$$

$$P_{\mathrm{tr,s}} = \frac{E_{\mathrm{H,S}} - \eta P_{\mathrm{s}} T + (1-\eta) P_{\mathrm{s}}\left(\tau_{\mathrm{p}} + \tau_{\mathrm{s}}\right) - E_{\mathrm{R}}}{P(H_0) P_{\mathrm{nf}}\left(T - \tau_{\mathrm{p}} - \tau_{\mathrm{s}} - T_{\mathrm{c}}\right)} \tag{12.56}$$

$$P_{\mathrm{tr,M}} = \frac{E_{\mathrm{H,M}} - M\left[\eta P_{\mathrm{s}} T + (1-\eta) P_{\mathrm{s}}\left(\tau_{\mathrm{p}} + \tau_{\mathrm{s}}\right)\right] - E_{\mathrm{R}}}{\left[P(H_1)(P_{\mathrm{d}} + P_{\mathrm{AM1}}) + P(H_0)(P_{\mathrm{f}} + P_{\mathrm{AM2}})\right]^{(M-1)} P_{\mathrm{nf}}\left(T - \tau_{\mathrm{p}} - \tau_{\mathrm{s}} - T_{\mathrm{c}}\right)} \tag{12.57}$$

Here, $E_{\mathrm{R}}$ gives the residual energy.

**Proposition 2** *The CRN decides the power required for transmission by preserving $E_{\mathrm{R}}$ volume of energy for the subsequent detection cycles so that the succeeding detection cycles are not subjected to energy outage.*

## 12.4 Results and Discussions

In this segment, the mathematical results and discussions are described for the projected scheme. We investigate the performance of CRN under the following system parameters: $P_{\mathrm{pu}} = 1$ W, $E_{\mathrm{R}} = 0.01$ J, $\alpha = 0.75$, $a = 0.75$, $\beta = 0.25$, $b = 0.25$, $N = 5$, $\eta = 0.1$, $\sigma_{\mathrm{s}}^2 = 1$, $\sigma_{\mathrm{w}}^2 = 1$, $\gamma_{\mathrm{s}} = -16$ dB, $f_{\mathrm{s}} = 6$ MHz, $E\_\mathrm{nrf} = 2$ J, $T = 100$ ms, $P_{\mathrm{s}} = 0.1$ W, and $\overline{P}_{\mathrm{d}} = 0.9$. Detection probability contributes significantly in determining the channel status. The channel utilization efficiency rises with the rise in detection probability.

Detection probability contributes significantly to determining the channel status. The channel utilization efficiency rises with the rise in detection probability. In Fig. 12.3, the detection probability about the number of samples for a definite prediction error is presented. It is seen that the detection probability rises with the rise in the number of samples. From Fig. 12.3, we note that ENP-based ED outperforms CED. The number of CR nodes in cooperation plays a major part in determining the detection probability and therefore is a key metric in assessing the performance of the proposed system.

Figure 12.4 shows that there is a rise in the detection probability with the rise in the number of CR nodes that lie under cooperation which is intuitively satisfying. False alarm probability brings about a substantial effect on the quality of service of the PU. The quality of service of PU improves as the false alarm probability drops.

Fig. 12.5 clearly demonstrates the probability of false alarm considering a particular prediction error that varies with the number of samples. The false alarm probability falls with the rise in the number of samples, and we found that ENP-based ED outperforms CED.

Also, the false alarm probability varying with the number of nodes under cooperation is simulated, and the findings are displayed in Fig. 12.6, which indicates that it falls with the increase in the number of CR nodes under cooperation.

The spectrum reuse pattern for both the proposed prediction-based scheme and the conventional scheme is shown in Fig. 12.7. It is found that the spectrum utiliza-

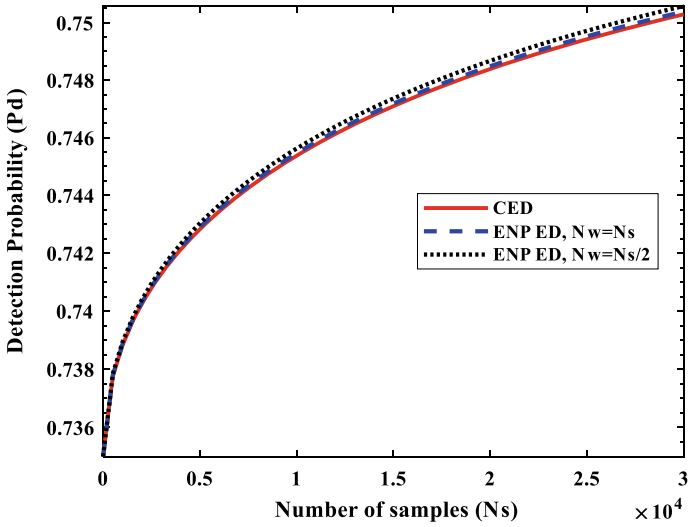



**Fig. 12.3** Detection probability versus number of samples for different sensing schemes, i.e., CED and ENP ED

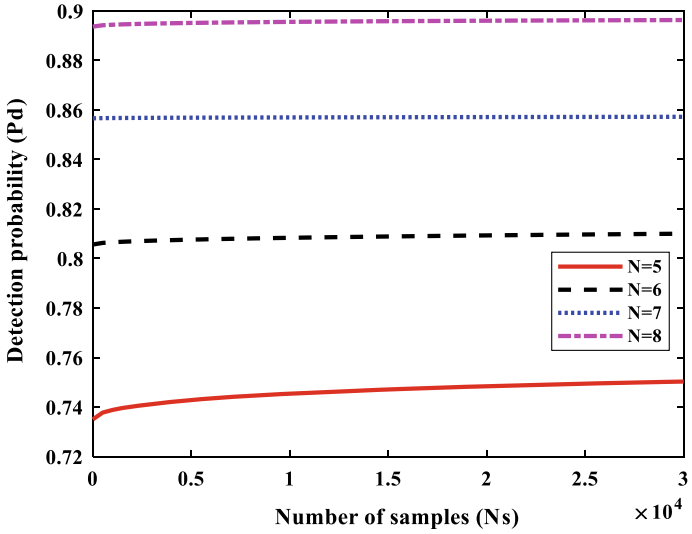



**Fig. 12.4** Detection probability for CED and ENP ED for different number of nodes

tion is much better in the prediction-based model with respect to the conventional scheme. From Fig. 12.7, it is seen that for a particular prediction error $P_e = 0.1$, the improvement in spectrum reuse increases by 200% in the prediction-based model.

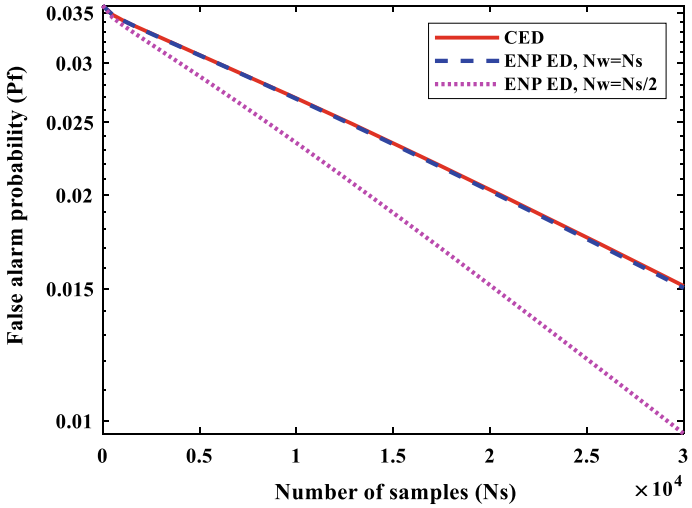



**Fig. 12.5** False alarm probability for the different sensing approaches (CED and ENP ED)

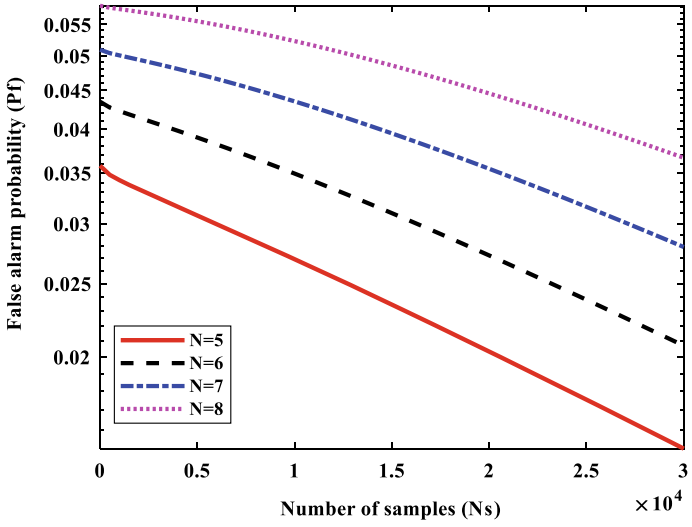



**Fig. 12.6** False alarm probability for different node counts

However, with higher prediction error, the improvement is decreasing because of the decrease in detection probability, which is intuitively satisfying.

The consequence of prediction error on the energy-harvesting procedure for Weibull and Erlang distribution with diverse shape parameters is depicted in Fig. 12.8. It can be viewed that the amount of energy harvested falls with the rise in prediction

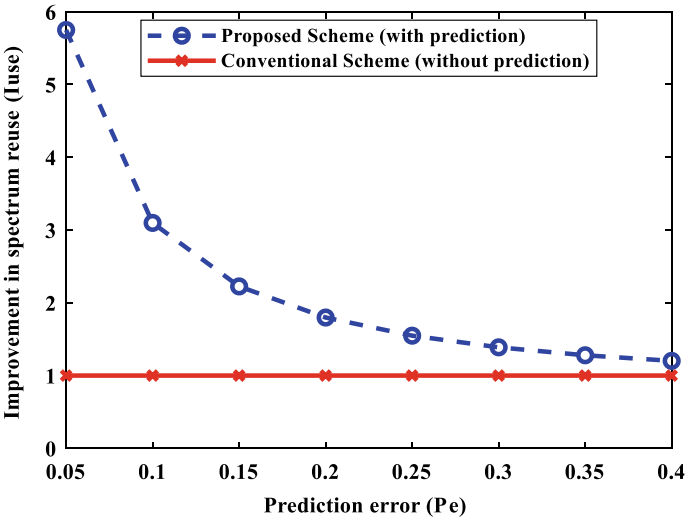



**Fig. 12.7** Enhancing spectrum reuse

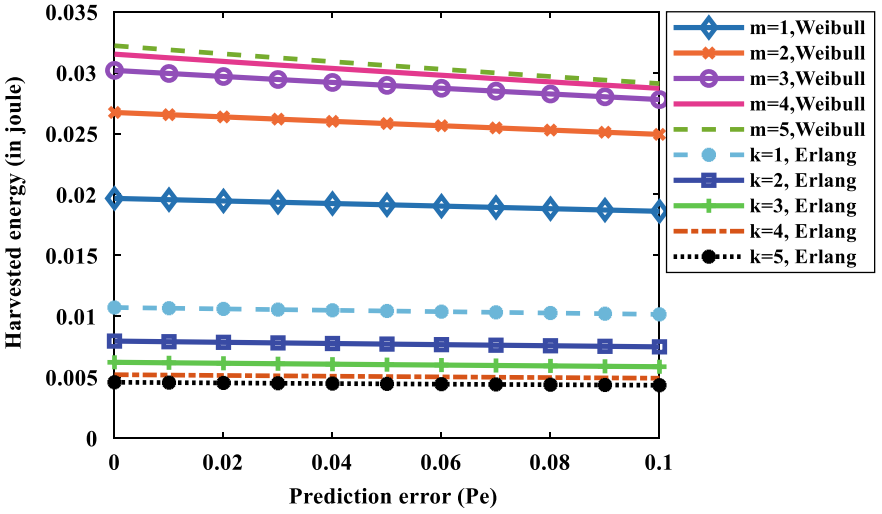



**Fig. 12.8** Implication of prediction error over the harvested energy

error for both the distribution function models. However, in the case of Weibull distribution with a higher value of *m,* the harvested energy increases because $E_{\mathrm{rf}}$ increases with $m$ as per Eq. (12.43). In the case of Erlang distribution, the harvested energy decreases with the increase in type ($k$) of distribution. For a particular prediction error $P_{\mathrm{e}} = 0.1$, as the type of Weibull distribution increases $m = 1$–2, the harvested energy increases by 33.87%, and as type rises from $m = 4$ to 5, the harvested energy increases

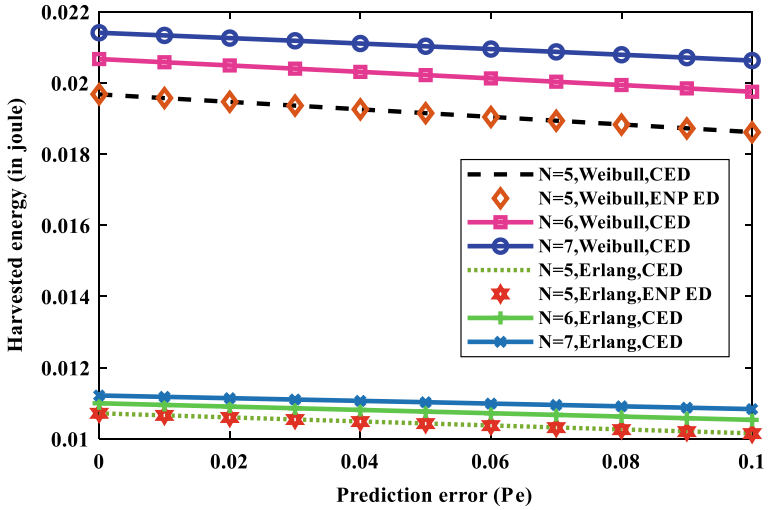



**Fig. 12.9** Impact of $N$ on harvested energy

by 1.39%. In the case of Erlang distribution for a particular prediction error $P_e = 0.1$, as the type increases $k = 1$–2, the harvested energy decreases by 26.47%, whereas for $k = 4$–5, the harvested energy decreases by 12.24%. We observed that harvested energy is higher in Weibull distribution as compared with Erlang distribution. We can conclude that Weibull distribution outperforms exponential distribution.

The influence of the number of CR nodes under cooperation on the harvested energy is demonstrated in Fig. 12.9. We noticed that as the number of CR nodes under cooperation goes up, the overall harvested energy increases, which is intuitively valid. Given a fixed $N$, the harvested energy is almost the same for CED as well as ENP-based ED.

Figure 12.10 exhibits the influence of the number of frames upon the harvested energy. The harvested energy ascends with an increase in the number of frames. The harvested energy tends to decrease concerning the increasing prediction error. It may be noted that as more frames are transmitted, the harvested energy also increases. The harvested energy for $M = 3$ and Weibull distributed PU activity model harvest the highest amount of energy.

The throughput of the network is one of the key measuring parameters for assessment of its performance under the presented scenario as a function of other system parameters. The affect of prediction error over the network throughput is depicted in Fig. 12.11. The throughput drops with the rising prediction error. We note that the throughput drops with the rise in type ($m$ or $k$) of the distribution for both the Weibull and Erlang distribution. In the case of Weibull distribution for $P_e = 0.1$, the throughput decreases by 41.06% as the type increases from one to two (i.e., $m = 1$–2), whereas in the case of Erlang distribution, the same is decreased by 41.17% with an increase in $k$ from 1 to 2. From Fig. 12.11, we found that the throughput of

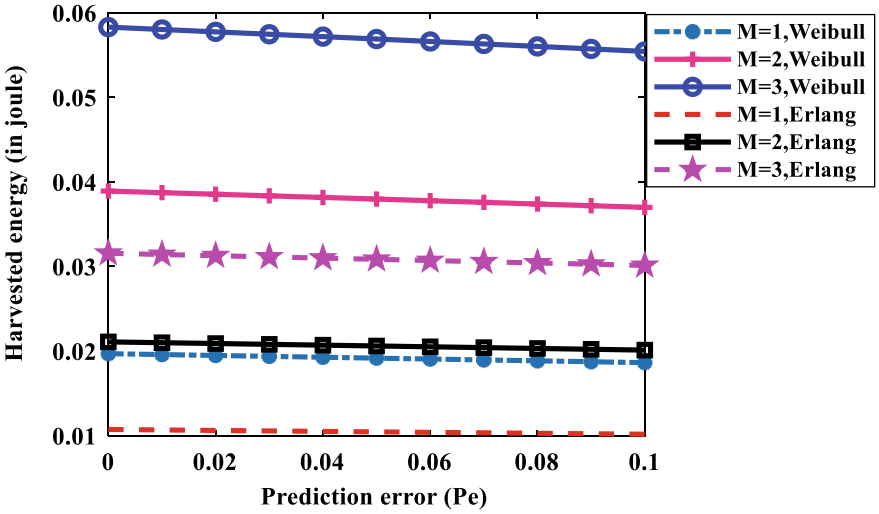



**Fig. 12.10** Variation of harvested energy with the number of frames

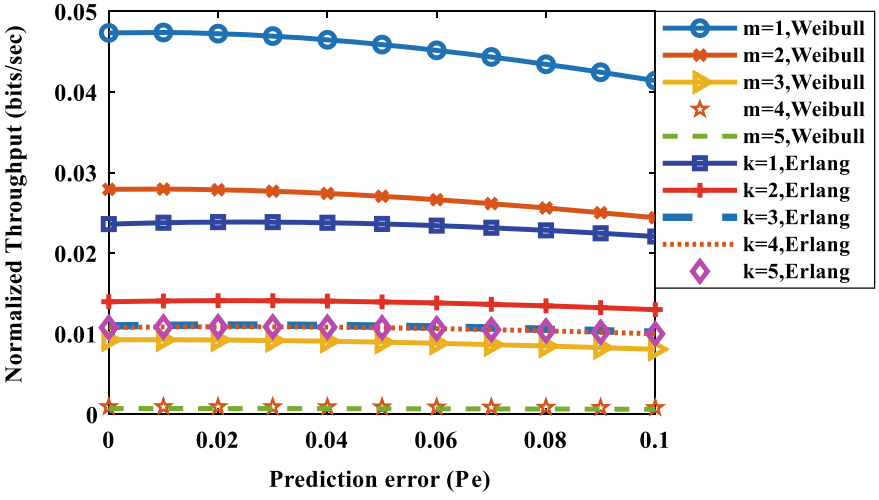



**Fig. 12.11** Variation of the normalized throughput concerning prediction error

CR node is greater when PU behavior follows Weibull distribution as compared with Erlang distribution for a particular type.

Normalized throughput for a fixed number of frames is demonstrated as a function of sensing period in Fig. 12.12. From the results shown in Fig. 12.12, we note that the normalized throughput drops with sensing interval for a specific frame. We also observed that as the type of the distribution function increases, the throughput of CR node decreases for both Weibull and Erlang distribution models.

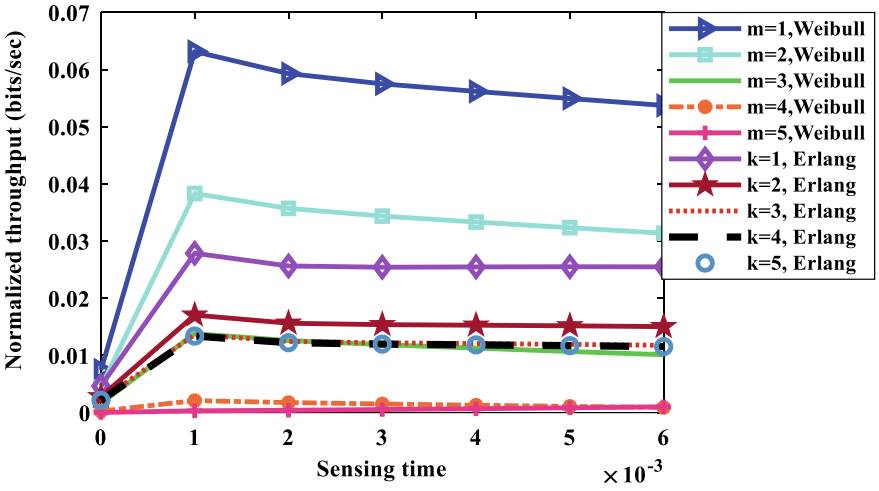



**Fig. 12.12** Variation of throughput with respect to sensing time

In Fig. 12.13, the effect on the throughput as a result of the number of frames is shown. We notice that as the number of frame rises, the throughput drops because the transmitted power of the CR node falls with the number of frames. The throughput of ENP-based ED performs better than CED for a specific number of frames.

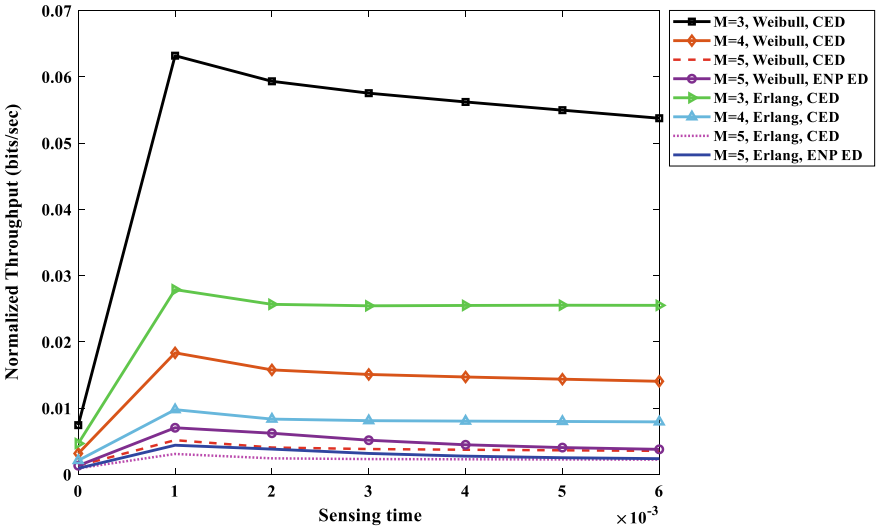



**Fig. 12.13** Impact of a number of the frame on throughput

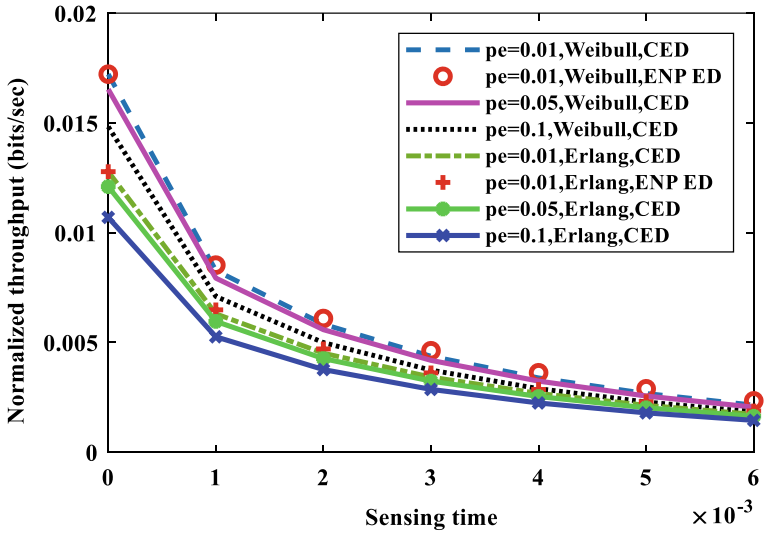



**Fig. 12.14** Effect of prediction error on throughput

The influence of prediction error on throughput for a particular frame is also simulated, and the result is illustrated in Fig. 12.14, which asserts that an increasing prediction error results in the drop of throughput of a CR node.

The energy penalty for various frames concerning the sensing interval is also obtained as shown in Fig. 12.15. It can be noted that the energy penalty upsurges with the number of frames for a specific sensing duration. The energy penalty is greater in ENP-based ED as against CED for a particular number of frames.

## 12.5 Conclusion

In this chapter, the performance of an EH-enabled cooperative CRN is investigated for Weibull and Erlang distribution, and a comprehensive comparative analysis is made. The mathematical equations for the probability of detection and false alarm, improvement in spectrum reuse, energy harvested, energy penalty throughout harvesting, and the normalized throughput are established for both the distribution models. Impact of prediction error, collision constraint, number of CR nodes under cooperation, number of frames, and shape parameter of the distribution functions on the performance of EH-CRN is investigated, and simulation is performed to obtain detailed results of the proposed model under specific situations. It is seen that the energy harvested and the throughput of the EH-enabled CR in a centralized cooperative network are better in the case of the Weibull distribution model when compared with the Erlang distribution model. As the type of the distribution grows, the harvested energy rises, and the throughput of the CR node decreases in the case of Weibull distribution.

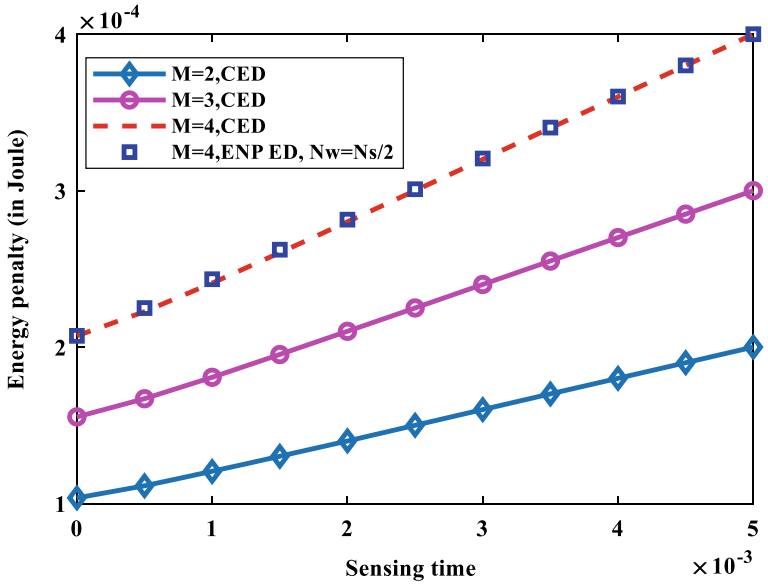



**Fig. 12.15** Energy penalty during harvesting

However, in the case of Erlang distribution, the harvested energy and throughput of the CR nodes both decrease. It is seen that for type-II Weibull, the throughput and harvested energy are considerably good and therefore may be useful in designing EH–CRN. The detailed results for the majority of the performance measuring parameters are presented in this chapter. The results are encouraging and are expected to be useful for the practical implementation of EH-CRN. The paper does not consider the decentralized cooperation scheme and is kept as future work.

**Acknowledgements** This work is also partially supported by 'Project NDS and DiCoBa,' sanction order no.: 21(1)/2015-CCBT dated 30.09.2015 from MEITY, Government of India.

## References

Ahmed ME, Kim DI, Kim JY, Shin Y (2017) Energy-arrival-aware detection threshold in wireless-powered cognitive radio networks. IEEE Trans Veh Technol 66(10):9201–9213

Akhtar AN, Arif F, Siddique AM (2018) Spectrum decision framework to support cognitive radio based IoT in 5G. In: Cognitive radio in 4G/5G wireless communication systems. IntechOpen. https://doi.org/10.5772/intechopen.80991

Akyildiz F, Lee W-Y, Vuran MC, Mohanty S (2006) Next generation/dynamic spectrum access/cognitive radio wireless networks—a survey. Comput Netw 50(13):2127–2159

Akyildiz IF, Lee W-Y, Chowdhury KR (2009) CRAHNs: cognitive radio ad hoc networks. Ad Hoc Netw 7(5):810–836

Akyildiz IF, Lo BF, Balakrishnan R (2011) Cooperative spectrum sensing in cognitive radio networks: a survey. Phys Commun 4(1):40–62
Arif W, Hoque S, Sen D, Baishya S (2015) A comprehensive analysis of spectrum handoff under different distribution models for cognitive radio networks. Wirel Pers Commun 85(4):2519–2548
Barani B (2020) Beyond 5G, a European perspective. In: Sixth ITU workshop on network 2030
Bhandari S, Joshi S (2018) Cognitive radio technology in 5G wireless communications. In: Second IEEE international conference on power electronics, intelligent control and energy systems (ICPEICES). https://doi.org/10.1109/icpeices.2018.8897345
Bhowmick A, Roy S, Kundu S (2016) Throughput of a cognitive radio network with energy harvesting based on primary user signal. IEEE Wirel Commun Lett 5(2):136–139. https://doi.org/10.1109/LWC.2015.2508806
Bhowmick A, Yadav K, Roy SD, Kundu S (2017) Throughput of an energy harvesting cognitive radio network based on prediction of primary user. IEEE Trans Veh Technol 66(9):8119–8128
Borgne YL, Santini S, Bontempi G (2007) Adaptive model selection for time series prediction in wireless sensor networks. Signal Process 87(12):3010–3020
Digham FF, Alouini M-S, Simon MK (2007) On the energy detection of unknown signals over fading channels. IEEE Trans Commun 55(1):21–24
Federal Communications Commission (FCC) (2003) Notice for proposed rulemaking (NPRM 03 322): facilitating opportunities for flexible, efficient, and reliable spectrum use employing spectrum agile radio technologies. ET Docket No. 03 108, Dec 2003
Gardner W (1988) Signal interception: a unifying theoretical framework for feature detection. IEEE Trans Commun 36(8):897–906
Gardner WA (1994) Cyclostationarity in Communications and Signal Processing. IEEE Press, IEEE Communications Society, Sponsor
Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23:201–220
Hoque S, Arif W (2017) Performance analysis of cognitive radio networks with generalized call holding time distribution of secondary user. Telecommun Syst 66(1):95–108
Hoque S, Arif W (2018) Impact of secondary user mobility on spectrum handoff under generalized residual time distributions in cognitive radio networks. AEU Int J Electron Commun 1(86):185–194
Hoque S, Arif W, Sen D, Baishya S (2018) Analysis of spectrum handoff under general residual time distributions of spectrum holes in cognitive radio networks. J Inf Sci Eng 34(4)
Jondral FK (2005) Software-defined radio-basics and evolution to cognitive radio. EURASIP J Wirel Commun Netw 3:275–283
Khalili S, Jamshidi A, Derakhtian M (2013) Total throughput maximisation in a cognitive radio network. IET Commun 7(11):1051–1060
Kim SW (2019) Simultaneous spectrum sensing and energy harvesting. IEEE Trans Wirel Commun 18(2):769–779
Kumar D, Talukdar B, Arif W (2019a) Performance analysis of prediction based sensing in energy harvesting cooperative CRN. In: 2019 second international conference on advanced computational and communication paradigms (ICACCP), pp 1–6. https://doi.org/10.1109/icaccp.2019.8882991
Kumar D, Talukdar B, Arif W (2019b) Impact of Weibull distribution on prediction based sensing in energy harvesting cooperative CRN. In: Sixth international conference on signal processing and integrated networks (SPIN), pp 704–709. https://doi.org/10.1109/spin.2019.8711729
Lee S, Zhang R, Huang K (2013) Opportunistic wireless energy harvesting in cognitive radio networks. IEEE Trans Wirel Commun 12(9):4788–4799
Lee HS, Ahmed ME, Kim DI (2018) Optimal spectrum sensing policy in RF-powered cognitive radio networks. IEEE Trans Veh Technol 67(10):9557–9570
Liang Y-C, Zeng Y, Peh ECY, Tuan Hong A (2008) Sensing-throughput tradeoff for cognitive radio networks. IEEE Trans Wirel Commun 7(4):1326–1337
Ma J, Li G, Juang BH (2009) Signal processing in cognitive radio. Proc IEEE 97(5):805–823

Mao S, Cheung MH, Wong V (2012) An optimal energy allocation algorithm for energy harvesting wireless sensor networks. In: Proceedings of IEEE international conference on communications, pp 265–270
Mariani A, Giorgetti A, Chiani M (2011) Effects of noise power estimation on energy detection for cognitive radio applications. IEEE Trans Commun 59(12):3410–3420
Mariani A, Kandeepan S, Giorgetti A (2015) Periodic spectrum sensing with non-continuous primary user transmissions. IEEE Trans Wirel Commun 14(3):1636–1649
Mitola J III, Maguire GQ Jr (1999) Cognitive radio: making software radios more personal. IEEE Pers Commun Mag 6(4):13–18
Park S, Hong D (2013) Achievable throughput of energy harvesting cognitive radio networks. IEEE Trans Wirel Commun 12(12):6166–6179
Park S, Hong D (2014) Optimal spectrum access for energy harvesting cognitive radio networks. IEEE Trans Wirel Commun 13(5):2601–2613
Shafie AE (2014) Space-time coding for an energy harvesting cooperative secondary terminal. IEEE Commun Lett 18(9):1571–1574
Shafie AE, Khattab T (2014) Maximum throughput of a cooperative energy harvesting cognitive radio user. In: Proceedings of 25th annual international symposium on personal, indoor, and mobile radio communication, pp 1067–1072
Shafie AE, Ashour M, Khattab T, Mohamed A (2015) On spectrum sharing between energy harvesting cognitive radio users and primary users. In: Proceedings of international conference on computing, networking and communications, pp 214–220
Shared Spectrum Company. [Online]. http://www.sharedspectrum.com/
Sharma SK, Rawat DB (2016) How can 5G wireless benefit from cognitive radio principles? EAI. [Online]. Available: https://blog.eai.eu/how-can-5g-wireless-benefit-from-cognitive-radio-principles/
Shi Z, Teh KC, Li KH (2013) Energy-efficient joint design of sensing and transmission durations for protection of primary user in cognitive radio systems. IEEE Commun Lett 17(3):565–568
Singh A, Bhatnagar M, Mallik R (2016) Performance of an improved energy detector in multi-hop cognitive radio networks. IEEE Trans Veh Technol 65(2):732–743
Tandra R, Sahai A (2008) SNR walls for signal detection. IEEE J Sel Top Signal Process 2(1):4–17
Xing X, Jing T, Cheng W, Huo Y, Cheng X (2012) Spectrum prediction in cognitive radio networks. In: Proceedings of IEEE GLOBECOM, pp 3982–3987
Yingxi L, Ahmed T (2014) Primary traffic characterization and secondary transmissions. IEEE Trans Wirel Commun 13(6):3003–3016
Yuan G, He H, Deng Z, Zhang X (2018) Cognitive radio network with energy harvesting based on primary and secondary user signals. IEEE Access 6:9081–9090
Yucek T, Arslan H (2009) A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun Surv Tutor 11(1):116–130
Zhang R, Ho CK (2013) MIMO broadcasting for simultaneous wireless information and power transfer. IEEE Trans Wirel Commun 12(5):1989–2001
Zhang J, Zheng F-C, Gao X-Q, Zhu H-B (2014) Sensing-energy efficiency tradeoff for cognitive radio networks. IET Commun 8(18):3414–3423

# Chapter 13
# Impact of Buffer Size on Proactive Spectrum Handoff Delay in Cognitive Radio Networks


**Shanidul Hoque, Banani Talukdar, and Wasim Arif**

**Abstract** Spectrum handoff (SH) is a vital process to guarantee seamless and effective services of secondary users (SUs) in cognitive radio networks (CRNs). SH delay has a negative impact on the performance of SUs. For simplicity, the PRP M/G/1 queuing model is used in literature to evaluate the SH delay parameters of CUs in a CRN. However, the design of an infinite buffer size queue in a real-time teletraffic system is not feasible. We present pre-emptive resume priority (PRP) M/G/1/K queuing model comprising of three priority queues: primary user (PU) queue for higher priority PUs, interrupted user (IU) queue for moderate priority interrupted SUs and SU queue for lower priority newly arrived SUs, to derive the SH performance metrcs such as blocking probability and cumulative handoff delay (CHD) of SUs. This chapter analyses the impact of buffer length ($K$) on blocking probability and CHD for various proactive SH schemes: non-switching, switching and random SH schemes in CRNs. We present and summarise the detailed comparison of results for blocking probability and CHD in terms of PUs' arrival rate and mobility parameter of spectrum holes for different $K$ under PRP M/G/1/K queuing network model. Results show that the blocking probability decreases and the CHD increases with increasing value of $K$. For an optimal value of $K$, the proposed model offers similar performance to the PRP M/G/1 queuing network model.

**Keywords** Cognitive radio · Handoff delay · M/G/1/K queuing network model · Proactive spectrum handoff

S. Hoque
Department of Electronics and Communication Engineering, Madanapalle Institute of Technology & Science, Madanapalle, India
e-mail: shanidulhoque257@gmail.com

B. Talukdar · W. Arif (✉)
Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, India
e-mail: arif.ece.nits@gmail.com

B. Talukdar
e-mail: talukdarbanani.smit@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_13

## 13.1 Introduction

The accelerated progress in wireless services and users causes the spectrum scarcity problem in the network, and hence, the network requires more spectrum band to fulfil the user demands (Wang et al. 2014; Zheng et al. 2015; Kumar et al. 2016). The data rate and the data traffic in mobile network are expected to expand by three and seven times, respectively, by 2021 as per the report of CISCO virtual network indexing (Cisco Visual Networking Index 2016, 2019). In addition, owing to the fixed spectrum allotment policy, there exists a great discrepancy between spectrum allocation and spectrum usage. The Federal Communication Commission (FCC) has shown that the spectrum occupancy between 0 and 6 GHz spectrum bands differs by 15–85% (FCC 2002), which leads to spectrum inefficiency problem. The Shared Spectrum Company performed the experiment on spectrum usage in the radio spectrum below 3 GHz near downtown Washington DC in Shared Spectrum Company. The agency concluded that the frequency utilisation varies up to 35% in a most crowded area, and hence, there is a frequency underutilisation problem (Shared Spectrum Company). In 2002, National Radio Astronomy Observatory (NRAO) made an experiment on spectrum measurement in New York City (McHenry 2003). As per the report of NRAO (McHenry 2003), the minimum, average and maximum spectrum utilisation are 1%, 5.2% and 13.1%, respectively, over all of the regions in New York City. To solve this problem, FCC approved the use of lower priority secondary users (SUs) in the licensed band in the absence of higher priority licensed or primary users (PUs) (Federal Communications Commission (FCC) 2003). The SU is defined as cognitive radio (CR) (Mitola and Maguire 1999), which should have cognitive property and can solve spectrum scarcity and inefficiency problems via dynamic spectrum access policy with key issue spectrum handoff (Mitola 2001; Haykin 2005; Akyildiz et al. 2006). The word "cognitive radio" was familiarised by Mitola and was defined as "A radio that employs model based on reasoning to achieve a specified level of competence in radio-related domains" (Mitola and Maguire 1999; Mitola 2001). Haykin also defined CR in his invited paper by means of two main objectives: (i) extremely reliable and seamless communication and (ii) effective utilisation of radio spectrum band whenever and wherever needed to retain the quality of service (QoS) of the CRN (Haykin 2005). CR is a smart device that can vary its radio operating parameter in the current scenario according to the situation of surroundings.

The spectrum management framework of CR is composed of four key functions (Akyildiz et al. 2006, 2008): spectrum sensing, spectrum decision, spectrum sharing and spectrum mobility. Spectrum sensing is a vital function of spectrum management framework, which identifies a spectrum band whether it is available or using by a user (Akyildiz et al. 2006). If the spectrum is available, the CR can access that, otherwise, the CR will perform spectrum sensing to other spectrum band in search of spectrum hole over the network. Subsequent to identifying the available spectrum holes, the spectrum decision function is accountable for selecting the best spectrum hole for CR, which is suitable for its transmission (Akyildiz et al. 2008, 2009). If more number of CRs are interested in same spectrum band, the spectrum sharing function

shares the spectrum fairly among CRs according to their requirements (Akyildiz et al. 2008). Spectrum mobility is a key function of CR which is the final step of spectrum management framework. Figure 13.1 illustrates the concept of mobility of spectrum holes at a particular spectrum band in different time $T_1$, $T_2$, and $T_3$, respectively. Here, the green boxes specify the spectrum that is being used by the users and vacant spaces indicate the white spaces or spectrum holes, which can be utilised by CRs. As shown in Fig. 13.1, the availability of spectrum in a particular band depends on the arrival and departure of the users in the system at that particular moment. In the absence of a PU, an SU (or CR) can acquire that channel to start its transmission. Upon sensing the appearance of a PU in the same channel, the SU immediately handovers that channel to the PU on its arrival and transfers the ongoing communication to another spectrum hole. Therefore, the prime objective of spectrum mobility is to sustain the ongoing service of a SU by offering seamless switchover from a channel to a different available channel for better QoS and throughput of the CRN (Kumar et al. 2016; Christian et al. 2012). The spectrum mobility function consists of two phases (Christian et al. 2012): spectrum handoff (SH) and connection management. This chapter focuses on SH parameters of an SU in CRN.

When a PU attains a channel that is currently in use by an SU, the SU immediately leaves that channel to the PU and shifts its ongoing communication to an empty channel. This method of transferring is called spectrum handoff (SH) process and is responsible for the seamless connectivity and robust services of the SUs in CRNs (Akyildiz et al. 2008, 2009). The successful spectrum handoff process improves the performance of both the SU and CRN; however, this process also offers the additional

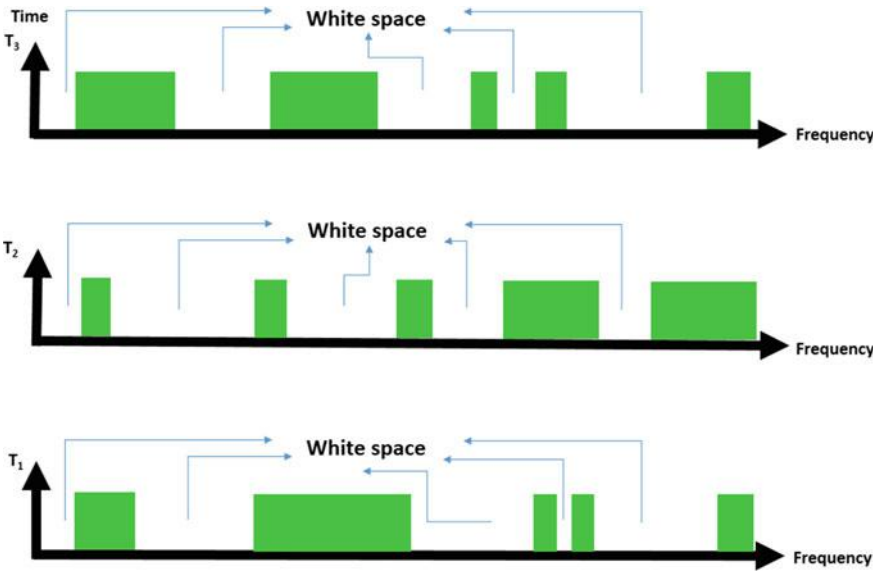



**Fig. 13.1** Concept of spectrum mobility in a particular spectrum band

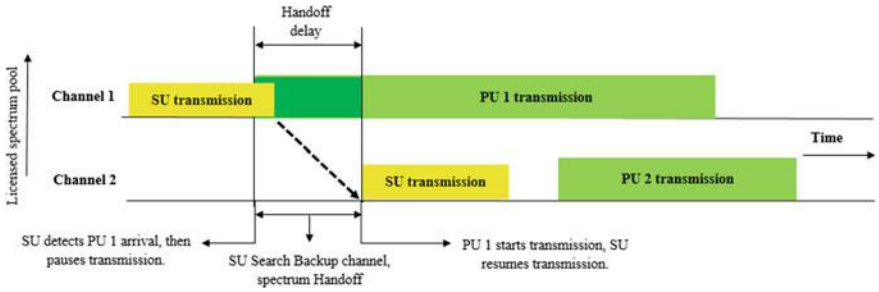



**Fig. 13.2** Illustration of reactive SH scheme

delay to SU service that has an adverse effect on the operation of a SU (Christian et al. 2012). The SH process can be usually classified as reactive and proactive SH schemes depending on the SH triggering decision time (Kumar et al. 2016; Christian et al. 2012; Wang and Wang 2008). In reactive SH scheme (Christian et al. 2012), the SU searches for available spectrum after the SH triggering event is initiated and thereafter transfers the remaining transmission to the new available channel as shown in Fig. 13.2. However, in proactive SH scheme (Christian et al. 2012), the SU makes a list of available channels during its service before the SH triggering event is initiated as illustrated in Fig. 13.3. Therefore, the sensing time is minimised in proactive decision SH scheme which leads to shorter handoff delay as compared to reactive SH scheme (Wang and Wang 2008). In Wang and Wang (2008, 2009) and Wang et al. (2012), the authors explored the delay performance of SUs in case of proactive and reactive SH schemes using pre-emptive resume priority (PRP) M/G/1 queuing model. In Wang and Wang (2009), a target channel selection-based approach is presented for minimising the SH delay and total service period of SUs in CRNs. In Lee and Yeo (2015), the authors investigated the SH delay and channel availability due to the arrival of PUs in CRNs. To minimise the SH delay and the packet loss, an improved handoff algorithm is proposed in Mathonsi and Kogeda (2016). Bayrakdar and Çalhan (2016) investigated the performance of proactive SH process for MAC protocol in CRNs. A hybrid SH model of proactive and reactive SH schemes is

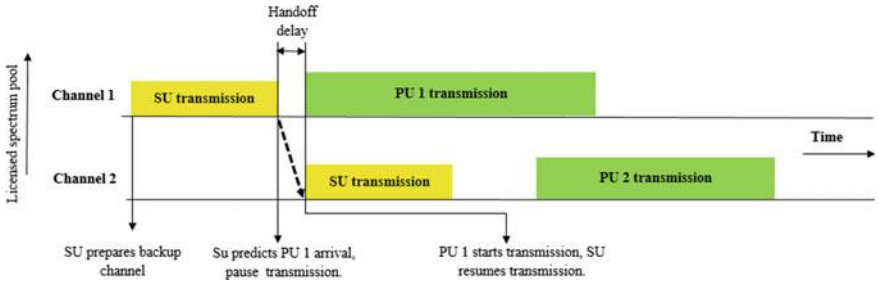



**Fig. 13.3** Illustration of proactive SH scheme

proposed for effective and robust communication of SUs in Ad Hoc-CRN (Gkionis et al. 2017). In Zakariya and Rabia (2016) and Zahed et al. (2013), the authors proposed a PRP M/G/1 queuing model based on the priority of the PUs (highest priority to PU queue) and SUs (based on the number of interruption) in CRN. The authors have classified the interrupted SU queues based on the interruption number and given higher priority to the SU with higher number of disruption so as to lower the handoff delay. In Wu et al. (2013, 2017), the authors proposed a hybrid priority queuing model consist of PRP and non-PRP M/G/1 for PUs and SUs, respectively, to analyse the delay performance of SUs in CRNs. In these literature, the authors analysed the delay performance considering infinite allowable interruption of SUs using infinite length M/G/1 queuing model in CRN. Nevertheless, the consideration of infinite buffer size by no means is feasible for the implementation of practical queuing network.

In Hoque et al. (2019), the authors analysed the delay performance in terms of cumulative handoff delay (CHD) for non-switching, switching and random proactive SH schemes considering finite allowable interruption for SUs using PRP M/G/1/K queuing network. However, it is important to offer services to an interrupted SU than starting of a newly arrived SU service. In Shekhar et al. (2019), the authors modelled the CHD for proactive decision switching SH schemes using PRP M/G/1/K queuing network allowing any number of interruptions for SUs. In this chapter, we consider a PRP M/G/1/K queuing model consists of three queues (higher priority PU queue, moderated priority interrupted user (IU) queue and lower priority SU queue) which allows any number of interruptions to the SUs throughout its total service period. This chapter models the performance measuring metrics and analyses the impact of buffer length ($K$) on blocking probability and CHD for various proactive SH schemes: non-switching, switching and random SH schemes using PRP M/G/1/K queuing network.

## 13.2 System Model

Figure 13.4 shows a PRP M/G/1/K queuing network model comprising of two wireless channels in a system. Both wireless channels consist of three queues with different priorities: higher priority PU queue, moderate priority interrupted user (IU) queue and lower priority SU queue. In the presented system, PU queue has the highest priority where $K$ PUs can wait for the service. The IU queue and SU queue are for the interrupted SUs and newly arrived SUs in the system, respectively. The IU queue has the lower priority than PU queue and higher priority than the SU queue. We consider finite buffer size $K$ for the all priority queues. Since the PU has the higher authorisation to access a channel, therefore, a PU can pre-empt the ongoing service of an SU in a CRN. If the channel is available and there is no PU in the PU queue, an SU can start its service by accessing the channel. However, when a PU appears in that channel, the SU will immediately leave the channel to arrival PU and searching for new available channel.

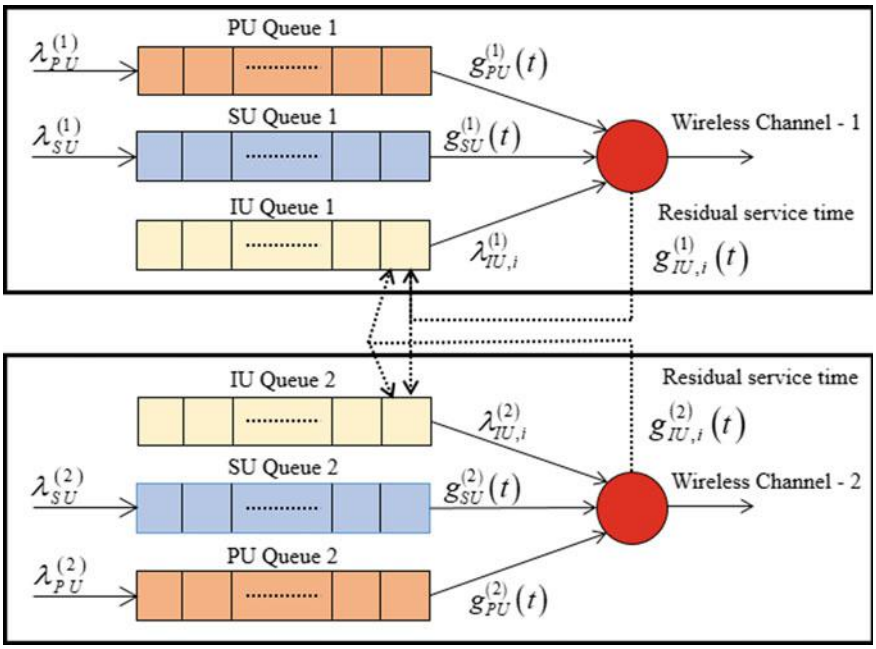



**Fig. 13.4** PRP M/G/1/K queuing network model for PUs and SUs with two wireless channel

If all the channels are busy in the system, the IU will join to the head of interrupted queue of the same channel and this process is called non-switching SH process. Otherwise, it will join the end of the interrupted queue of another channel and this process is called switching SH process.

We consider that the appearance of PUs and SUs in the system obey follow the Poisson random process with the mean arrival rates $\lambda_{\mathrm{PU}}^{(m)}$ and $\lambda_{\mathrm{SU}}^{(m)}$ of PUs and SUs in $m$th default channel (in Fig. 13.4, $m = 1$ or 2), respectively. The terms $g_{\mathrm{PU}}^{(m)}(t)$ and $g_{\mathrm{SU}}^{(m)}(t)$ denote the probability density functions (PDFs) of PUs' and SUs' service period, respectively, with mean $E\left[X_{\mathrm{PU}}^{(m)}\right]$ and $E\left[X_{\mathrm{SU}}^{(m)}\right]$; cumulative distribution function (CDF) $G_{\mathrm{PU}}(t)$ and $G_{\mathrm{SU}}(t)$, respectively. As the switching SH process also taken in consideration, therefore, say an SU with i interruption arrives with arrival rate $\lambda_{\mathrm{IU},i}^{(m)}$; mean effective service time $E\left[X_{\mathrm{IU}}^{(m)}(i)\right]$ and effective service period distribution $g_{\mathrm{IU},i}^{(m)}(t)$ at a particular channel m. We consider that the service period of both the PUs and SUs obey the exponential distributions with mean $\mu_{\mathrm{PU}} = 1/E[X_{\mathrm{PU}}]$ and $\mu_{\mathrm{su}} = 1/E[X_{\mathrm{SU}}]$, respectively. For simplicity, all channels in the system are considered as identical for modelling the delay performance measuring metric of SUs.

The service of an SU may be disrupted multiple times by higher priority PUs throughout its service and which will arise delay in the complete service period. This

delay is known as spectrum handoff delay which has negative impact on the performance of the system. Let $E[N]$ and $E[D]$ denote the average number of interruption and handoff delay (HD) experienced by an SU throughout its total service period. The $E[N]$ can be calculated as

$$E[N] = \lambda_{\mathrm{PU,eff}} E[X_{\mathrm{SU}}] \tag{13.1}$$

where $\lambda_{\mathrm{PU,eff}}$ is PUs' effective mean arrival rate and evaluated according to Bose (2002) as,

$$\lambda_{\mathrm{PU,eff}} = \sum_{r=0}^{K-1} \lambda_r p_r = \lambda_{\mathrm{PU}} \sum_{r=0}^{K-1} p_i = \lambda_{\mathrm{PU}} \left(1 - P_{\mathrm{B_{PU},}K}\right) \tag{13.2}$$

In (13.2), $p_i$ denotes the probability of $i$ number of PUs present in the system and $P_{\mathrm{B_{PU},}K}$ denotes the blocking probability of PUs in the queue with finite buffer length $K$. The $P_{\mathrm{B_{PU},}K}$ can be calculated as,

$$P_{\mathrm{B_{PU},}K} = \frac{(1 - \rho_{\mathrm{PU}})}{\left(1 - \rho_{\mathrm{PU}}^{K+1}\right)} \rho_{\mathrm{PU}}^{K} \tag{13.3}$$

where $\rho_{\mathrm{PU}}$ denotes the traffic offered by PUs which is obtained as,

$$\rho_{\mathrm{PU}} = \lambda_{\mathrm{PU}} E[X_{\mathrm{PU}}] = \frac{\lambda_{\mathrm{PU}}}{\mu_{\mathrm{PU}}} \tag{13.4}$$

Now, the CHD ($E[D_{\mathrm{cum}}]$) can be obtained by using the formula

$$E[D_{\mathrm{cum}}] = E[N]E[D] \tag{13.5}$$

## 13.3 CHD for Non-switching, Switching and Random Switching Proactive SH Schemes

In this section, we model the CHDs for various proactive SH schemes such as non-switching, switching and random SH schemes. For the non-switching SH scheme, the user will always join the IU queue of the same channel upon interruption, and hence, it is called *always stay* case (Zahed et al. 2013). However, for switching SH scheme, the IU will *always change* its channel and will join the IU queue of other channels (Zahed et al. 2013). In random SH schemes, an interrupted user may wait in the current channel or can shift to another channel.

### 13.3.1 Non-switching SH

Figure 13.5 illustrates the concept of non-switching proactive SH scheme. Here, an SU can begin its transmission by accessing the channel when there is no PU present in that channel as well as in PU queue (Wu et al. 2013, 2017). However, if a PU appears in that channel, the SU service will be disrupted and it vacates that channel to PU and join the IU queue of the same channel. The interrupted SU will stay in the IU queue up to the completion of the PU service. If the PU service ends and no other PUs present in the PU queue, the interrupted SU can access the channel to complete its remaining transmission. Therefore, the handoff delay of the SU will be the service period of the PU in that channel as shown in Fig. 13.5.

Therefore, the CHD of a SU due to the PUs service for the $K$ finite buffer size queue in a non-switching SH scheme is modelled as,

$$\begin{aligned} E[D_{\text{cum}}] &= E[N]E[D] = E[N]\text{BP}_K \\ &= \lambda_{\text{PU,eff}} E[X_{\text{SU}}] \left( E[X_{\text{PU}}] + \sum_{i=1}^{K-1} \text{BP}_i \alpha_{K-i} \right) \Big/ \beta_0 \end{aligned} \tag{13.6}$$

where $E[D] = \text{BP}_K$ represents the average SH delay of the SU due to the busy period of PUs service in the same channel (Miller 1975) and

$$\alpha_i = 1 - \sum_{j=0}^{i} \beta_j \tag{13.7}$$

When the service time of PUs follows the exponential distribution, $\beta_i$ can be calculated as (Hoque et al. 2019)

$$\beta_i = \frac{\lambda_{\text{PU}}^i \mu_{\text{PU}}}{(\lambda_{\text{PU}} + \mu_{\text{PU}})^{i+1}} = \frac{\mu_{\text{PU}}}{\lambda_{\text{PU}}} \left( \frac{\lambda_{\text{PU}}}{\mu_{\text{PU}} + \lambda_{\text{PU}}} \right)^{i+1} = \frac{\rho_{\text{PU}}^i}{(1 + \rho_{\text{PU}})^{i+1}} \tag{13.8}$$

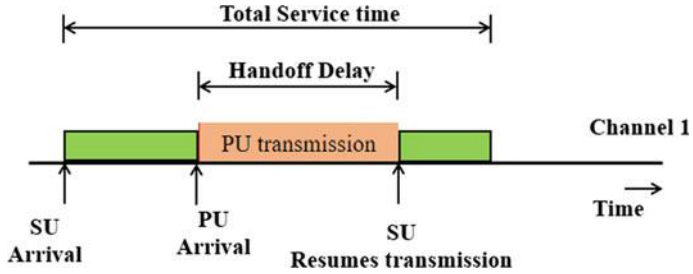



**Fig. 13.5** Concept of non-switching SH process

By calculating the values of $\beta_i$ and $\alpha_i$, we can obtain $E[D]$ and also CHD ($E[D_{\text{cum}}]$) of an SU in non-switching proactive SH scheme, respectively.

### 13.3.2 Switching Spectrum Handoff

The concept of switching SH process is shown in Fig. 13.6. In this scheme, whenever an SU service is interrupted by higher priority PU, it always switches its service to another new channel. If the channel is busy, the SU joins the end of IU queue of that new channel and waits for the service. Therefore, this spectrum handoff is also known as *always change scheme* (Zahed et al. 2013).

From Fig. 13.4, it is observed that an SU requires to wait in the queue due to: the residual service times of PUs and IUs, the PUs and IUs present in the PU and IU queues, respectively, and the newly appeared PUs during the process. Therefore, the waiting time ($W_{\text{su}}$) for an SU before beginning its service can be obtained as (Shekhar et al. 2019)

$$W_{\text{SU}} = R_{\text{SU}} + \sum_{i=1}^{\infty} Q_{\text{IU}_i} E\left[X_{\text{IU}_i}\right] + Q_{\text{PU}} E[X_{\text{PU}}] + \lambda_{\text{PU,eff}} W_{\text{SU}} E[X_{\text{PU}}] \tag{13.9}$$

where $R_{\text{SU}}$ denotes the delay results from the residual service time of a PU or SU at that instant and can be evaluated as (Bose 2002)

$$R_{\text{SU}} = \frac{1}{2}\lambda_{\text{PU,eff}} E\left[(X_{\text{PU}})^2\right] + \frac{1}{2}\sum_{i=0}^{\infty} \lambda_{\text{IU}_i,\text{eff}} E\left[\left(X_{\text{IU}_i}\right)^2\right] \tag{13.10}$$

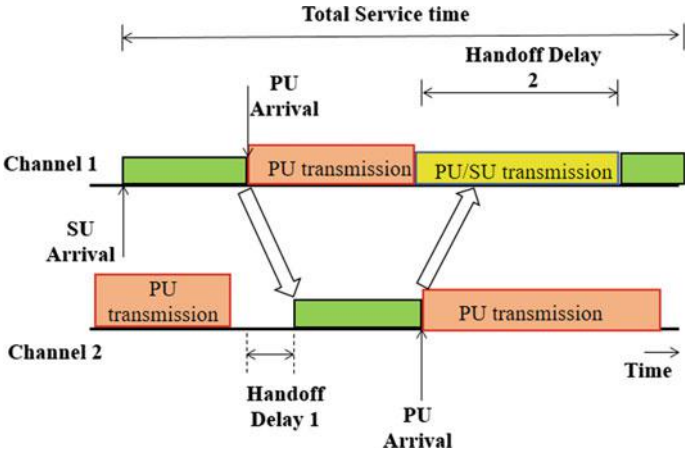



**Fig. 13.6** Concept of switching SH process with two channels

The terms $\lambda_{\mathrm{IU}_i}$, $E\left[\left(X_{\mathrm{IU}_i}\right)^2\right]$ and $E\left[X_{\mathrm{IU}_i}\right]$ in (13.10) can be given as (Wang and Wang 2009; Wang et al. 2012),

$$\lambda_{\mathrm{IU}_i} = \lambda_{\mathrm{SU}}\left(\frac{\lambda_{\mathrm{PU}}}{\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}}\right)^i;\ E\left[\left(X_{\mathrm{IU}_i}\right)^2\right] = \frac{2}{\left(\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}\right)^2} \text{ and}$$
$$E\left[\left(X_{\mathrm{IU}_i}\right)\right] = \frac{1}{\left(\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}\right)} \tag{13.11}$$

Similar to $\lambda_{PU,eff}$, we can also express the term $\lambda_{\mathrm{IU,eff}}$ in (13.10) as,

$$\begin{aligned}\lambda_{\mathrm{IU}_i,\mathrm{eff}} &= \lambda_{\mathrm{IU}_i}\left(1-P_{\mathrm{B}_{\mathrm{IU}},K}\right)\\ &= \lambda_{\mathrm{SU}}\left(\frac{\lambda_{\mathrm{PU}}}{\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}}\right)^i\left(1-\frac{\left(1-\left(\frac{\rho_{\mathrm{SU}}\lambda_{\mathrm{PU}}}{\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}}\right)\right)}{\left(1-\left(\frac{\rho_{\mathrm{SU}}\lambda_{\mathrm{PU}}}{\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}}\right)^{K+1}\right)}\left(\frac{\rho_{\mathrm{SU}}\lambda_{\mathrm{PU}}}{\lambda_{\mathrm{PU}}+\mu_{\mathrm{su}}}\right)^K\right)\end{aligned} \tag{13.12}$$

The second and third terms in (13.9) denote the delay arises because of the interrupted users present in the IU queue and PUs present in PU queue, respectively. According to Little's law, we can calculate the average number of IUs ($Q_{\mathrm{IU}}$) and PUs ($Q_{\mathrm{PU}}$) as given in (13.13) and (13.14), respectively.

$$Q_{\mathrm{IU}_i} = \lambda_{\mathrm{IU}_i,\mathrm{eff}} W_{\mathrm{CU}} \tag{13.13}$$

$$Q_{\mathrm{PU}} = \lambda_{\mathrm{PU,eff}} W_{\mathrm{PU}} \tag{13.14}$$

In (13.14), the term $W_{\mathrm{PU}}$ represents the waiting time for PU in higher priority PU queue before starting its service and which can be expressed as (Bose 2002):

$$\begin{aligned}W_{\mathrm{PU}} &= R_{\mathrm{PU}} + \lambda_{\mathrm{PU}}\left(1-P_{\mathrm{B}_{\mathrm{PU}},K}\right)W_{\mathrm{PU}}E[X_{\mathrm{PU}}]\\ &= \frac{R_{\mathrm{PU}}}{\left(1-\lambda_{\mathrm{PU}}\left(1-P_{\mathrm{B}_{\mathrm{PU}},K}\right)E[X_{\mathrm{PU}}]\right)}\end{aligned} \tag{13.15}$$

where $R_{PU}$ represents the mean residual service period for a PU and can be evaluated as (Wang and Wang 2009),

$$R_{\mathrm{PU}} = \frac{1}{2}\lambda_{\mathrm{PU}}\left(1-P_{\mathrm{B}_{\mathrm{PU}},K}\right)E\left[\left(X_{\mathrm{PU}}\right)^2\right] \tag{13.16}$$

The final term in (13.9) represents the delay due to the newly arrived PU in the system during the ongoing process of the CU and can be expressed as:

$$\lambda_{\mathrm{PU,eff}} W_{\mathrm{CU}} E[X_{\mathrm{PU}}] = \rho_{\mathrm{PU}}\left(1-P_{\mathrm{B}_{\mathrm{PU}},K}\right)W_{\mathrm{CU}} \tag{13.17}$$

By replacing all the terms from (13.10) to (13.17) into (13.9), the standard form of $W_{SU}$ can be obtained as

$$W_{SU} = \frac{\frac{1}{2}\lambda_{PU}\left(1-P_{B_{PU},K}\right)E\left[(X_{PU})^2\right] + \frac{\rho_{SU}}{(\lambda_{PU}+\mu_{su})}\left(1-P_{B_{PU},K}\right) + \frac{1}{2}\frac{\lambda_{PU}^2\left(1-P_{B_{PU},K}\right)^2 E\left[(X_{PU})^2\right]}{\left(1-\rho_{PU}\left(1-P_{B_{PU},K}\right)\right)}E[X_{PU}]}{\left(1-\rho_{SU}\left(\frac{\lambda_{PU}}{\lambda_{PU}+\mu_{su}}\right)\left(1-P_{B_{IU},K}\right)-\rho_{PU}\left(1-P_{B_{PU},K}\right)\right)} \tag{13.18}$$

For the switching SH schemes, the average handoff delay $E[D]$ can be obtained as

$$E[D] = W_{SU} + T_{sw} \tag{13.19}$$

Now, the cumulative handoff delay $E[D_{cum}]$ for switching SH schemes can be obtained as,

$$E[D_{cum}] = \lambda_{PU,eff} E[X_{SU}]\left(\frac{\frac{1}{2}\lambda_{PU}\left(1-P_{B_{PU},K}\right)E\left[(X_{PU})^2\right] + \frac{\rho_{SU}}{(\lambda_{PU}+\mu_{su})}\left(1-P_{B_{PU},K}\right) + \frac{1}{2}\frac{\lambda_{PU}^2\left(1-P_{B_{PU},K}\right)^2 E\left[(X_{PU})^2\right]}{\left(1-\rho_{PU}\left(1-P_{B_{PU},K}\right)\right)}E[X_{PU}]}{\left(1-\rho_{SU}\left(\frac{\lambda_{PU}}{\lambda_{PU}+\mu_{su}}\right)\left(1-P_{B_{IU},K}\right)-\rho_{PU}\left(1-P_{B_{PU},K}\right)\right)} + T_{sw}\right) \tag{13.20}$$

### 13.3.3 Random Spectrum Handoff

This spectrum handoff scheme is consist of both switching and non-switching SH schemes. Therefore, an SU performs either switching or non-switching SH to continue its remaining service. In the presented model, we consider that there will be 50% chance that the SU stays on its own channel and 50% chance that it changes its channel to completes its transmission. The CHD of random SH scheme can be obtained by considering both non-switching and switching SH schemes as

$$\begin{aligned} E[D_{cum}] &= \frac{E[N]}{2}E[D] + \frac{E[N]}{2}(W_{SU} + T_{sw}) \\ &= \frac{E[N]}{2}(BP_K + W_{SU} + T_{sw}) \end{aligned} \tag{13.21}$$

## 13.4 Results and Discussion

The blocking probability and handoff delay are vital parameters in the performance evaluation of a queuing network. Here, the results of blocking probability and CHD of an SU for various proactive SH schemes are presented under the impact of K using the PRP M/G/1/K queuing model in CRNs. In order to analyse the handoff performance metrics, the following parameters specification are considered: $\lambda_{PU} = 0.15$, $\mu_{PU} = 0.60$, $\mu_{su} = 0.40$, $p = 0.50$ and $T_{sw} = 0$.

The blocking probability of higher priority PUs is depicted in Fig. 13.7for various length of $K = 1, 2, 3, 5, 7$ with respect to arrival rate of PUs ($\lambda_{PU}$). For the fixed service

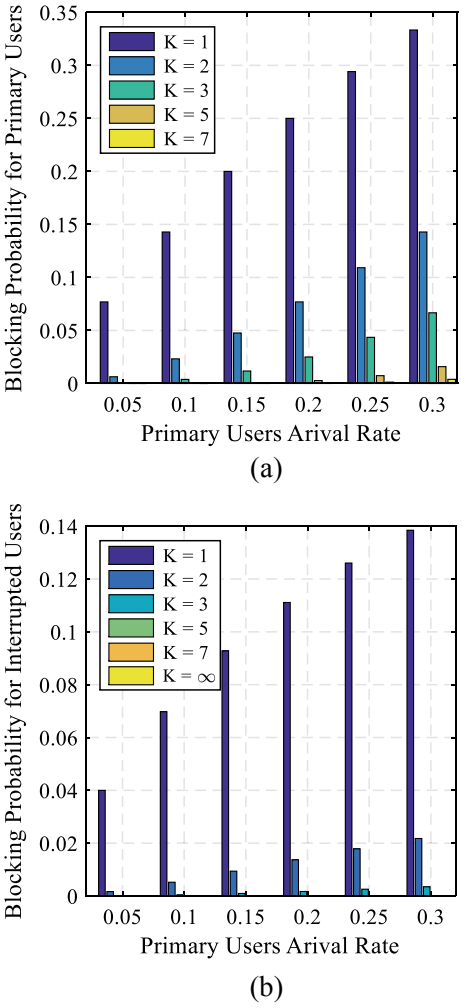



**Fig. 13.7** Blocking probability of **a** PUs and **b** IUs in terms of $\lambda_{PU}$

time of PUs, as the $\lambda_{PU}$ increases, the number of PUs in system also increases. Therefore, the blocking probability increases for the PUs with increasing $\lambda_{PU}$ as depicted in Fig. 13.7a. The increasing value of $K$ in higher priority queue implies additional number of waiting place for PUs in the queue. Hence, the blocking probability of PUs reduces with the higher values of $K$.

Figure 13.7b presents the impact of buffer size on blocking probability of interrupted SU in terms of $\lambda_{PU}$. Increasing values of $\lambda_{PU}$ implies higher number of PUs in the system and which leads to more number of interruptions to SUs. Hence, the number of interrupted SUs also increases with increasing $\lambda_{PU}$ and which leads to increasing in blocking probability. It is also noticed that the blocking probability of interrupted SUs in Fig. 13.7b decreases with increasing length of buffer size and this is because of similar reason discussed in Fig. 13.7a. As the interrupted SUs have the preference over the newly arrived SUs and join the IU queue only, the number of interrupted SUs in the interrupted queue will be lesser as compared to the other users presented in other queues. Hence, the system offers lower blocking probability for interrupted SUs as compared to PUs as shown in Fig. 13.7a, b.

Figures 13.8 and 13.9 depict the results of CHD of a CU for non-switching spectrum handoff scheme with varying arrival rate of the primary user ($\lambda_{PU}$) and mobility factor of spectrum holes ($\lambda_{PU}/\mu_{su}$), respectively. As $\lambda_{PU}$ increases, the chances for the interruption in service of a SU increase, i.e., a high arrival rate of PUs causes higher number of interruption in comparison with low arrival rate of primary users. Each interruption causes a delay to SU service which results in increasing of cumulative handoff delay as shown in Figs. 13.8 and 13.9, respectively. As the length of the queue decreases, the number of users waiting for service also decreases and hence, the CHD is reduced and an SU can complete its service within lesser time.

As the value of $K$ is increasing, the performance of M/G/1/K model in terms of CHD is approaching towards the performance of infinite buffer length M/G/1 queuing model. It is clearly observed that the proposed model with $K > 5$ offers very

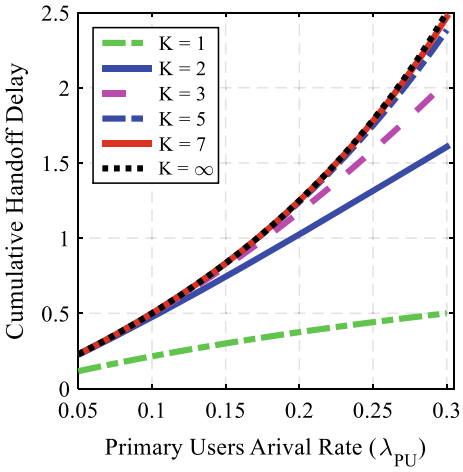



**Fig. 13.8** CHD for non-switching SH with varying $\lambda_{PU}$

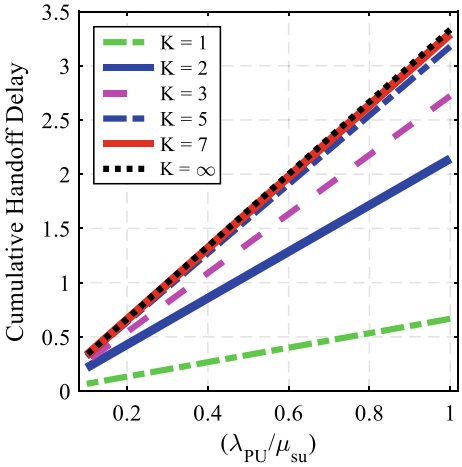



**Fig. 13.9** CHD for non-switching SH with varying $\lambda_{PU}/\mu_{su}$

similar performance to the M/G/1 queuing model, and hence, we can be reduced the infinite buffer size to an optimal buffer size for the better performance of a practical queuing system.

The CHDs of an SU for switching SH scheme with various buffer size are presented in Figs. 13.10 and 13.11 in term of $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$, respectively. For this handoff scheme also, the CHD increases with increasing values of $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$. It is observed that the non-switching SH performs better than the switching SH for higher values of $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$.

In case of non-switching SH, an SU requires to join the head of IU queue upon the detection of PU arrival and stays there until the channel becomes available. But in switching SH, when an SU switches to the other channel, it has to join the tail of

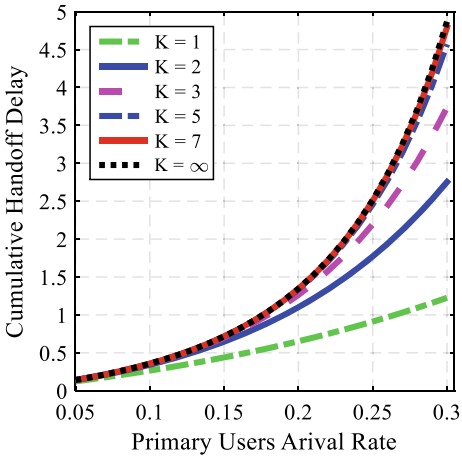



**Fig. 13.10** CHD for switching SH scheme with varying $\lambda_{PU}$

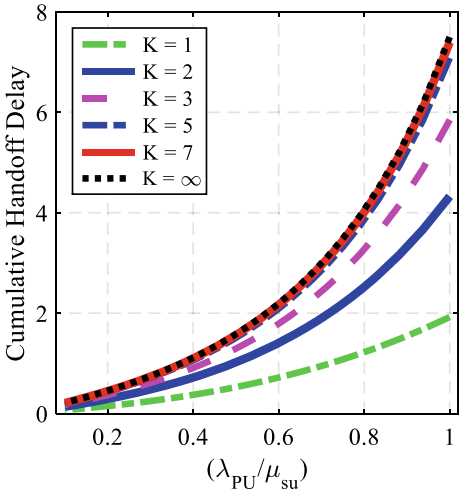



**Fig. 13.11** CHD for switching SH scheme with varying $\lambda_{PU}/\mu_{su}$

interrupted queue. During this waiting time, a newly PU can arrive in that channel which further increases the waiting time for SU to resume its transmission.

Under the impact of buffer size, the CHDs of an SU for random SH scheme with varying $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$ are presented in Figs. 13.12 and 13.13, respectively. The CHD for random spectrum handoff scheme is obtained by averaging same for non-switching and switching SH schemes. As the values of $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$ increase, the CHDs of SUs also increase for different values of K in PRP M/G/1/K queuing network. For lower K, CHD is reduced for our proposed model in comparison with the PRP M/G/1 queuing network.

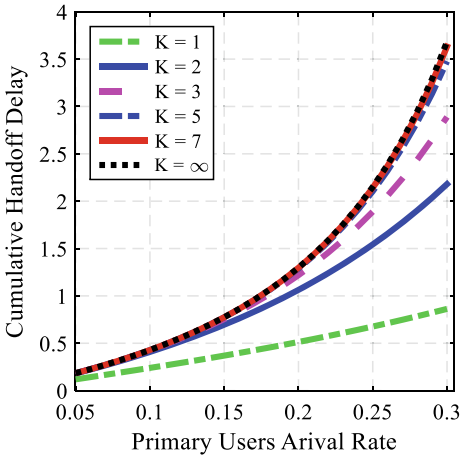



**Fig. 13.12** CHD for random SH scheme with varying $\lambda_{PU}$

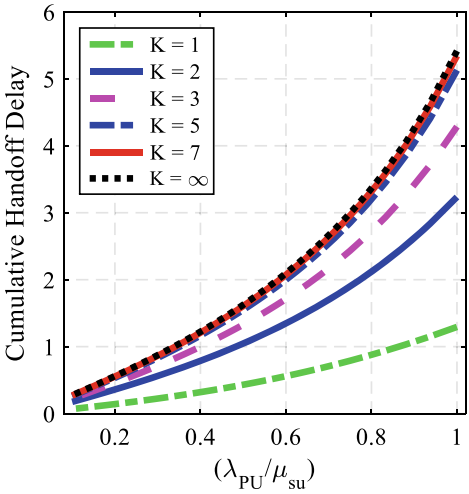



**Fig. 13.13** CHD for random SH scheme with varying $\lambda_{PU}/\mu_{su}$

## 13.5 Conclusion

This chapter presents a PRP M/G/1/K queuing model to explore the effect of buffer size on the handoff performance of SUs in terms of blocking probability and CHD in CRNs. We model CHDs for various proactive SH schemes: non-switching, switching and random SH schemes considering PRP M/G/1/K queuing model and investigates the impact of buffer size on them in terms of arrival rate of the PUs ($\lambda_{PU}$) and mobility parameter ($\lambda_{PU}/\mu_{su}$), respectively. From the results, it is observed that the CHDs for all SH schemes always increase with increasing values of $\lambda_{PU}$ and $\lambda_{PU}/\mu_{su}$. Results also show that there is a significant impact of buffer size ($K$) on blocking probability and CHD of SUs with respect to $\lambda_{PU}$ under various SH schemes. The model with lower buffer size offers lower CHD and as we increase the buffer size, the performance of the PRP M/G/1/K queuing model approaches to the performance of infinite buffer length PRP M/G/1 queuing model. The methods used and results obtained in this chapter will be useful in the implementation of practical queuing model for the better performance of CRNs.

**Acknowledgements** The authors highly acknowledge the Project (File Number: SRG/2019/001744 dated 17-Dec-2019), Science and Engineering Research Board (SERB), Government of India for the resources provided and their never-ending support and motivation for the research work.

## References

Akyildiz IF, Lee W-Y, Vuran MC, Mohanty S (2006) Next generation/dynamic spectrum access/cognitive radio wireless networks—a survey. Comput Netw 50(13):2127–2159

Akyildiz IF, Lee W-Y, Vuran MC, Mohanty S (2008) A survey on spectrum management in cognitive radio networks. IEEE Commun Mag 46(4):40–48

Akyildiz IF, Lee WY, Chowdhury KR (2009) Spectrum management in cognitive radio ad hoc networks. IEEE Netw 23(4):6–12

Bayrakdar ME, Çalhan A (2016) Performance analysis of proactive decision spectrum handoff for MAC protocols in cognitive radio networks. In: 24th signal processing and communication application conference (SIU), pp 481–484

Bose SK (2002) An introduction to queuing systems. Kluwer Academic/Plenum Publishers, New York

Christian J, Moh S, Chung I, Lee J (2012) Spectrum mobility in cognitive radio networks. IEEE Commun Mag 50(6):114–121

Cisco Visual Networking Index (2016) Mobile data traffic forecast update, 2015–2020. CISCO, San Jose

Cisco Visual Networking Index (2019) Mobile data traffic forecast update, 2017–2022. CISCO, San Jose

FCC (2002) Spectrum policy task force report. ET Docket 02-155

Federal Communications Commission (FCC) (2003) Notice for proposed rulemaking (NPRM 03 322): facilitating opportunities for flexible, efficient, and reliable spectrum use employing spectrum agile radio technologies. ET Docket No. 03 108

Gkionis G, Sgora A, Vergados DD, Michalas A (2017) An effective spectrum handoff scheme for cognitive radio ad hoc networks. In: Wireless telecommunications symposium (WTS), pp 1–7

Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23:201–220

Hoque S, Shekhar S, Sen D, Arif W (2019) Analysis of handoff delay for proactive spectrum handoff scheme M with PRP/G/1/K queuing system in cognitive radio networks. IET Commun 13(6):706–711

Kumar K, Prakash A, Tripathi R (2016) Spectrum handoff in cognitive radio networks: a classification and comprehensive survey. J Netw Comput Appl 61:161–188

Lee DJ, Yeo WY (2015) Channel availability analysis of spectrum handoff in cognitive radio networks. IEEE Commun Lett 19(3):435–438

Mathonsi TE, Kogeda OP (2016) Handoff delay reduction model for heterogeneous wireless networks. In: IST-Africa week conference, pp 1–7

McHenry M (2003) Spectrum white space measurements. New America Foundation Broadband Forum

Miller LW (1975) Technical note—a note on the busy period of an M/G/1 finite queue. Oper Res 23(6):1179–1182

Mitola J (2001) Cognitive radio for flexible mobile multimedia communications. Mob Netw Appl 6(5):435–441

Mitola J, Maguire GQ (1999) Cognitive radio: making software radios more personal. IEEE Pers Commun Mag 6(4):13–18

Shared Spectrum Company. [Online]. http://www.sharedspectrum.com/

Shekhar S, Hoque S, Arif W (2019) Analysis of spectrum handoff delay using finite queuing model in cognitive radio networks. Int J Commun Netw Distrib Syst Indersci 25(1). https://doi.org/10.1504/ijcnds.2020.10023989

Wang LC, Wang CW (2008) Spectrum handoff for cognitive radio networks: reactive-sensing or proactive-sensing. In: IEEE international performance, computing and communications conference, pp 343–348

Wang CW, Wang LC (2009) Modeling and analysis for proactive-decision spectrum handoff in cognitive radio networks. In: IEEE international conference on communications, pp 1–6

Wang LC, Wang CW, Chang CJ (2012) Modeling and analysis for spectrum handoffs in cognitive radio networks. IEEE Trans Mob Comput 11(9):1499–1513
Wang C-X et al (2014) Cellular architecture and key technologies for 5G wireless communication networks. IEEE Commun Mag 52(2):122–130
Wu Y, Hu F, Kumar S, Guo M, Bao K (2013) Spectrum handoffs with mixed priority queuing model over cognitive radio networks. In: IEEE global conference on signal and information processing, pp 1194–1197
Wu Y, Hu F, Zhu Y, Kumar S (2017) Optimal spectrum handoff control for CRN based on hybrid priority queuing and multi-teacher apprentice learning. IEEE Trans Veh Technol 66(3):2630–2642
Zahed S, Awan I, Cullen A (2013) Analytical modeling for spectrum handoff decision in cognitive radio networks. Simul Model Pract Theory 38:98–114
Zakariya AY, Rabia SI (2016) Analysis of an interruption-based priority for multi-class secondary users in cognitive radio networks. In: IEEE international conference on communications (ICC), pp 1–6
Zheng MA, Zhengquan Z, Zhiguo D, Pingzhi F, Hengchao L (2015) Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives. Sci China Inform Sci 58(4):1–20

# Chapter 14
# Overlay Multi-user Satellite-Terrestrial Networks for IoT in the Presence of Interference


**Pankaj K. Sharma, Budharam Yogesh, and Deepika Gupta**

**Abstract** The fifth-generation (5G) wireless systems have recently reached to the deployment stage, and research toward next-generation wireless communications has already started. The integration of Internet of things (IoT) to satellite networks is one of the key focuses in 5G and beyond systems. Since the spectrum resources are not abundant to accommodate billions of IoT devices, the cognitive satellite-terrestrial networks supporting IoT communications are of great importance. In this chapter, we present the outage performance analysis of a cognitive overlay multi-user satellite-terrestrial network (OMSTN) where a primary satellite communicates with a selected terrestrial receiver via a secondary IoT network in the presence of interference from extra-terrestrial sources (ETSs) and terrestrial sources (TSs). Hereby, we consider the following two cases for system performance analysis: (a) *Case* 1: when interference occurs due to TSs only; (b) *Case* 2: when interference occurs due to both ETSs and TSs. We specifically derive the tight closed-form outage probability (OP) expressions of both the primary satellite and secondary IoT networks. We further carry out the asymptotic high signal-to-noise ratio (SNR) analysis to obtain the diversity order of both satellite and IoT networks. For asymptotic analysis, we consider two scenarios, namely when interferers' power is fixed and when it varies proportionally to the transmit power of satellite and IoT nodes. We further devise a scheme for the adaptive power splitting factor under a guaranteed quality-of-service (QoS) of the primary satellite network. We provide various insights on the considered OMSTN based on our analysis and numerical results.

P. K. Sharma (✉) · B. Yogesh
Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha 769008, India
e-mail: sharmap@nitrkl.ac.in

B. Yogesh
e-mail: 218ec5123@nitrkl.ac.in

D. Gupta
Department of Electronics and Communication Engineering, Dr S P M International Institute of Information Technology, Naya Raipur, Chhattisgarh 493661, India
e-mail: deepika@iiitnr.edu.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_14

**Keywords** Satellite-terrestrial networks · Internet-of-things (IoT) · Spectrum sharing · Performance analysis · Outage probability

## 14.1 Introduction

Satellite communications have attracted a great deal of recent research interests for wide range of applications viz., broadcasting, remote sensing, navigation, disaster management, etc., (Vanelli-Corali et al. 2007). The low earth orbit (LEO) and geostationary earth orbit (GEO) satellites have been integrated to mobile terrestrial networks to extend the wireless coverage. Further, the medium earth orbit (MEO) satellites are the popular choice for certain applications. The typical classification of satellites based on the orbital data is given in Table 14.1 (Riebeek and Simmon 2009). In addition, the frequency bands in use for various satellite applications according to the data provided by European Space Agency (ESA) are listed in Table 14.2 (Satellite Frequency Bands 2020).

**Table 14.1** Satellite orbits based on altitudes (Riebeek and Simmon 2009)

| Orbit | Altitude (km) | Orbital period | Satellites |
|---|---|---|---|
| Low Earth orbit (LEO) | <2000 | 84–127 min | Hubble space telescope, RISAT-2B, etc. |
| Medium Earth orbit (MEO) | 2000–35,780 | 2–12 h | GLONASS, Galileo, etc. |
| Geostationary orbit (GEO) | 35,780 | 23 h and 56 min | GSAT-11, South Asia satellite, etc. |
| High Earth orbit (HEO) | >35,780 | >24 h | Vela 1A, IBEX |

**Table 14.2** Satellite frequencies and their applications (Satellite Frequency Bands 2020)

| Bands | Frequencies (GHz) | Applications |
|---|---|---|
| L-band | 1–2 | Global positioning system (GPS) |
| S-band | 2–4 | Communication satellites for space shuttles, weather radar, etc. |
| C-band | 4–8 | Satellite TV networks |
| X-band | 8–12 | Military radar applications |
| Ku-band | 12–18 | Broadcast satellite services |
| Ka-band | 26–40 | Close-range targeting radars on military aircraft |

Unlike the terrestrial networks, the satellite networks suffer from severe blockage effects due to variety of reasons, e.g., heavy clouds, rain attenuation, terrestrial user is indoor, etc. Consequently, the direct link from satellite to ground user is often masked (Jo 2011). To combat the masking effect, the terrestrial cooperative relaying has been integrated to satellite networks for reliable communications (Evans et al. 2005; Paillassa et al. 2011). As a consequence, the dual-hop satellite-terrestrial networks (STNs) with cooperative relaying infrastructure have been evolved. The major classification of STNs based on spectrum utilization can be given as two types, viz., hybrid STNs and integrated STNs (Sreng et al. 2013; Kim 2011). In the former STN type, orthogonal spectrum resources are utilized for transmission over the two hops. However, in the latter STN type, same spectrum resources are utilized for transmission in both hops. Note that as compared to hybrid STNs, the integrated STNs may subject to co-channel interference not only from the terrestrial sources (TSs), but also from extra-terrestrial sources (ETSs). Hereby, the common interfering ETSs are the co-channel satellites and the modern high altitude platforms (HAPs) (e.g., unmanned aerial vehicles, balloons, etc.). Further, with respect to STNs, it is customary to have satellite links affected by the shadowed-Rician fading and the terrestrial links affected by the Rayleigh fading.

In fifth-generation (5G) networks, a terrestrial ecosystem of billions of connected devices (e.g., home appliances, smart vehicles, industry equipment, etc.), referred to as Internet of things (IoT), is expected to increase tremendously the demand of spectrum resources (Chen et al. 2017; Khan et al. 2017; Salman et al. 2018). The large number of low-cost IoT devices in future must be operated in unlicensed spectrum to meet their spectrum requirements (Lysogor et al. 2018). It is predicted that the end-user spending on IoT applications will be 6.3 times higher in 2025 than in 2020. To this end, cognitive radio is viable solution to cater the future spectrum demands of IoT devices. Cognitive radio allows the coexistence of unlicensed secondary devices in the licensed spectrum of a primary network based on spectrum sharing as long as the quality-of-service (QoS) of primary network is protected. The underlay and overlay are the two most popular models of cognitive radio (Manna et al. 2011; Zou et al. 2010). In underlay model, the transmit power of secondary devices is safely constrained to limit the harmful interference at primary network. In contrast, in overlay model, the secondary devices communicate with its own receiver by cooperatively assisting primary communications based on a less restrictive network coding techniques. In this chapter, we discuss the integration of IoT devices to satellites based on overlay approach. Some practical systems have been devised to integrate IoT devices to satellite such as receiver autonomous integrity monitoring (RAIM), satellite-based augmentation system (SBAS), multi-global navigation satellite system (Multi-GNSS), advanced-RAIM (ARAIM), etc., [please refer to (Chen et al. 2017)]. More recently, the integration of IoT devices to satellite networks for millimeter (mm) wave communications in 5G systems has been considered based on cognitive networking (Evans et al. 2015). It is worth mentioning that the integration of IoT to low earth orbit (LEO)/high earth orbit (HEO) satellite systems has already been envisioned in existing standards, e.g., digital video broadcast-satellite-second generation extension (DVBS2X) (ETSI 2014). Vodafone has recently initiated a

project for extending satellite communications to terrestrial IoT devices (Vodafone Group 2018). A similar project (known as Starlink) has been initiated by SpaceX for terrestrial internet applications (Starlink Network 2020). Cognitive STNs have recently gained significant research interests (Jia et al. 2016; Kandeepan et al. 2010; Sharma et al. 2013). In particular, for IoT applications, an overlay STN is of great importance where the primary satellite network can share the spectrum (e.g., direct-to-home television bands, etc.) with secondary IoT devices. Herein, an IoT device cooperatively assists satellite communications based on network-coded primary and secondary signal transmissions. More importantly, the densification of STNs aiming futuristic sixth-generation (6G) services leads to an unavoidable interference problem.

In light of the above, in this chapter, we consider an overlay multi-user STN (OMSTN) where a primary satellite communicates with a selected terrestrial receiver with the help of a secondary IoT network in the presence of interference. Hereby, to incorporate the hybrid as well as integrated STN scenarios, we consider the following two cases for system performance analysis: (a) *Case* 1: when interference occurs due to TSs only; (b) *Case* 2: when interference occurs due to both ETSs and TSs. We specifically derive the tight closed-form outage probability (OP) expressions of both the primary satellite and secondary IoT networks. We further carry out the asymptotic high signal-to-noise ratio (SNR) analysis to obtain the diversity order of respective networks. For high SNR analysis, we consider two scenarios, namely when interferers' power is fixed and when it varies proportional to the transmit power of satellite and IoT nodes. We further devise a scheme for adaptive power splitting factor under a guaranteed QoS of primary satellite network. We disclose various insights on the considered OMSTN based on our analysis in numerical results section.

### *14.1.1 Prior Works*

In general, the performance of both decode-and-forward (DF) and amplify-and-forward (AF)-based STNs has been analyzed in Sreng et al. (2013), Bhatnagar and Arti (2013), An et al. (2015), Upadhyay and Sharma (2016). The authors in Miridakis et al. (2015) have considered the dual-hop multi-antenna STNs. Most recently, STNs with mobile unmanned aerial vehicle (UAV) relaying has been investigated in Sharma et al. (2020). The secrecy performance of STNs has been analyzed in An et al. (2016). In Sharma and Kim (2020), the authors have considered the secrecy performance analysis of mobile UAV relaying for STNs. Specifically, the performance of cognitive STNs have been analyzed in Guo et al. (2018), Sharma et al. (2017), Vassaki et al. (2013). We highlight that the aforementioned works have neglected the impact of interference on the performance of STNs. Several works in literature have taken into account the interference from TSs for the performance analysis of STNs. In particular, the works in An et al. (2014, 2015) have analyzed the performance of STNs with interference. Further, the performance of AF-based STNs for generalized integer and non-integer SR fading parameters has been analyzed in Yang and Hasna

(2015) with interference. In addition, the performance of multi-user STNs multi-user scheduling was assessed in Upadhyay and Sharma (2016) under interference. The outage performance of integrated STNs with interference has been analyzed in Ruan et al. (2017). Furthermore, the impact of hardware impairments on the performance of STNs has been considered in Guo et al. (2019) with interference. Moreover, the outage performance of underlay cognitive STNs under interference has been analyzed in An et al. (2016). Hereby, we emphasize that the majority of aforementioned works have considered the performance analysis of STNs with interfering TSs only by neglecting completely the interference from ETSs. It is quite intuitive that the interference from ETSs in addition to the TSs in next-generation STNs is inevitable. Thus, this chapter deals with a comprehensive performance analysis of OMSTNs under *Case 1* and *Case 2* as described previously.

### *14.1.2 Chapter Organization*

The rest of the chapter is composed of the following sections. In Sect. 14.2, the considered OMSTN system model is discussed. In Sects. 14.3 and 14.4, we assess the outage performance of satellite network and IoT network, respectively. In Sect. 14.5, we formulate a scheme for power splitting factor at IoT transmitter to guarantee the QoS of primary satellite network. In Sect. 14.6, we provide the numerical and simulation results. Finally, in Sect. 14.7, we discuss the summary of this chapter.

## 14.2 System Description

### *14.2.1 System Model*

We consider an OMSTN system model as shown in Fig. 14.1. Herein, a satellite transmitter $A$ intends to communicate with a selected terrestrial user $B_n$ from $N$ terrestrial users $\{B_n\}_{n=1}^{N}$. We assume that the direct links from $A$ to $B_n$ are unavailable due to either severe shadowing or due to indoor location of $B_n$. Hence, for reliable communications from $A$ to $B_n$, the primary satellite network allows spectrum sharing to a secondary IoT network comprises of a transmitter $C$ and receiver $D$. The IoT transmitter is assumed to be capable of applying superposition coding to multiplex the primary satellite and secondary IoT signal in power domain based on overlay cognitive radio principle. Consequently, the IoT network access the primary satellite spectrum for its own secondary communications alongside cooperatively assisting the primary satellite communications. Further, we assume that the terrestrial satellite receivers $\{B_n\}_{n=1}^{N}$ as well as the IoT network $C - D$ are operating under the mixed interference received from the extra-terrestrial sources (ETSs) and terrestrial sources (TSs). While the ETSs include various satellite transmitters or high altitude

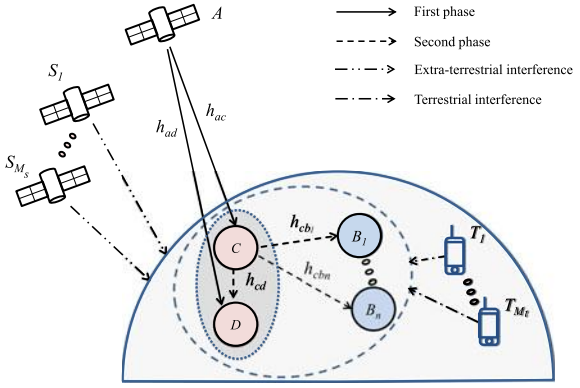



**Fig. 14.1** OMSTN system model

platforms (HAPs) (e.g., balloons, etc.) sharing the common frequency spectrum as of considered primary satellite network, the TSs include the cellular users. In this chapter, we analyze the outage performance of considered OMSTN system model under the following two cases: *Case* 1: when interference occurs due to TSs only and *Case* 2: when interference occurs due to both ETSs and TSs. We consider $M_t$ as the number of interfering TSs $\{T_l\}_{l=1}^{M_t}$ and $M_s$ as the number of interfering ETSs $\{S_j\}_{j=1}^{M_s}$. The channel coefficients corresponding to $A-C$, $A-D$, $C-D$, and $C-B_n$ links are $h_{ac}$, $h_{ad}$, $h_{cd}$, and $h_{cb_n}$, respectively.

### *14.2.2 Propagation Model*

The satellite $A$ communicates with the selected terrestrial receiver $B_n$ in two consecutive time phases via the IoT transmitter $C$ which first superposes an amplified version of primary satellite signal received in first phase and its own secondary signal (based on power domain multiplexing technique) and then, broadcasts the combined signal in the second phase. Hereby, the amplification of primary satellite signal is achieved using conventional AF principle. Specifically, in the first phase, satellite $A$ transmits its signal $x_a$ with power $P_a$ which is received by the IoT transmitter $C$ as well as by the IoT receiver $D$. In addition, the interference $I$ corresponding to *Cases* 1 and 2 is received by the IoT network. The received signals at $C$ and $D$ can be represented in a compact form as

$$y_{ai} = \sqrt{P_a} h_{ai} x_a + I + n_{ai}, \tag{14.1}$$

where $i \in \{c, d\}$ and $n_{ai}$ is the AWGN with zero mean and variance $\sigma^2$. Note that, in (14.1), the term $I = I_t$ for *Case 1* where $I_t = \sum_{l=1}^{M_t} \sqrt{P_t} h_{tl} x_l$, with $P_t$, $h_{tl}$, and $x_l$ are respectively, the transmit power, channel, and signal of interfering TSs. Further,

the term $I = I_t + I_s$ for *Case 2* where $I_s = \sum_{j=1}^{M_s} \sqrt{P_s} h_{sj} x_j$, with $P_s$, $h_{sj}$, and $x_j$ are respectively, the transmit power, channel, and signal of interfering ETSs.

In the second phase, the IoT transmitter $C$ first combines the amplified primary satellite signal $y_{ac}$ and its own secondary signal $x_c$ by splitting its total available power as $\mu P_c$ and $(1-\mu)P_c$ among these signals, respectively. Consequently, the combined signal can be expressed as

$$z_c = \sqrt{\mu P_c} \frac{y_{ac}}{\sqrt{|y_{ac}|^2}} + \sqrt{(1-\mu)P_c} x_c, \tag{14.2}$$

where $\mu \in (0, 1)$ is defined as a power splitting factor. Then, the IoT transmitter $C$ broadcasts the superposed signal $z_c$ toward $B_n$ and $D$. The signals received at $B_n$ and $D$ can be given in a compact form as

$$y_{c\upsilon} = h_{c\upsilon} z_c + I + n_{c\upsilon}, \tag{14.3}$$

where $\upsilon \in \{b_n, d\}$, $I$ is the same as defined previously, and $n_{c\upsilon}$ is the AWGN. Thus, the signal-to-interference-plus-noise ratio (SINR) at $B_n$ via node $C$ is given by

$$\Lambda_{acb_n} = \frac{\mu \hat{\Lambda}_{ac} \hat{\Lambda}_{cb_n}}{(1-\mu)\hat{\Lambda}_{ac}\hat{\Lambda}_{cb_n} + \hat{\Lambda}_{ac} + \hat{\Lambda}_{cb_n} + 1}, \tag{14.4}$$

where $\hat{\Lambda}_{ac} = \frac{\Lambda_{ac}}{W_c+1}$, $\hat{\Lambda}_{cb_n} = \frac{\Lambda_{cb_n}}{W_c+1}$, $\Lambda_{ac} = \eta_a |h_{ac}|^2$, $\Lambda_{cb_n} = \eta_c |h_{cb_n}|^2$, $\eta_a = \frac{P_a}{\sigma^2}$, and $\eta_c = \frac{P_c}{\sigma^2}$. Note that, in (14.4), the term $W_c = W_t$ for *Case* 1 where $W_t = \sum_{l=1}^{M_t} \Lambda_{tl}$, $\Lambda_{tl} = \eta_t |h_{tl}|^2$, $\eta_t = \frac{P_t}{\sigma^2}$. Whereas, for *Case* 2, $W_c = W_t + W_s$ where $W_s = \sum_{j=1}^{M_s} \Lambda_{sj}$, $\Lambda_{sj} = \eta_s |h_{sj}|^2$, $\eta_s = \frac{P_s}{\sigma^2}$.

Furthermore, we observe that the received signal $y_{cd}$ at $D$ in (14.3) contains primary satellite signal $x_a$ which can be straightforwardly canceled by $D$ using a copy of $x_a$ received by it in the first phase. Therefore, the resultant SINR at IoT receiver $D$ after successful primary interference cancellation can be given as

$$\Lambda_{acd} = \frac{(1-\mu)\hat{\Lambda}_{cd}(\hat{\Lambda}_{ac} + 1)}{\mu \hat{\Lambda}_{cd} + \hat{\Lambda}_{ac} + 1}, \tag{14.5}$$

where $\hat{\Lambda}_{cd} = \frac{\Lambda_{cd}}{W_c+1}$, $\Lambda_{cd} = \eta_c |h_{cd}|^2$, and $W_c$ is the same as defined previously under the *Cases* 1 and 2.

### 14.2.3 Multi-user Selection Criterion for Primary Terrestrial Receivers

To proceed further, we introduce an opportunistic multi-user selection criterion for the primary terrestrial receiver $B_{n^*}$ as given by

$$n^* = \arg \max_{n \in \{1,\ldots,N\}} \Lambda_{acb_n}. \tag{14.6}$$

It is worth emphasizing that the aforementioned multi-user selection criterion provides the best connection to $B_{n^*}$ from IoT transmitter $C$. Further, the aforementioned selection criterion can be implemented in a distributed timer-based mechanism. Here, the IoT transmitter first broadcasts the training symbols which are collected by each receiver $B_n$ to set their individual timers inversely proportional to the received signal strength of training symbols corresponding to $C - B_n$ link. The terrestrial receiver $B_n$ whose timer expires first would be the selected.

### 14.2.4 Channel Models

#### 14.2.4.1 Main Satellite and Extra-Terrestrial Interference Channels

We assume that the channel for main satellite links ($h_{ai}, i \in \{c, d\}$) follows shadowed-Rician (SR) distribution. The probability density function (pdf) of independently and identically distributed (i.i.d.) squared channels $|h_{ai}|^2$ can be given as

$$f_{|h_{ai}|^2}(x) = \alpha_{ai} \mathrm{e}^{-\beta_{ai} x} \, {}_1F_1(m_{ai}; 1; \delta_{ai} x), x \geq 0, \tag{14.7}$$

where $\alpha_{ai} = (2\mathrm{b}_{ai} m_{ai}/(2\mathrm{b}_{ai} m_{ai} + \Omega_{ai}))^{m_{ai}}/2\mathrm{b}_{ai}$, $\beta_{ai} = 1/2\mathrm{b}_{ai}$, $\delta_{ai} = \Omega_{ai}/(2\mathrm{b}_{ai})(2\mathrm{b}_{ai} m_{ai} + \Omega_{ai})$, $\Omega_{ai}$ and $2\mathrm{b}_{ai}$ are, respectively, the average power of the line-of-sight and multipath components, $m_{ai}$ denotes the fading severity, and ${}_1F_1(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function of the first kind (Gradshteyn and Ryzhik 2000, Eq. 9.210.1). The aforementioned pdf in (14.7) for integer $m_{ai}$ can be simplified as (Sharma et al. 2017)

$$f_{|h_{ai}|^2}(x) = \alpha_{ai} \sum_{\kappa=0}^{m_{ai}-1} \zeta_{ai}(\kappa) x^{\kappa} \mathrm{e}^{-(\beta_{ai}-\delta_{ai})x}, \tag{14.8}$$

where $\zeta_{ai}(\kappa) = (-1)^{\kappa}(1 - m_{ai})_{\kappa} \delta_{ai}^{\kappa}/(\kappa!)^2$, with $(\cdot)_{\kappa}$ as the Pochhammer symbol (Gradshteyn and Ryzhik 2000, p. xliii). Further, by the transformation of variates, the pdf of random variable $\Lambda_{ai} = \eta_a |h_{ai}|^2$ can be obtained using (14.8) as

$$f_{\Lambda_{ai}}(x) = \alpha_{ai} \sum_{\kappa=0}^{m_{ai}-1} \frac{\zeta_{ai}(\kappa)}{(\eta_a)^{\kappa+1}} x^{\kappa} \mathrm{e}^{-\frac{\beta_{ai}-\delta_{ai}}{\eta_a} x}. \tag{14.9}$$

Moreover, one can determine the cumulative distribution function (cdf) $F_{\Lambda_{ai}}(x)$ based on integrating (14.9) with the aid of (Gradshteyn and Ryzhik 2000, Eq. 3.351.2) as

$$F_{\Lambda_{ai}}(x) = 1 - \alpha_{ai} \sum_{\kappa=0}^{m_{ai}-1} \frac{\zeta_{ai}(\kappa)}{(\eta_a)^{\kappa+1}} \sum_{p=0}^{\kappa} \frac{\kappa!}{p!} \left( \frac{\beta_{ai} - \delta_{ai}}{\eta_a} \right)^{-(\kappa+1-p)} x^{p} \mathrm{e}^{-\frac{\beta_{ai}-\delta_{ai}}{\eta_a} x}. \tag{14.10}$$

Using (14.8), the pdf of i.i.d. squared SR channels $|h_{sj}|^2$ (i.e., $f_{|h_{sj}|^2}(x)$) pertaining to interfering ETSs can be obtained by replacing the parameters $\{\alpha_{ai}, \beta_{ai}, \flat_{ai}, \delta_{ai}, m_{ai}, \Omega_{ai}, \zeta_{ai}(\kappa)\}$ by $\{\alpha_s, \beta_s, \flat_s, \delta_s, m_s, \Omega_s, \zeta_s(\kappa)\}$, respectively, for $j = 1, ..., M_s$. Similarly, the pdf $f_{\Lambda_{sj}}(x)$ and cdf $F_{\Lambda_{sj}}(x)$ of random variable $\Lambda_{sj} = \eta_s |h_{sj}|^2$ can be obtained using (14.9) and (14.10), respectively, by replacing the parameters $\{\alpha_{ai}, \beta_{ai}, \flat_{ai}, \delta_{ai}, m_{ai}, \Omega_{ai}, \zeta_{ai}(\kappa)\}$ therein by $\{\alpha_s, \beta_s, \flat_s, \delta_s, m_s, \Omega_s, \zeta_s(\kappa)\}$ and $\eta_a$ by $\eta_s$.

#### 14.2.4.2 Main Terrestrial and Terrestrial Interference Channels

The channels for main terrestrial links (i.e., $h_{c\upsilon}$, $\upsilon \in \{b_n, d\}$) and interfering TSs follow Rayleigh fading. Therefore, the respective pdf $f_{\Lambda_{c\upsilon}}(x)$ and cdf $F_{\Lambda_{c\upsilon}}(x)$ of $\Lambda_{c\upsilon} = \eta_c |h_{c\upsilon}|^2$ can be given, based on the exponential distribution, as

$$f_{\Lambda_{c\upsilon}}(x) = \frac{1}{\Omega_{c\upsilon}\eta_c} \mathrm{e}^{-\frac{x}{\Omega_{c\upsilon}\eta_c}} \tag{14.11}$$

and

$$F_{\Lambda_{c\upsilon}}(x) = 1 - \mathrm{e}^{-\frac{x}{\Omega_{c\upsilon}\eta_c}}, \tag{14.12}$$

where $\upsilon \in \{b, d\}$ (here index $n$ is dropped under i.i.d. fading links) and $\Omega_{c\upsilon}$ is the average fading power.

Likewise, the pdf $f_{\Lambda_{c\upsilon}}(x)$ and cdf $F_{\Lambda_{c\upsilon}}(x)$ of i.i.d. random variables $\Lambda_{tl} = \eta_t |h_{tl}|^2$ corresponding to interfering TSs' links can be expressed, respectively, by replacing the parameters $\{\Omega_{c\upsilon}, \eta_c\}$ with $\{\Omega_t, \eta_t\}$ in (14.11) and (14.12), for $l = 1, ..., M_t$.

### 14.2.5 Statistical Characterizations

In this section, we first characterize statistically the random variable $\Lambda_{cb_{n^\star}}$ related to the multi-user selection criterion. By making use of (14.6), (14.12) and applying the order statistics for i.i.d. random variables $\Lambda_{cb_n}$, for $n = 1, ..., N$, the cdf of $\Lambda_{cb_{n^\star}}$ can be determined as

$$F_{\Lambda_{cb_{n^\star}}}(x) = \left(1 - \mathrm{e}^{-\frac{x}{\Omega_{cb}\eta_c}}\right)^n. \tag{14.13}$$

We further characterize statistically the interference variable $W_c = W_s + W_t$ under *Case 2*. Hereby, to proceed, we require the pdf of sum of i.i.d. SR variates (i.e., $W_s$) as well as the sum of i.i.d. Rayleigh variates (i.e., $W_t$). The pdf of $W_s$ is given as (Upadhyay and Sharma 2016)

$$f_{W_s}(w) = \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^\Lambda} w^\Lambda \mathrm{e}^{-\Theta_{ss} w} \tag{14.14}$$

where $\Xi_s(M_s) = \alpha_s^{M_s} \prod_{\kappa=1}^{M_s} \zeta_s(i_\kappa) \prod_{j=1}^{M_s-1} \Phi(\sum_{l=1}^{j} i_l + j, i_{j+1} + 1)$, $\Lambda = \sum_{\kappa=1}^{M_s} i_\kappa + M_s$, $\Theta_{ss} = \frac{\beta_s - \delta_s}{\eta_s}$, $\widetilde{\sum} = \sum_{i_1}^{m_s-1} \ldots \sum_{i_{M_s}}^{m_s-1}$ and $\Phi(\cdot,\cdot)$ denotes the Beta function (Gradshteyn and Ryzhik 2000, Eq. 8.384.1). Next, the pdf of $W_t$ (i.e., sum of i.i.d. and equal power terrestrial Rayleigh interferers) is given as (An et al. 2016)

$$f_{W_t}(w) = \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{w^{M_t-1}}{\Gamma(M_t)} \mathrm{e}^{-\frac{w}{\Omega_t \eta_t}}. \tag{14.15}$$

Now, with pdfs of $W_s$ and $W_t$ off-the-shelf, we can determine the pdf of $W_c$ according to the following lemma.

**Lemma 1** *The pdf of $W_c$ can be given as*

$$f_{W_c}(w) = \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^\Lambda} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \times w^{\Lambda + M_t - 1} \mathrm{e}^{-\frac{w}{\Omega_t \eta_t}} {}_1F_1\left(\Lambda; \Lambda + M_t; -\tilde{\Theta}_{ss} w\right), \tag{14.16}$$

*where* $\tilde{\Theta}_{ss} = \Theta_{ss} - \frac{1}{\Omega_t \eta_t}$.

***Proof*** The pdf of $W_c$ can be written as the statistical convolution of the pdf's of $W_s$ and $W_t$ as per the following expression (Miridakis et al. 2015)

$$f_{W_c}(w) = \int_0^w f_{W_s}(x) f_{W_t}(w-x)\mathrm{d}x. \tag{14.17}$$

On plugging (14.14) and (14.15) in (14.17), it results

$$f_{W_c}(w) = \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \frac{1}{\Gamma(M_t)} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \times \mathrm{e}^{-\frac{w}{\Omega_t \eta_t}} \int_0^w x^{\Lambda-1}(w-x)^{M_t-1} \mathrm{e}^{-\tilde{\Theta}_{ss} x}\mathrm{d}x. \tag{14.18}$$

Finally, evaluating the integral in (14.18), using (Gradshteyn and Ryzhik 2000, Eq. 3.383.1), one can attain (14.16). □

## 14.3 Outage Performance of Satellite Network

In this section, we assess the outage performance of primary satellite network and carry out asymptotic OP analysis to reveal its achievable diversity order.

For a target rate $\mathcal{R}_p$, the OP of primary satellite network with selected primary receiver $n^*$ is determined as

$$\begin{aligned} \mathcal{P}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) &= \Pr\left[\Lambda_{acb_{n^*}} < \gamma_p\right] \\ &= \mathbb{E}\{\Pr\left[\Lambda_{acb_{n^*}} < \gamma_p | W_c = w\right]\}, \end{aligned} \tag{14.19}$$

where $\gamma_p = 2^{2\mathcal{R}_p} - 1$ and $\mathbb{E}\{\cdot\}$ denotes the statistical expectation. Since the evaluation of OP in (14.19) is cumbersome based on the exact SINR in (14.4), we proceed to derive a tight lower bound on the exact OP of satellite network for the *Cases* 1 and 2 in following subsections.

### *14.3.1 Case 1*

We derive the OP of satellite network for *Case* 1 in the following theorem.

**Theorem 1** *For Case 1, the tight lower bound on OP of satellite network can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \Psi(\mathcal{R}_p), \text{if}\, \gamma_p < \mu' \\ 1, \text{if}\, \gamma_p \geq \mu' \end{cases}, \tag{14.20}$$

*where* $\Psi(\mathcal{R}_p)$ *is given by*

$$\Psi(\mathcal{R}_p) = 1 - \alpha_{ac} \sum_{\kappa=0}^{m_{ac}-1} \frac{\zeta_{ac}(\kappa)}{\eta_a^{\kappa+1}} \sum_{n=0}^{\kappa} \frac{\kappa!}{n!} \Theta_{ac}^{-(\kappa+1-n)} \tilde{\gamma}_p^n \mathrm{e}^{-\Theta_{ac}\tilde{\gamma}_p} \times \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{1}{\Gamma(M_t)} \sum_{q=0}^{n} \binom{n}{q} \left( \frac{\Gamma(M_t+q)}{\left(\Theta_{ac}\tilde{\gamma}_p + \frac{1}{\Omega_t\eta_t}\right)^{M_t+q}} - \sum_{l=0}^{N} \binom{N}{l} (-1)^l \, \mathrm{e}^{\frac{-\tilde{\gamma}_p l}{\Omega_{cb}\eta_c}} \frac{\Gamma(M_t+q)}{\left(\Theta_{ac}\tilde{\gamma}_p + \frac{\tilde{\gamma}_p l}{\Omega_{cb}\eta_c} + \frac{1}{\Omega_t\eta_t}\right)^{M_t+q}} \right), \tag{14.21}$$

*with* $\mu' = \frac{\mu}{1-\mu}$, $\tilde{\gamma}_p = \frac{\gamma_p}{\mu-(1-\mu)\gamma_p}$ *and* $\Theta_{ac} = \frac{\beta_{ac}-\delta_{ac}}{\eta_a}$.

***Proof*** Please refer to Appendix 1 for proof. □

#### 14.3.1.1 Asymptotic OP

Here, we derive the asymptotic OP of satellite network at high SNR ($\eta_a, \eta_c \to \infty$) for *Case* 1 to reveal its diversity order.

**Corollary 1** *For Case 1, the asymptotic OP of satellite network under* $\gamma_p < \mu'$ *and* $\eta_a = \eta_c = \eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta} + \frac{\tilde{\gamma}_p}{\Omega_{cb}\eta}\right) \sum_{q=0}^{1} \binom{1}{q} \psi_1(q), \text{if} N = 1, \\ \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta}\right) \sum_{q=0}^{1} \binom{1}{q} \psi_1(q), \text{if} N > 1, \end{cases} \tag{14.22}$$

*where* $\psi_1(q) = \left(\frac{1}{\Omega_t\eta_t}\right)^{-q} \frac{\Gamma(M_t+q)}{\Gamma(M_t)}$.

***Proof*** The proof is in line with that given in Appendix 1. Considering $\eta \to \infty$, we can approximate (14.51) by neglecting the term $F_{\hat{\Lambda}_{ac}}(\tilde{\gamma}_p|w) F_{\hat{\Lambda}_{cb_{n^*}}}(\tilde{\gamma}_p|w)$ resulting in higher order as

$$\tilde{\mathcal{P}}_{\text{out},\infty}^{\text{sat}}(\mathcal{R}_p) = \mathbb{E}\{F_{\hat{\Lambda}_{ac}}(\tilde{\gamma}_p|w) + F_{\hat{\Lambda}_{cb_{n^*}}}(\tilde{\gamma}_p|w)\}. \tag{14.23}$$

Further, under $\eta \to \infty$, we can simplify the cdfs $F_{\hat{\Lambda}_{ac}}(x|w)$ and $F_{\hat{\Lambda}_{cb_{n^*}}}(x|w)$, for small argument $x$ as

$$F_{\hat{\Lambda}_{ac}}(x|w) \simeq \frac{\alpha_{ac} x(w+1)}{\eta} \tag{14.24}$$

and

$$F_{\hat{\Lambda}_{cb_{n^*}}}(x|w) \simeq \left(\frac{x(w+1)}{\Omega_{cb}\eta}\right)^N. \tag{14.25}$$

Similarly, one can have $F_{\hat{\Lambda}_{cd}}(x|w) \simeq \frac{x(w+1)}{\Omega_{cd}\eta}$ pertaining to $C$-$D$ link under high SNR. Subsequently, substituting these cdfs into (14.23), we can re-express (14.23) as

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta} + \frac{\tilde{\gamma}_p}{\Omega_{cb}\eta}\right)\mathbb{E}\{(w+1)\}, \text{if}\, N = 1, \\ \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta}\right)\mathbb{E}\{(w+1)\}, \text{if}\, N > 1, \end{cases} \tag{14.26}$$

Now, taking the expectation in (14.26) with respect to the pdf $f_{W_t}(w)$ given by (14.15) similar to the evaluation of $\mathcal{I}_1$ in Appendix 1, we can obtain (14.22). □

Note that in Corollary 1, the power $\eta_t$ of interfering TSs is assumed to be fixed. Unlike this, in the following, we derive the OP under the condition when the power $\eta_t$ of interfering TSs varies proportional to $\eta$ (i.e., $\eta_t = \nu\eta$), for some constant $\nu$.

**Corollary 2** *For Case 1, the asymptotic OP of satellite network under the conditions* $\gamma_p < \mu'$ *and* $\eta_t = \nu\eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \left(\alpha_{ac}\tilde{\gamma}_p + \frac{\tilde{\gamma}_p}{\Omega_{cb}}\right)\tilde{\psi}_1, \text{if}\, N = 1, \\ \left(\alpha_{ac}\tilde{\gamma}_p\right)\tilde{\psi}_1, \text{if}\, N > 1, \end{cases} \tag{14.27}$$

*where* $\tilde{\psi}_1 = \nu\Omega_t \frac{\Gamma(M_t+1)}{\Gamma(M_t)}$.

***Proof*** Considering $\eta_t = \nu\eta$ in (14.22), (14.27) can be obtained by neglecting the higher order terms. □

### *14.3.2 Case 2*

We derive the OP of satellite network for *Case* 2 in the following theorem.

**Theorem 2** *For Case 2, the tight lower bound on OP of satellite network can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \tilde{\Psi}(\mathcal{R}_p), \text{if}\, \gamma_p < \mu', \\ 1, \text{if}\, \gamma_p \geq \mu', \end{cases} \tag{14.28}$$

*where* $\tilde{\Psi}(\mathcal{R}_p)$ *is given by*

$$\Psi(\mathcal{R}_p) = 1 - \alpha_{ac} \sum_{\kappa=0}^{m_{ac}-1} \frac{\zeta_{ac}(\kappa)}{\eta_a^{\kappa+1}} \sum_{n=0}^{\kappa} \frac{\kappa!}{n!} \Theta_{ac}^{-(\kappa+1-n)} \tilde{\gamma}_p^n \mathrm{e}^{-\Theta_{ac}\tilde{\gamma}_p} \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \tag{14.29}$$
$$\times \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \frac{\Gamma(M_t + \Lambda)}{\Gamma(\Lambda)} \sum_{q=0}^{n} \binom{n}{q} \left(\mathcal{T}\left(q, \Theta_{ac}\tilde{\gamma}_p + \frac{1}{\Omega_t \eta_t}\right)\right.$$
$$\left. - \sum_{l=0}^{N} \binom{N}{l} (-1)^l \, \mathrm{e}^{\frac{-\tilde{\gamma}_p l}{\Omega_{cb}\eta_c}} \mathcal{T}\left(q, \Theta_{ac}\tilde{\gamma}_p + \frac{\tilde{\gamma}_p l}{\Omega_{cb}\eta_c} + \frac{1}{\Omega_t \eta_t}\right)\right).$$

*with function* $\mathcal{T}(x, y)$ *represented as*

$$\mathcal{T}(x, y) = y^{-\Lambda - M_t - x} G_{2,2}^{1,2}\left[\frac{\tilde{\Theta}_{ss}}{y} \middle| \begin{matrix} 1 - \Lambda - M_t - x, 1 - \Lambda \\ 0, 1 - \Lambda - M_t \end{matrix}\right]. \tag{14.30}$$

***Proof*** The proof follows the procedure in Appendix 1. Here, for *Case* 2, we re-express $\mathcal{I}_1$ taking into account the pdf $f_{W_c}(w)$ in (14.16) as

$$\mathcal{I}_1 = \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \int_0^{\infty} w^{M_t - 1} (w + 1)^n \, \mathrm{e}^{-\left(\Theta_{ac}\tilde{\gamma}_p + \frac{1}{\Omega_t \eta_t}\right)w} \tag{14.31}$$
$$\times \left(1 - \sum_{l=0}^{N} \binom{N}{l} (-1)^l \mathrm{e}^{\frac{-\tilde{\gamma}_p l(w+1)}{\Omega_{cb}\eta_c}}\right) {}_1F_1\left(\Lambda; \Lambda + M_t; -\tilde{\Theta}_{ss} w\right) dw.$$

Next, we first apply the binomial expansion for $(w + 1)^n$ along with (Wolfram's functions site 2020, Eq. 07.20.26.0006.01) for ${}_1F_1(\cdot; \cdot; \cdot)$ and then, evaluate the resulting integral using (Gradshteyn and Ryzhik 2000, Eq. 7.813.1), to get (14.28) □

#### 14.3.2.1 Asymptotic OP

Here, we derive the asymptotic OP of satellite network at high SNR ($\eta_a, \eta_c \to \infty$) for *Case* 2 to reveal its diversity order.

**Corollary 3** *For Case 2, the asymptotic OP of satellite network under* $\gamma_p < \mu'$ *and* $\eta_a = \eta_c = \eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta} + \frac{\tilde{\gamma}_p}{\Omega_{cb}\eta}\right) \sum_{q=0}^{1} \binom{1}{q} \psi_2(q), \text{if} N = 1, \\ \left(\frac{\alpha_{ac}\tilde{\gamma}_p}{\eta}\right) \sum_{q=0}^{1} \binom{1}{q} \psi_2(q), \text{if} N > 1, \end{cases} \tag{14.32}$$

*where*

$$\psi_2(q) = \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \frac{\Gamma(M_t + \Lambda)}{\Gamma(\Lambda)} \mathcal{T}\left(q, \frac{1}{\Omega_t \eta_t}\right). \quad (14.33)$$

***Proof*** The proof is similar to that in Corollary 1. Taking the required expectation in (14.26) using the pdf $f_{W_c}(w)$ given by (14.16) and performing the resulting integration as in the proof of Theorem 2, we can achieve (14.32). □

Note that in Corollary 3, the powers $\eta_t$ and $\eta_s$ of interfering TSs and ETSs are assumed to be fixed. Unlike this, in the following, we derive the OP under the condition when the powers $\eta_t$ and $\eta_s$ of interfering TSs and ETSs vary proportional to $\eta$ (i.e., $\eta_t = \eta_s = \nu\eta$), for some constant $\nu$.

**Corollary 4** *For Case 2, the asymptotic OP of satellite network under the conditions* $\gamma_p < \mu'$ *and* $\eta_t = \eta_s = \nu\eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) = \begin{cases} \left(\alpha_{ac}\tilde{\gamma}_p + \frac{\tilde{\gamma}_p}{\Omega_{cb}}\right)\tilde{\psi}_2, \text{if } N = 1, \\ \left(\alpha_{ac}\tilde{\gamma}_p\right)\tilde{\psi}_2, \text{if } N > 1, \end{cases} \quad (14.34)$$

*where*

$$\tilde{\psi}_2 = \nu \widetilde{\sum} \Xi_s(M_s) \left(\frac{1}{\Omega_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \frac{\Gamma(M_t + \Lambda)}{\Gamma(\Lambda)} \mathcal{T}\left(1, \frac{1}{\Omega_t \eta_t}\right). \quad (14.35)$$

***Proof*** Considering $\eta_t = \eta_s = \nu\eta$ in (14.32) and neglecting the higher order terms, one can determine (14.34). □

***Remark 1*** Upon re-expressing (14.22) and (14.32) as $\mathcal{G}_c \eta^{-\mathcal{G}_d}$ with $\eta_s = \eta_t$ set fixed, and ignoring the higher order terms, the achievable diversity order $\mathcal{G}_d$ of satellite network is unity. However, from (14.27) and (14.34), under the case that interferers' power varies proportional to $\eta$, i.e., $\eta_s = \eta_t = \nu\eta$ for some constant $\nu$, $\mathcal{G}_d$ reduces to zero. Here, $\mathcal{G}_c$ represents the achievable coding gain.

## 14.4 Outage Performance of IoT Network

In this section, we assess the outage performance of secondary IoT network and carry out asymptotic OP analysis to reveal its achievable diversity order.

For a target rate $\mathcal{R}_S$, the OP of IoT network can be calculated, based on SINR in (14.5), as

$$\begin{aligned} \mathcal{P}_{\text{out}}^{\text{IoT}}(\mathcal{R}_S) &= \Pr[\Lambda_{acd} < \gamma_s] \quad (14.36) \\ &= \mathbb{E}\left\{\Pr\left[\frac{\mu\hat{\Lambda}_{cd}(\hat{\Lambda}_{ac} + 1)}{\mu\hat{\Lambda}_{cd} + \hat{\Lambda}_{ac} + 1} < \mu'\gamma_s \middle| w\right]\right\}, \end{aligned}$$

where $\gamma_s = 2^{2\mathcal{R}_s} - 1$. Since the exact analysis of OP in (14.36) is intractable, we derive a tight lower bound on the exact OP of IoT network (say $\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_\text{S})$) based on the bound $\frac{XY}{X+Y} \leq \min(X, Y)$ as

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_\text{S}) = \mathbb{E}\{\underbrace{\Pr[\min(\mu\hat{\Lambda}_{cd}, \hat{\Lambda}_{ac} + 1) < \mu'\gamma_s|w]}_{\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_\text{S}|w)}\}. \tag{14.37}$$

After the transformation of variates for $\hat{\Lambda}_{ac} + 1$ and some manipulation, $\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_\text{S}|w)$ in (14.37) can be expressed as

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_\text{S}|w) = \begin{cases} \mathbb{E}\{F_{\mu\hat{\Lambda}_{cd}}(\mu'\gamma_s|w)\}, \text{if}\, \gamma_s < \frac{1}{\mu'}, \\ \mathbb{E}\{1 - \overline{F}_{\hat{\Lambda}_{ac}}(\mu'\gamma_s - 1|w)\overline{F}_{\mu\hat{\Lambda}_{cd}}(\mu'\gamma_s|w)\}, \text{if}\, \gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.38}$$

Next, we evaluate the OP of IoT network based on (14.37) for *Cases* 1 and 2.

### *14.4.1 Case 1*

We derive the OP of IoT network for *Case* 1 in the following theorem.

**Theorem 3** *For Case 1, the tight lower bound on OP of IoT network can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_s) = \begin{cases} \Psi_1(\mathcal{R}_s), \text{if}\, \gamma_s < \frac{1}{\mu'}, \\ \Psi_2(\mathcal{R}_s), \text{if}\, \gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.39}$$

*where* $\gamma_s = 2^{2\mathcal{R}_s} - 1$, *and the equations* $\Psi_1(\mathcal{R}_s)$ *and* $\Psi_2(\mathcal{R}_s)$ *are given, respectively, by*

$$\Psi_1(\mathcal{R}_s) = 1 - \mathrm{e}^{\frac{-\gamma_s}{(1-\mu)\Omega_{cd}\eta_c}} \left(\frac{1}{\Omega_t\eta_t}\right)^{M_t} \left(\frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c} + \frac{1}{\Omega_t\eta_t}\right)^{-M_t} \tag{14.40}$$

*and*

$$\Psi_2(\mathcal{R}_s) = 1 - \alpha_{ac} \sum_{\kappa=0}^{m_{ac}-1} \frac{\zeta_{ac}(\kappa)}{\eta_a^{\kappa+1}} \sum_{n=0}^{\kappa} \frac{\kappa!}{n!} \Theta_{ac}^{-(\kappa+1-n)} \tilde{\gamma}_s^n \mathrm{e}^{-\left(\Theta_{ac}\tilde{\gamma}_s + \frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c}\right)} \sum_{q=0}^{n} \binom{n}{q} \tag{14.41}$$

$$\times \left(\frac{1}{\Omega_t\eta_t}\right)^{M_t} \frac{\Gamma(M_t + q)}{\Gamma(M_t)} \left(\Theta_{ac}\tilde{\gamma}_s + \frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c} + \frac{1}{\Omega_t\eta_t}\right)^{-M_t-q},$$

*with* $\tilde{\gamma}_s = \mu'\gamma_s - 1$.

***Proof*** Please refer to Appendix 2 for proof. □

#### 14.4.1.1 Asymptotic OP

Here, we derive the asymptotic OP of IoT network at high SNR ($\eta_a, \eta_c \to \infty$) for *Case* 1 to reveal its diversity order.

**Corollary 5** *For Case 1, the asymptotic OP of IoT network under $\eta_a = \eta_c = \eta$ can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{IoT}(\mathcal{R}_S) = \begin{cases} \frac{1}{\Omega_{cd}\eta}\left(\frac{\gamma_s}{(1-\mu)}\right)\sum_{n=0}^{1}\binom{1}{n}\psi_1(n), \text{if}\,\gamma_s < \frac{1}{\mu'}, \\ \left(\frac{\alpha_{ac}\tilde{\gamma}_s}{\eta} + \frac{1}{\Omega_{cd}\eta}\left(\frac{\gamma_s}{(1-\mu)}\right)\right)\sum_{n=0}^{1}\binom{1}{n}\psi_1(n), \text{if}\,\gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.42}$$

***Proof*** The proof follows the similar steps as given in Appendix 2. By substituting the high SNR approximation cdfs (14.24) and (14.25) in (14.54) and (14.55), and taking the expectation as in Appendix 1, we can reach (14.42). □

Note that in Corollary 5, the power $\eta_t$ of interfering TSs is assumed to be fixed. Unlike this, in the following, we derive the OP under the condition when the power $\eta_t$ of interfering TSs varies proportional to $\eta$ (i.e., $\eta_t = \nu\eta$), for some constant $\nu$.

**Corollary 6** *For Case 1, the asymptotic OP of IoT network under the conditions $\eta_t = \nu\eta$ can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_S) = \begin{cases} \frac{1}{\Omega_{cd}}\left(\frac{\gamma_s}{(1-\mu)}\right)\tilde{\psi}_1, \text{if}\,\gamma_s < \frac{1}{\mu'}, \\ \left(\alpha_{ac}\tilde{\gamma}_s + \frac{1}{\Omega_{cd}}\left(\frac{\gamma_s}{(1-\mu)}\right)\right)\tilde{\psi}_1, \text{if}\,\gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.43}$$

***Proof*** Considering $\eta_t = \nu\eta$ in (14.42) and neglecting the higher order terms, one can determine (14.43). □

### *14.4.2 Case 2*

We derive the OP of IoT network for *Case* 2 in the following theorem.

**Theorem 4** *For Case 2, the tight lower bound on OP of IoT network can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_s) = \begin{cases} \tilde{\Psi}_1(\mathcal{R}_s), \text{if}\,\gamma_s < \frac{1}{\mu'}, \\ \tilde{\Psi}_2(\mathcal{R}_s), \text{if}\,\gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.44}$$

*where* $\tilde{\Psi}_1(\mathcal{R}_s)$ *and* $\tilde{\Psi}_2(\mathcal{R}_s)$ *are given, respectively, by*

$$\tilde{\Psi}_1(\mathcal{R}_s) = 1 - \mathrm{e}^{\frac{-\gamma_s}{(1-\mu)\Omega_{cd}\eta_c}} \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \frac{\Gamma(M_t + \Lambda)}{\Gamma(\Lambda)} \times \mathcal{T}\left(0, \frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c} + \frac{1}{\Omega_t \eta_t}\right) \tag{14.45}$$

*and*

$$\tilde{\Psi}_2(\mathcal{R}_s) = 1 - \alpha_{ac} \sum_{\kappa=0}^{m_{ac}-1} \frac{\zeta_{ac}(\kappa)}{\eta_a^{\kappa+1}} \sum_{n=0}^{\kappa} \frac{\kappa!}{n!} \Theta_{ac}^{-(\kappa+1-n)} \tilde{\gamma}_s^n \mathrm{e}^{-\left(\Theta_{ac}\tilde{\gamma}_s + \frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c}\right)} \sum_{q=0}^{n} \binom{n}{q} \times \widetilde{\sum} \frac{\Xi_s(M_s)}{\eta_s^{\Lambda}} \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{\Phi(M_t, \Lambda)}{\Gamma(M_t)} \frac{\Gamma(M_t + \Lambda)}{\Gamma(\Lambda)} \times \mathcal{T}\left(q, \Theta_{ac}\tilde{\gamma}_s + \frac{\gamma_s}{(1-\mu)\Omega_{cd}\eta_c} + \frac{1}{\Omega_t \eta_t}\right). \tag{14.46}$$

***Proof*** The proof follows that of Theorem 3. □

#### 14.4.2.1 Asymptotic OP

Here, we derive the asymptotic OP of IoT network at high SNR ($\eta_a, \eta_c \to \infty$) for *Case* 2 to reveal its diversity order.

**Corollary 7** *For Case 2, the asymptotic OP of IoT network under* $\eta_a = \eta_c = \eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_s) = \begin{cases} \frac{1}{\Omega_{cd}\eta} \left(\frac{\gamma_s}{(1-\mu)}\right) \sum_{n=0}^{1} \binom{1}{n} \psi_2(n), \text{if} \gamma_s < \frac{1}{\mu'}, \\ \left(\frac{\alpha_{ac}\tilde{\gamma}_s}{\eta} + \frac{1}{\Omega_{cd}\eta} \left(\frac{\gamma_s}{(1-\mu)}\right)\right) \sum_{n=0}^{1} \binom{1}{n} \psi_2(n), \text{if} \gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.47}$$

***Proof*** The proof follows that of Corollary 5. □

Note that in Corollary 7, the powers $\eta_t$ and $\eta_s$ of interfering TSs and ETSs are assumed to be fixed. Unlike this, in the following, we derive the OP under the condition when the powers $\eta_t$ and $\eta_s$ of interfering TSs and ETSs vary proportional to $\eta$ (i.e., $\eta_t = \eta_s = \nu\eta$), for some constant $\nu$.

**Corollary 8** *For Case 2, the asymptotic OP of IoT network under the conditions* $\eta_t = \eta_s = \nu\eta$ *can be given as*

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{IoT}}(\mathcal{R}_s) = \begin{cases} \frac{1}{\Omega_{cd}}\left(\frac{\gamma_s}{(1-\mu)}\right)\tilde{\psi}_2, \text{if}\, \gamma_s < \frac{1}{\mu'}, \\ \left(\alpha_{ac}\tilde{\gamma}_s + \frac{1}{\Omega_{cd}}\left(\frac{\gamma_s}{(1-\mu)}\right)\right)\tilde{\psi}_2, \text{if}\, \gamma_s \geq \frac{1}{\mu'}, \end{cases} \tag{14.48}$$

***Proof*** The proof follows the proof of Corollary 6. □

***Remark 2*** Upon re-expressing (14.42) and (14.47) as $\mathcal{G}_c\eta^{-\mathcal{G}_d}$ with $\eta_s = \eta_t$ set fixed, and ignoring the higher order terms, the achievable diversity order $\mathcal{G}_d$ of IoT network is unity. However, from (14.43) and (14.48), under the case that interferers' power varies proportional to $\eta$, i.e., $\eta_s = \eta_t = \nu\eta$ for some constant $\nu$, $\mathcal{G}_d$ reduces to zero.

## 14.5 Adaptive Power Splitting Factor

In this section, we give an adaptive scheme to obtain the effective value of power splitting factor $\mu$ for spectrum sharing. Recalling the mathematical condition $\gamma_p < \mu'$ in Theorem 1, the feasible range of $\mu$ can be given as $\frac{\gamma_p}{1+\gamma_p} \leq \mu \leq 1$. Furthermore, to calculate $\mu$, we require a QoS constraint to protect the satellite network from IoT transmissions. For this, the OP of satellite network $\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p)$ is guaranteed below a predetermined QoS level $\epsilon$, i.e.,

$$\tilde{\mathcal{P}}_{\text{out}}^{\text{sat}}(\mathcal{R}_p) \leq \epsilon. \tag{14.49}$$

It may be intuitively seen that the QoS constraint at equality yields the value of $\mu$ that minimizes the OP of IoT network. Note that the value of $\mu$ can be determined through exhaustive search method as there is no closed-form solution available. For comparison, we also consider the assignment of an arbitrary fixed value of $\mu$.

## 14.6 Numerical and Simulation Results

In this section, we present the numerical results for considered OMSTN. Here, we validate our theoretical results by simulations for $10^6$ independent channel realizations. We set the rate parameters $\mathcal{R}_p = \mathcal{R}_S = 0.5$ bps/Hz in order to have $\gamma_p = \gamma_s = 1$ (unless stated otherwise). We further set $\Omega_{cb} = \Omega_{cd} = 1$, and SNR $\eta_a = \eta_c = \eta$. The SR fading parameters for satellite link $A - C$ are considered as $(m_{ac}, \flat_{ac}, \Omega_{ac} = 5, 0.251, 0.279)$ (i.e., light shadowing) and $(m_{ac}, \flat_{ac}, \Omega_{ac} = 2, 0.063, 0.0005)$ (i.e., heavy shadowing) for *Cases* 1 and 2, respectively (Upadhyay and Sharma 2016). The SR fading parameters for interfering ETSs are considered as $(m_s, \flat_s, \Omega_s = 1, 0.063, 0.0007)$ for heavy shadowing. We set $\Omega_t = 0.1$ for interfering TSs, and interferers' power $\eta_s = \eta_t$. Here, we consider two scenarios viz., (a) when interferers' power is fixed, i.e., $\eta_s = \eta_t = 5$ dB, and (b) when interferers'

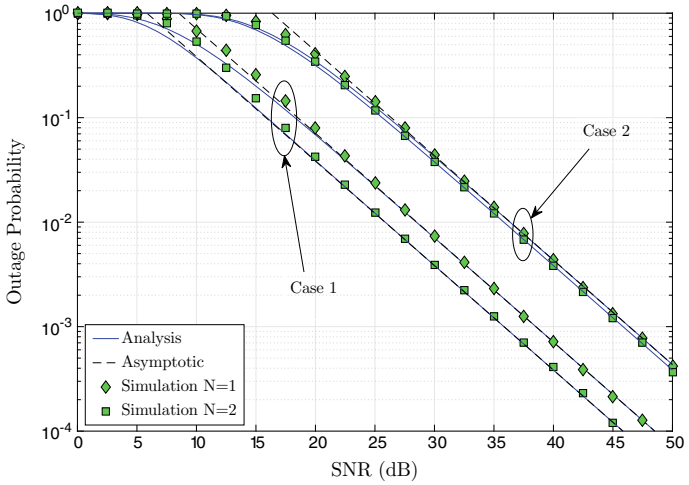



**Fig. 14.2** Outage probability of satellite network

power is proportional to $\eta$, i.e., $\eta_s = \eta_t = \nu\eta$, with $\nu = -20$ dB. Moreover, we set the number of interferers $M_t = 2$ for *Case* 1 and $M_s = M_t = 2$ for *Case* 2.

In Fig. 14.2, we plot the OP versus SNR curves for primary satellite network for different number of primary terrestrial receivers. In this figure, the set of curves are obtained for a fixed value of power splitting factor (i.e., $\mu = 0.75$). Here, we can observe that our derived theoretical lower bound on exact OP are well-aligned to the simulation results over medium-to-high SNR regime. Even, in low SNR region, the theoretical lower bound is acceptably tight. Furthermore, by observing the slope of curves when $N$ changes from 1 to 2, we conclude that the diversity order of the considered OMSTN remains unaffected of the choice of $N$.

In Fig. 14.3, we plot the OP versus SNR curves for secondary IoT network for different $\gamma_s$. Hereby, we set the value of $\gamma_s = 1$ and 0.3 to realize the scenarios $\gamma_s \geq \frac{1}{\mu'}$ and $\gamma_s < \frac{1}{\mu'}$, respectively, as observed in Theorem 3. Here also, we assume that power splitting factor has a fixed value 0.75 and $N = 2$. Likewise satellite network, we can observe that the theoretical lower bound on exact OP of IoT networks are in close proximity to simulation results. Further, one can confirm a diversity order of unity for the IoT network based on the slope of various OP curves.

In Fig. 14.4, we plot the OP versus SNR curves for primary satellite network for different number of primary terrestrial receivers under the scenario where interferers' power is proportional to SNR $\eta$, i.e., $\eta_s = \eta_t = \nu\eta$ for $\nu = -20$ dB. Here, we set fixed value of $\mu$ as 0.75. Note that unlike the scenario discussed in Fig. 14.2, hereby, the slope of OP curves becomes zero in medium-to-high SNR regime. This indicates the loss of diversity order of the satellite network to zero. Thus, no diversity gain is achievable under this scenario for any choice of $N$, i.e., 1 or 2.

In Fig. 14.5, we plot the OP versus SNR curves for secondary IoT network for different $\gamma_s$ under the scenario where interferers' power is proportional to SNR $\eta$,

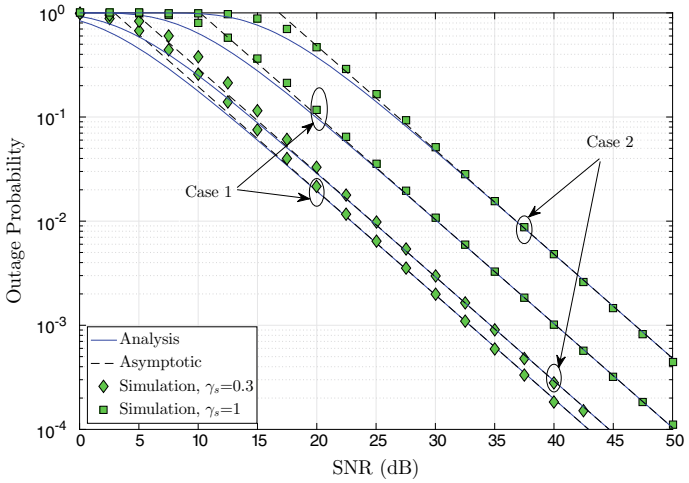



**Fig. 14.3** Outage probability of IoT network

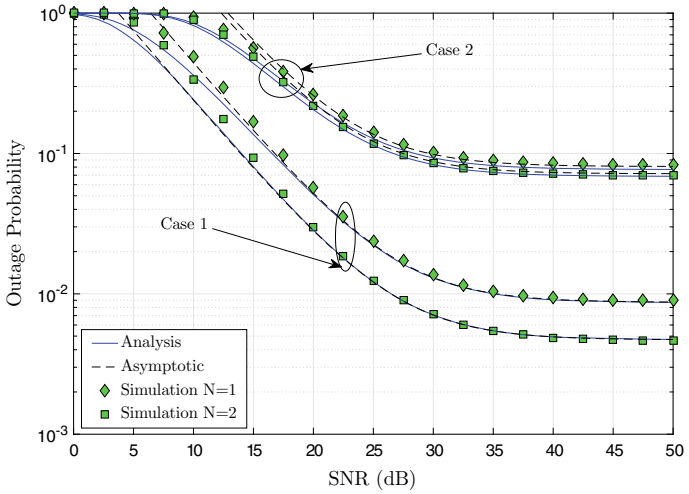



**Fig. 14.4** Outage probability of satellite network

i.e., $\eta_s = \eta_t = \nu\eta$ for $\nu = -20$ dB. Likewise the satellite network, there is the loss of diversity order of IoT network to zero in this scenario as apparent from the flat OP curves in medium-to-high SNR regime.

In Fig. 14.6, we plot the OP versus SNR curves for secondary IoT network by adaptively choosing the $\mu$ according to the procedure described in Sect. 14.5. Hereby, we set $\epsilon = 0.1$ owing to the OP of 10%. It is interesting to observe here that OP can be slightly improved in low SNR regime through adaptive power splitting factor when $N$ changes from 1 to 2. The performance gap is more noticeable in $Case$ 1 for given SR

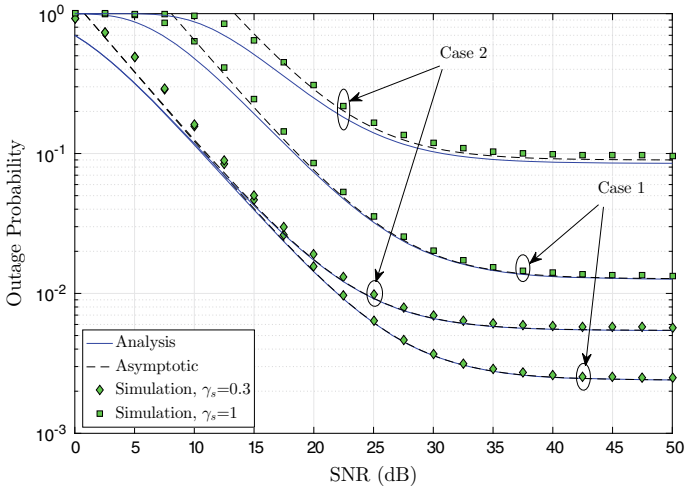



**Fig. 14.5** Outage probability of IoT network

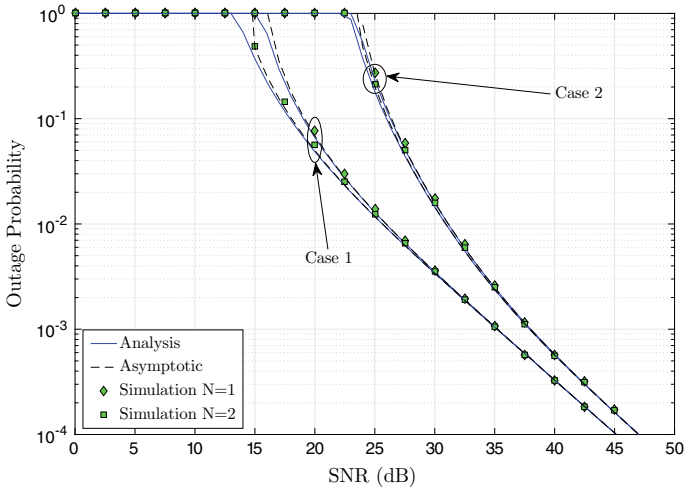



**Fig. 14.6** Outage probability of IoT network

fading parameters. However, the diversity order of the OMSTN remains unaffected of $N$. More importantly, the adaptive choice of $\mu$ always protect the satellite network by guaranteeing a desired QoS.

In Fig. 14.7, we plot the OP versus SNR curves for secondary IoT network by adaptively choosing $\mu$ under the scenario when interferers' power is proportional to $\eta$, i.e., $\eta_s = \eta_t = \nu\eta$ for $\nu = -20\,\text{dB}$. Here, the zero diversity order at high SNR is apparent through flat OP curves. However, the QoS for primary satellite network can be effectively guaranteed under this scenario.

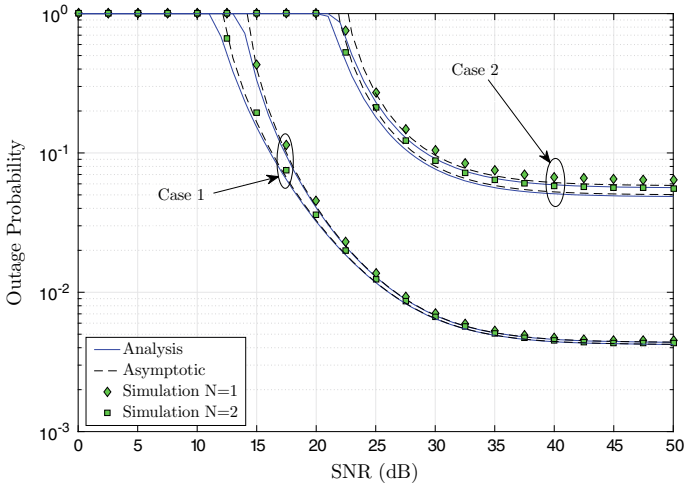



**Fig. 14.7** Outage probability of IoT network

## 14.7 Summary and Future Extensions

This chapter presented a comprehensive outage performance analysis of an OMSTN where terrestrial secondary IoT devices cooperatively assist the multi-user primary satellite communications to access spectrum for secondary communications. We derived the tight closed-form OP expressions of both satellite and IoT networks for *Case* 1 and *Case* 2 in the presence of interference from TSs and ETSs. We have also performed achievable diversity order analysis of the considered OMSTN under two scenarios, viz., when interferers' power is fixed and when it varies with the transmit power of main satellite and IoT nodes. We have also presented the scheme for adaptively choosing the power splitting factor. We depicted that in general adaptive power splitting factor improves the performance of IoT network while guaranteeing the QoS of satellite network. The presence of interference degrades the performance of both the satellite and IoT networks. Moreover, the diversity order of both satellite and IoT networks reduces to zero under the scenario when interferers' power varies proportionally with that of main satellite and IoT nodes. In this chapter, we analyzed the performance of OMSTN with single-antenna nodes. The performance of such OMSTN with multi-antenna nodes may be considered as a future extension. Further, the analysis of multi-antenna satellite nodes with correlated fading scenario is a practical problem. Other major research extensions include the techniques for secrecy enhancement, radio frequency energy harvesting, reducing the impact of hardware impairments, etc.

## Appendix 1

To derive the OP of satellite network, we first apply the bound $\frac{XY}{X+Y} \leq \min(X, Y)$ to re-express the SINR in (14.4) as (Sharma et al. 2017)

$$\Lambda_{acb_{n^*}} \leq \Lambda^{\star}_{acb_{n^*}} = \frac{\mu}{(1-\mu) + 1/\min(\hat{\Lambda}_{ac}, \hat{\Lambda}_{cb_{n^*}})}. \tag{14.50}$$

Then, substituting the upper bound SINR $\Lambda^{\star}_{acb_{n^*}}$ in (14.19) yields

$$\tilde{\mathcal{P}}^{\text{sat}}_{\text{out}}(\mathcal{R}_p) = \mathbb{E}\{1 - \overline{F}_{\hat{\Lambda}_{ac}}(\tilde{\gamma}_p|w)\overline{F}_{\hat{\Lambda}_{cb_{n^*}}}(\tilde{\gamma}_p|w)\} \tag{14.51}$$

where $\overline{F}_X(\cdot|w) = 1 - F_X(\cdot|w)$ is the complementary cdf. Now, invoking (14.10) and (14.13) in (14.51) after applying binomial expansion in (14.13), we can get

$$\begin{aligned}\tilde{\mathcal{P}}^{\text{sat}}_{\text{out}}(\mathcal{R}_p) =& 1 - \alpha_{ac} \sum_{\kappa=0}^{m_{ac}-1} \frac{\zeta_{ac}(\kappa)}{\eta_a^{\kappa+1}} \sum_{n=0}^{\kappa} \frac{\kappa!}{n!} \Theta_{ac}^{-(\kappa+1-n)} \tilde{\gamma}_p^n \mathrm{e}^{-\Theta_{ac}\tilde{\gamma}_p} \\ &\times \underbrace{\mathbb{E}\left\{(w+1)^n \mathrm{e}^{-\Theta_{ac}\tilde{\gamma}_p w}\left(1 - \sum_{l=0}^{N}\binom{N}{l}(-1)^l \mathrm{e}^{\frac{-\tilde{\gamma}_p l(w+1)}{\Omega_{cb}\eta_c}}\right)\right\}}_{\mathcal{I}_1}.\end{aligned} \tag{14.52}$$

We further re-express the term $\mathcal{I}_1$ in (14.52) with the help of pdf $f_{W_t}(w)$ given by (14.15) as

$$\begin{aligned}\mathcal{I}_1 =& \left(\frac{1}{\Omega_t \eta_t}\right)^{M_t} \frac{1}{\Gamma(M_t)} \int_0^{\infty} w^{M_t-1} (w+1)^n \, \mathrm{e}^{-\left(\Theta_{ac}\tilde{\gamma}_p + \frac{1}{\Omega_t\eta_t}\right)w} \\ &\times \left(1 - \sum_{l=0}^{N}\binom{N}{l}(-1)^l \mathrm{e}^{\frac{-\tilde{\gamma}_p l(w+1)}{\Omega_{cb}\eta_c}}\right) dw.\end{aligned} \tag{14.53}$$

Finally, by applying the binomial expansion for $(w+1)^n$ and evaluating the resulting integral using (Gradshteyn and Ryzhik 2000, Eq. 3.351.3), we can obtain (14.20).

## Appendix 2

We evaluate (14.38) based on the cdfs given by (14.10) and (14.12), where

$$\Psi_1(\mathcal{R}_s) = \mathbb{E}\{F_{\mu\hat{\Lambda}_{cd}}(\mu'\gamma_s|w)\}, \tag{14.54}$$

and

$$\Psi_2(\mathcal{R}_s) = \mathbb{E}\{1 - \overline{F}_{\hat{\Lambda}_{ac}}(\mu'\gamma_s - 1|w)\overline{F}_{\mu\hat{\Lambda}_{cd}}(\mu'\gamma_s|w)\}. \tag{14.55}$$

Here, $\Psi_1(\mathcal{R}_s)$ and $\Psi_2(\mathcal{R}_s)$ can be computed by following the approach similar to that in $\mathcal{I}_1$ in Appendix 1 with the pdf $f_{W_t}(w)$ given by (14.15).

## References

An K, Lin M, Ouyang J, Huang Y, Zheng G (2014) Symbol error analysis of hybrid satellite terrestrial cooperative networks with cochannel interference. IEEE Commun Lett 18(11):1947–1950

An K, Lin M, Liang T (2015) On the performance of multiuser hybrid satellite terrestrial relay networks with opportunistic scheduling. IEEE Commun Lett 19(10):1722–1725

An K, Lin M, Liang T, Wang J, Wang J, Huang Y, Swindlehurst AL (2015) Performance analysis of multi-antenna hybrid satellite-terrestrial relay networks in the presence of interference. IEEE Trans Commun 63(11):4390–4404

An K, Lin M, Ouyang J, Zhu W (2016) Secure transmission in cognitive satellite terrestrial networks. IEEE J Sel Areas Commun 34(11):3025–3037

An K, Lin M, Zhu W, Huang Y, Zheng G (2016) Outage performance of cognitive hybrid satellite terrestrial networks with interference constraint. IEEE Trans Veh Technol 65(11):9397–9404

Bhatnagar MR, Arti MK (2013) Performance analysis of AF based hybrid satellite-terrestrial cooperative network over generalized fading channels. IEEE Commun Lett 17(10):1912–1915

Chen L, Thombre S, Järvinen K, Lohan ES, Alén-Savikko A, Leppäkoski H, Bhuiyan MZH, Bu-Pasha S, Ferrara GN, Honkala S, Lindqvist J, Ruotsalainen L, Korpisaari P, Kuusniemi H (2017) Robustness, security and privacy in location based services for future IoT: a survey. IEEE Access 5:8956–8977

ETSI EN 302 307-2 V1.1.1 (2014) Digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for interactive services, news gathering and other broadband satellite applications; Part 2: DVB-S2 extensions (DVB-S2X)

Evans B, Werner M, Lutz E, Bousquet M, Corazza GE, Maral G, Rumeau R (2005) Integration of satellite and terrestrial systems in future multimedia communications. IEEE Wirel Commun 12(5):72–80

Evans B, Onireti O, Spathopoulos T, Imran MA (2015) The role of satellites in 5G. In: 2015 23rd European signal processing conference (EUSIPCO), pp 2756–2760

Gradshteyn IS, Ryzhik IM (2000) Tables of integrals, series and products, 6th edn. Academic Press, New York

Guo K, An K, Zhang B, Huang Y, Zheng G (2018) Outage analysis of cognitive hybrid satellite-terrestrial networks with hardware impairments and multi-primary users. IEEE Wirel Commun Lett 7(5):816–819

Guo K, Lin M, Zhang B, Zhu W, Wang J, Tsiftsis TA (2019) On the performance of LMS communication with hardware impairments and interference. IEEE Trans Commun 67(2):1490–1505

Jia M, Gu X, Guo Q, Xiang W, Zhang N (2016) Broadband hybrid satellite terrestrial communication systems based on cognitive radio toward 5G. IEEE Wirel Commun 23(6):96–106

Jo KY (2011) Satellite communications network design and analysis. Artech House, Norwood

Kandeepan S, De Nardis L, Di Benedetto M, Guidotti A, Corazza GE (2010) Cognitive satellite terrestrial radios. In: 2010 IEEE global telecommunications conference (GLOBECOM 2010), pp 1–6

Khan AA, Rehmani MH, Rachedi A (2017) Cognitive-radio-based internet of things: applications, architectures, spectrum related functionalities, and future research directions. IEEE Wirel Commun 24(3):17–25

Kim S (2011) Evaluation of cooperative techniques for hybrid/integrated satellite systems. In: 2011 IEEE international conference on communications (ICC), pp 1–5
Lysogor II, Voskov LS, Efremov SG (2018) Survey of data exchange formats for heterogeneous LPWAN-satellite IoT networks. In: 2018 Moscow workshop on electronic and networking technologies (MWENT), pp 1–5
Manna R, Louie RHY, Li Y, Vucetic B (2011) Cooperative spectrum sharing in cognitive radio networks with multiple antennas. IEEE Trans Sig Process 59(11):5509–5522
Miridakis NI, Vergados DD, Michalas A (2015) Dual-hop communication over a satellite relay and shadowed Rician channels. IEEE Trans Veh Technol 64(9):4031–4040
Paillassa B, Escrig B, Dhaou R, Boucheret ML, Bes C (2011) Improving satellite services with cooperative communications. Int J Sat Commun 29(6):479–500
Riebeek H, Simmon R (2009) Catalog of earth satellite orbits. Available: https://earthobservatory.nasa.gov/features/OrbitsCatalog
Ruan Y, Li Y, Wang C, Zhang R, Zhang H (2017) Outage performance of integrated satellite-terrestrial networks with hybrid CCI. IEEE Commun Lett 21(7):1545–1548
Salman O, Elhajj I, Chehab A, Kayssi A (2018) IoT survey: an SDN and fog computing perspective. Comput Netw 143:221–246
Satellite frequency bands. Available: https://www.esa.int/Applications/Telecommunications_Integrated_Applications/Satellite_frequency_bands
Sharma PK, Yogesh B, Gupta D (2020) Internet of things-enabled overlay satellite-terrestrial networks in the presence of interference. In: 2020 national conference on communications (NCC). https://doi.org/10.1109/ncc48643.2020.9056048, (Online). Available: https://arxiv.org/abs/2001.05176
Sharma SK, Chatzinotas S, Ottersten B (2013) Cognitive radio techniques for satellite communication systems. In: 2013 IEEE 78th vehicular technology conference (VTC Fall), pp 1–5
Sharma PK, Kim DI (2020) Secure 3D mobile UAV relaying for hybrid satellite terrestrial networks. IEEE Trans Wirel Commun 19(4):2770–2784
Sharma PK, Upadhyay PK, da Costa DB, Bithas PS, Kanatas AG (2017) Performance analysis of overlay spectrum sharing in hybrid satellite-terrestrial systems with secondary network selection. IEEE Trans Wirel Commun 16(10):6586–6601
Sharma PK, Deepthi D, Kim DI (2020) Outage probability of 3-D mobile UAV relaying for hybrid satellite-terrestrial networks. IEEE Commun Lett 24(2):418–422
Sreng S, Escrig B, Boucheret M (2013) Exact symbol error probability of hybrid/integrated satellite-terrestrial cooperative network. IEEE Trans Wirel Commun 12(3):1310–1319
Starlink network. Available: https://www.starlink.com/
Upadhyay PK, Sharma PK (2016) Multiuser hybrid satellite-terrestrial relay networks with co-channel interference and feedback latency. In: 2016 European conference on networks and communications (EuCNC), pp 174–178
Upadhyay PK, Sharma PK (2016) Max-max user-relay selection scheme in multiuser and multirelay hybrid satellite-terrestrial relay systems. IEEE Commun Lett 20(2):268–271
Vanelli-Corali A, Corazza GE, Karagiannidis GK, Mathiopoulos PT, Michalopoulos DS, Mosquera C, Papaharalabos S, Scalise S (2007) Satellite communications: research trends and open issues. In: 2007 international workshop on satellite and space communications, pp 71–5
Vassaki S, Poulakis MI, Panagopoulos AD, Constantinou P (2013) Power allocation in cognitive satellite terrestrial networks with QoS constraints. IEEE Commun Lett 17(7):1344–1347
Vodafone Group (2018) Extending the reach of cellular networks, pp 1–3
Wolfram's functions site. Available: https://functions.wolfram.com
Yang L, Hasna MO (2015) Performance analysis of amplify-and-forward hybrid satellite-terrestrial networks with cochannel interference. IEEE Trans Commun 63(12):5052–5061
Zou Y, Zhu J, Zheng B, Yao Y (2010) An adaptive cooperation diversity scheme with best-relay selection in cognitive radio networks. IEEE Trans Sig Process 58(10):5438–5445

# Chapter 15
# Resource Allocation in D2D Communications


**Ajay Bhardwaj and Devendra Singh Gurjar**

**Abstract** The proliferation of mobile devices and data-hungry applications running on them leads to a massive growth of wireless data traffic. Supporting this ever-increasing data demands and communication rate requires reconsideration of the existing cellular network architecture. Device-to-device (D2D) communications, which allow two mobile devices in the proximity to communicate with each other, emerge as a potential solution to this challenge. It provides multifold gains in terms of transmission-rate gain, frequency-reuse gain, coverage-gain, and hop-gain. However, extensive deployment of D2D communications in the cellular networks poses several intrinsic challenges, such as severe interference to the primary cellular users, rapid battery depletion of D2D transmitters in relaying scenarios. Therefore, this chapter discusses various challenges in supporting D2D communications. It then surveys various existing resource allocation schemes to address these challenges. It also provides the achievable performance over different fading channels and computational complexity of the power allocation schemes.

**Keywords** 5G communication · D2D communication · LTE-A · Interference management · Resource allocation

## 15.1 Introduction

Owing to the unprecedented growth in the number of mobile users, the next generation of wireless networks, referred to as the fifth generation (5G), are envisaged to support enormous number of simultaneously connected users with access to numer-

A. Bhardwaj (✉)
School of Computing and Electrical Engineering, IIT Mandi,
Mandi, Himachal Pradesh 175005, India
e-mail: ajay_singh@students.iitmandi.ac.in

D. S. Gurjar
ECE, National Institute of Technology Silchar, Silchar, Assam, India
e-mail: devendra.gurjar@ieee.org



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_15

ous services and applications. The CISCO's latest visual network index (VNI) report forecasts that by 2021, the mobile users traffic will account for 71 % of the total IP traffic, and 82 % of it is video traffic (Index 2019). Media-rich applications such as 3D holography, high-definition video streaming, and cloud-based networking which require high data rate, and delay sensitive applications such as two-way gaming and tactile Internet (e.g., Google glass) (Fettweis 2014) which require a round-trip latency of about 1 ms cannot be supported by existing 4G networks. In summary, providing data at high rate and low latency to high number of users is a daunting challenge facing telecom researchers, service provider, and policymakers.

To cope up with these intense data demands of users, D2D communication, which allows proximate mobile users to communicate, has gained momentum as a prominent solution (Gamage and Shen 2018; Sobhi-Givi et al. 2018). However, despite having significant potential for providing higher throughput (Kai et al. 2019) and lower delay (Mi et al. 2015), implementing D2D communication poses several intrinsic challenges. For example, mutual interference among cellular users (CUs) and D2D users may deteriorate the performance of the network, and signaling overhead to set up D2D links. In order to address these issues, in this chapter, we study the resource allocation schemes where D2D users may share the cellular users channels.

Depending upon their frequency allocation, the operation of D2D communication can be divided into two parts: inband (Feng et al. 2013) and outband (Gui and Zhou 2018) D2D communication. In the inband communication, D2D users are allowed to share the channels allocated to the primary CU users with QoS control from the BS. While in outband communication, D2D users communicate on different frequency bands, such as ISM band, and may operate without any control of the BS. Specifically, inband D2D communication can be realized using two spectrum sharing techniques: underlay (Lee and Lee 2019; Zhang et al. 2018a) and overlay (Kazeminia et al. 2019). In underlay D2D communication, the D2D users are allowed to share the channels allocated to the CUs; however, the interference created to the CUs should be below certain thresholds. While in overlay communication, a dedicated part of the cellular radio resources is allocated to D2D communication, but the resource allocation is supposed to be done efficiently so that dedicated cellular resources are not underutilized (Zhao et al. 2017). The above-mentioned different ways of allocating the frequency channels to the D2D users are depicted in Fig. 15.1.

Recent surveys on D2D communication (Asadi et al. 2014; Mach et al. 2015) evaluate the state-of-the-art research activities and present various research challenges, such as spectrum management, links scheduling, interference handling, and resource allocation. Furthermore, many research works have been done to standardize the integration of D2D communication in LTE networks (Lien et al. 2016). In addition with academia, standardization efforts have been taken by many industries and standardization bodies, such as Qualcomm and 3GPP. The first architecture to implement D2D communication underlaying cellular networks is provided by Qualcomm's FlashLinQ (Wu et al. 2013). The 3GPP provides an analysis on the feasibility of proximity services (ProSe) and proposed the required architectural enhancement to accommodate the D2D use cases in LTE (Lin et al. 2014a). Furthermore, the 3GPP release-12 includes the ProSE as a public safety feature with specific focus on one

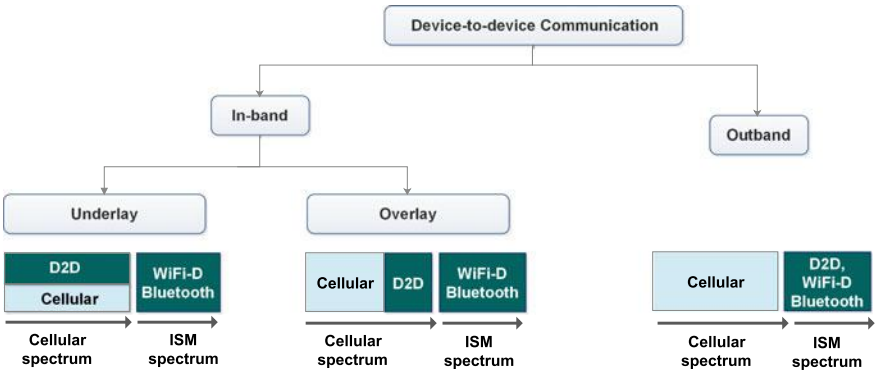



**Fig. 15.1** Different ways to allocate frequency channels to D2D communication

to many communication scenarios (Lin et al. 2014a). In next section, we address the challenges in designing of optimal resource allocation schemes.

**Chapter Outline**: The remaining of this chapter is structured as follows:

- Section 15.2 provides a brief overview of the challenges in the designing of an optimal resource allocation schemes for D2D-enabled cellular networks.
- Section 15.3 provides the classification of the existing resource allocation schemes based on the challenge they are addressing.
- Section 15.4 provides an example of the resource allocation scheme with an objective of sum-throughput maximization.
- Finally, Section 15.5 provides concluding remarks along with some of the possible future works.

The acronyms used in this chapter are listed in Table 15.1.

## 15.2 Challenges in Resource Allocation for D2D Communication

The integration of D2D communication in current LTE network poses many challenges. In the following, we provide a brief overview of the main challenges in designing of the resource allocation schemes for D2D communication in cellular networks.

**Spectrum sharing and interference management**: Due to limited spectrum available for the cellular communication, buying additional licensed band to enable D2D communication is not a feasible solution. Therefore, the allocation of spectrum should be done judiciously, and efficient interference avoidance mechanisms should be provided if the spectrum reuse factor is greater than one. Specifically for the scenarios where D2D users share the channels with cellular users (the underlay approach),

**Table 15.1** List of acronyms

| | |
|---|---|
| ASE | Area spectral efficiency |
| BS | Base station |
| CSI | Channel state information |
| CU | Cellular user |
| DL | Downlink |
| EE | Energy efficiency |
| eNB | Evolved node base station |
| HetNets | Heterogeneous networks |
| ISM band | The industrial, scientific, and medical band |
| LTE | Long-term evolution |
| MINLP | Mixed-integer nonlinear programming |
| PPP | Poisson point process |
| QoS | Quality of service |
| SE | Spectral efficiency |
| SINR | Signal-to-interference-plus-noise ratio |
| SIR | Signal-to-interference ratio |
| UL | Uplink |
| VNI | Visual network index |

providing efficient interference management techniques is of utmost importance. Depending upon the sharing of uplink or downlink CU channels, D2D users create interference to the CUs, or the CUs create interference to the D2D receivers, respectively. This intra-cell interference created to and from the CUs degrades the system throughput. In fact, because of higher chance of presence of proximate D2D links to the CUs, it can severely decrease the performance of the cellular network. Therefore, to enable underlay D2D communication, the optimal resource allocation schemes in presence of interference are needed.

In order to avoid the mutual interference between the CUs and D2D users, overlay communication can be employed (Kazeminia et al. 2019); however, to avoid any wastage of licensed band, the resource allocation should be done efficiently. Besides, in outband D2D communications, the interference is completely eliminated, however, the mobile device should be multi-homed. Further, it decreases the network control over D2D communications, and consequently reliability of the links. In additions, D2D communications need to consider the interference and adapt accordingly to other technologies operating in the same unlicensed band.

**Spectral and energy efficiency trade-off**: Spectral efficiency (SE) which is defined as achievable sum throughout per unit of bandwidth is one of the most prominent parameter in designing of current cellular networks. In D2D integrated cellular networks, SE is majorly increased by supporting more number of D2D users on a single frequency channel. With proliferation of data-hungry applications running on the mobile devices, there is staggering increase in the power consumption of

the wireless networks. Therefore, energy efficiency (EE) which measures how efficiently available transmission power is used for data communication is also becoming an important parameter in cellular network design. Ideally, the resource allocation schemes are needed to be designed in such a way that both the parameters are maximized simultaneously. However, often conflicting nature of these parameters hinders the maximization of both in a network. Therefore, it is imperative to provide resource allocation schemes which provide a good trade-off between spectral efficiency and energy efficiency.

**Optimal performance, computational complexity, and signaling trade-off**: For improving the performance of D2D-enabled cellular networks, resource allocation schemes need to solve the complex optimization problems that are non-convex and combinatorial in nature. In many cases, these optimization problems turn out to be NP-hard, which are intractable in nature, and have exponential time complexity. Thus, they are not manageable for large-scale networks. In addition, optimal solution often leverages the full channel state information (CSI) of all involved links, and in practical large cellular networks providing full CSI to all the users all the time is not a feasible solution. Therefore, the challenges in designing an optimal resource allocation scheme consist of finding a good trade-off between optimality and applicability, centralized versus distributed, and joint versus separate optimization solutions.

## 15.3 Classification of Existing Schemes for D2D Communication

In this section, we classify the existing resource allocation schemes based on the challenge or the objective of the formulated problem in D2D communication they are addressing.

### *15.3.1 Sum-Throughput Maximization*

Most of the previous works on D2D communication formulate a joint channel and power allocation problem with an objective of sum-throughput maximization and ensure a certain level of QoS to the CUs and the D2D users by imposing a SINR threshold on both type of users.

The work of Ye et al. (2014) formulates a sum-throughput maximization problem for D2D-enabled cellular networks. The system model consists of a single BS, with the CUs and D2D users are distributed as a Poisson point process, and it allows multiple D2D pairs to share the channel with a CU. To solve the formulated problem efficiently, authors reformulate it into a Stackelberg game, and a low-complexity low-overhead distributed algorithm is proposed. Through numerical solutions, it is shown that the proposed scheme increases the system throughput and expeditiously

manages the interference created by D2D users to CUs. However, the computational complexity of the proposed iterative algorithm increases exponentially with increase in number of D2D users.

The work of Yin et al. (2015) formulates a joint channel and power allocation problem with an objective of sum-throughput maximization of D2D users. Then, depending upon the availability of CSI at the BS, it provides centralized and distributed solutions for the resource allocation problem. For centralized solution, a successive convex approximation method is proposed, while for distributed solution the formulated problem is remodeled into a Stackelberg game where D2D pairs compete with each other in a noncooperative manner to maximize their individual data rate. The results obtained show that the average system throughput increases significantly with very less signaling overhead in distributed algorithm.

The work of Li et al. (2018) provides a framework to jointly improve the system fairness and throughput in a D2D-enabled underlay cellular network. An optimization problem is formulated to maximize the proportional fairness functions of all the users while considering the fairness and SINR requirement of the CUs and the D2D users. A two-stage joint power control and proportional scheduling algorithm is proposed, which optimally maximize the system throughput along with fairness. However, it only addresses the scenarios where a cellular channel is shared by only one D2D pair which is an underutilization of the spectrum.

The work of Lee and Lee (2019) proposes a joint channel and power allocation problem for maximizing the average achievable rate in a D2D-enabled underlay cellular network. The system model considers more number of D2D pairs than CUs. It derives the expression for outage probability for a CU, and putting it as a constraint, the optimal power allocation problem is solved. Through numerical simulations, it is shown that the provided framework significantly increases the achievable rate of admitted D2D users.

In Cai et al. (2015), authors formulate an overall system capacity maximization problem in a D2D-enabled cellular network. The system model allows multiple D2D pairs to share the single downlink channel. The formulated problem is MINLP, which is NP hard in nature. To solve the problem, authors utilize the graph-coloring approach, where D2D pairs are considered as vertices and the channels of cellular users are considered as a set of colors. The proposed heuristic scheme introduces two areas: interference limited area which identifies those D2D pairs that can transmit on the same channel, and signal-to-interference ratio limited area which identifies those D2D pairs that cannot share the same spectrum. The major limitation is sub-optimal solution at higher computational cost.

In Kai et al. (2019), an optimization problem is formulated with an objective of maximizing the sum-throughput of D2D users while maintaining the certain data rate thresholds to the CUs. The formulated problem turns out to be an instance of MINLP. To solve the formulated problem efficiently, it is decomposed into two subproblems: sub-carrier assignment and power allocation. In sub-carrier assignment problem, it is assumed that the maximum transmit power of a user is divided equally among all the sub-carrier he is assigned. Then, by exploiting the successive convex approximation technique, the power allocation problem is remodeled into sequence

of convex optimization problems. To solve it in an efficient manner, a low-complexity power allocation scheme is proposed. The numerical results show that there is significant gain in sum-rate, however not close to the sum-rate obtained using optimal (exhaustive) approach.

The work of Feng et al. (2013) addresses the problem of mutual interference minimization between CUs and D2D users. It formulates a sum-throughput maximization problem in a D2D-enabled underlay cellular networks where D2D users are sharing the uplink channels with uniformly distributed CUs. The objective of the optimization problem is to allocate the channels and powers to the CUs and the D2D users in order to maximize the sum-throughput of the system, as shown by the Eq. (15.1).

$$\max_{\rho, P_i^c, P_j^d} \left\{ \sum_{i \in C} \sum_{j \in D} [\log_2(1 + \gamma_i^c) + \rho_{i,j} \log_2(1 + \gamma_j^d)] \right\}, \tag{15.1}$$

where $\gamma_i^c$ and $\gamma_j^d$ denotes the SINR for the CUs and D2D users, respectively; $\rho_{i,j} \in \rho$ is a binary variable, and it indicates whether the channel of the $i$th CU is shared by the $j$th D2D user or not; variables $C$ and $D$ denote the number of the CUs and D2D users in the cell, respectively. This work also ensures the QoS of both type of users by maintaining SNR above certain thresholds. Since the above problem is again an instance of MINLP, which is hard to solve optimally, authors propose a three step solution: (1) QoS-aware admission control to decide whether a D2D pair is admissible or not, (2) power allocation to admissible D2D users, and (3) channel allocation to a D2D user by utilizing bipartite graph matching technique. Using numerical simulations, it is shown that the sum-throughput and number of D2D users admitted in a cell increases significantly. It compares their proposed scheme with other heuristic (Zulhasnine et al. 2010) and fixed margin scheme (Janis et al. 2009).

Figure 15.2 depicts that the sum-throughput decreases with increase in D2D geographical spread. The reason is as D2D receivers are far from their corresponding D2D transmitters, thus they have low SINR, and consequently lower contribution to the sum-throughput. It can also be observed that the gain in the sum-throughput increases if the channels are considered faded, and the resource allocation schemes are designed accordingly.

In addition to above-mentioned works, there exist many research works which solve the similar objective optimization problem with different approaches. Table 15.2 provides a brief summary of such works.

### *15.3.2 Spectral Efficiency Maximization*

Some of the previous works address the problem of maximizing the spectral efficiency of the D2D-enabled cellular networks, as listed in Table 15.3. These works either tries to increase the number of users sharing the channels, or increase the data transmission rate per unit of bandwidth by adequately managing the interference.

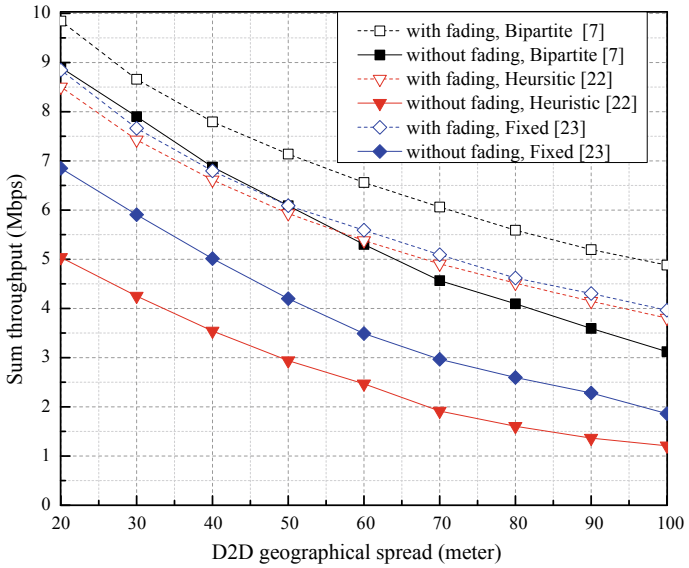



**Fig. 15.2** Sum-throughput variation with D2D geographical spread for different power allocation schemes with impact of fading

**Table 15.2** Resource allocation schemes for D2D communications that are addressing sum-throughput maximization, $N_c$ is numbers of CUs, and $N_D$ is number of D2D users

| References | Spectrum share | Channel sharing direction | Complexity | Analytical tool |
|---|---|---|---|---|
| Ye et al. (2014) | Underlay | UL | $O(N_D \log_2(\mu_{\max}/\epsilon) + N_D^2 K)$, $K \in [4, 8]$, $\log_2(\mu_{\max}/\epsilon) \in [5, 10]$ | Game theory |
| Yin et al. (2015) | Underlay | UL | | Non-convex optimization and Game theory |
| Li et al. (2018) | Underlay | UL | $O(T * N_c^3)$, T is number of time slots | Graph theory (Hungarian algorithm) |
| Feng et al. (2013) | Underlay | UL | $O(N_c^3)$, $N_c = N_D$ | Graph theory |
| Lee and Lee (2019) | Underlay | UL | non-trivial to determine | Exhaustive search, Heuristic scheme |
| Cai et al. (2015) | Underlay | DL | $O(2^{N_D} N_D)$, | Graph coloring theory |
| Kai et al. (2019) | Underlay | UL | $O((2N_D + N_c)K)$, K denote number of channels | Convex optimization |

The work of Mustafa et al. (2014) proposes a three-tier network as a hierarchical heterogeneous networks (HetNet), where D2D communication is enabled as tier-3 network within macro-cell BS (tier-1) and small-cell BS (tier-2). This framework increases the spectral efficiency in terms of percentage of mobile users engage in D2D communication instead of tier-1 and tier-2. The numerical results show that there is significant increase in achievable capacity by hierarchical HetNet in comparison with traditional HetNet where D2D communication is not enabled. It also shows that spectral efficiency increases for D2D-enabled ultra dense networks. The major limitation of this work is the signaling overhead to set up D2D links in HetNets.

Authors in Kim et al. (2018) formulate a joint channel and power allocation problem with an objective of maximizing the spectral efficiency in a D2D-enabled underlay cellular networks. By exploiting the inner approximation method, a low-complexity iterative algorithm is proposed. Using numerical results, it is shown that with increase in transmit power of a D2D user, the average spectral efficiency increases upto a certain level, then it saturates. The limitation of this work is that it always assumes that a channel is shared by maximum one D2D pair, which is an inefficient utilization of spectrum (Table 15.3).

The work of Trigui and Affes (2018) provides an analytical framework to evaluate the CUs and D2D users SINR distributions in general fading scenario. It unifies various fading models by utilizing the H-transform theory. An average area spectral efficiency (ASE) maximization problem is formulated, and the expression for the access probability and the optimal transmit power which maximizes the ASE is derived. Numerical results show that the proposed opportunistic access scheme requires only statistical CSI instead of complete CSI for centralized solution, which leads to less delay in the network. It is also shown that the proposed framework handled all type of complex fading models, such as Weibull, shadowed $\kappa - \mu$.

**Table 15.3** Resource allocation schemes for D2D communications that are addressing spectral efficiency maximization, $N_c$ is numbers of CUs, and $N_D$ is number of D2D users

| References | Spectrum share | Channel sharing direction | Complexity | Analytical tool |
|---|---|---|---|---|
| Mustafa et al. (2014) | Overlay and outband | DL/UL | Non-trivial to determine | Heuristic schemes |
| Kim et al. (2018) | Underlay | UL | $2(N_c * N_D)$ | Heuristic and inner approx. |
| Trigui and Affes (2018) | Underlay | UL | $O(T * N_c^3)$, $T$ is number of time slots | Hungarian algorithm |

### *15.3.3 Energy Efficiency Maximization*

Many research efforts have been put to address the problem of energy efficiency maximization both in cellular networks and D2D-enabled cellular networks. The optimization problem is formulated either to decrease the energy consumption of an individual transmitter or as whole of the system.

The work of Wu et al. (2015) formulates an optimization problem with an objective of maximizing the energy efficiency of D2D users with constraints on minimum data rate supported to each CU and D2D user. To make the system model more practical, it considers circuit power consumption alongside transmit power consumption. Depending upon the D2D users circuit power consumption, three different operative regions are identified, and corresponding power control mechanisms are provided. To make the system more computationally efficient, a distributed power control algorithm is proposed. The numerical results show that the energy efficiency increases significantly by controlling the transmit power of D2D users.

Authors in Della Penda et al. (2016) formulate an energy consumption problem in D2D-enabled cellular networks operating in dynamic time division duplex (TDD) mode. The formulated optimization framework jointly optimizes the mode selection, uplink/downlink transmission period, and transmission power to minimize the energy consumption. It is MINLP, therefore, heuristic schemes are proposed which provide the near-optimal solutions. The results obtained show that significant energy saving can be achieved by synergy of better channel gains of D2D links and the adaptive transmission time of dynamic TDD.

In Zhao et al. (2016), the authors formulate a mixed-integer optimization problem with an objective of minimizing the energy consumption in D2D underlay cellular networks. A more realistic scenario is considered where multiple D2D pairs are sharing the channel with cellular users, and they are moving within cells. By exploiting the dynamic graph modeling technique, theoretical lower bound on system energy consumption is derived, and it is revealed that adapting underlay D2D communication with proper transmission power control significantly decreases the energy consumption. Effect of various network parameters, such as flow delay, buffer size, and bandwidth is also shown.

In Khazali et al. (2018), the authors formulate an energy efficiency maximization problem for D2D users in a D2D-and-femtocell-enabled HetNets. They utilize the frequency fractional reuse structure to allocate the channels to the CUs and D2D users and show that it minimizes the interference among the simultaneously operating technologies. To solve the problem in an efficient manner, they remodeled the problem as a noncooperative game and proposed an iterative algorithm to allocate transmit powers to users. The results obtained show that by controlling the transmit power of CUs and D2D users, the energy consumption can be reduced significantly without much decrease in number of the CUs and the femtocell users. However, the complexity of the proposed algorithm increases exponentially with increase in number of CUs and D2D users.

The work of Jiang et al. (2015) formulates an energy efficiency maximization problem in a D2D-enabled cellular network, and it turns out to be an MINLP, which is NP hard and non-convex in nature. To find a tractable solution, it is then transformed into a fractional programming problem and an equivalent subtractive form. A two-layer iterative power and channel allocation scheme is proposed which converges fast to the optimal solution. The objective function of the energy efficiency maximization problems is expressed as follows:

$$\max_{\rho_{i,j}, P_{i,j}} U_{\text{EE}} \tag{15.2}$$

$$U_{\text{EE}} = \frac{\sum_{i=1}^{D} \sum_{j=1}^{C} \rho_{i,j} R(P_{i,j})}{\sum_{i=1}^{D} \sum_{j=1}^{C} \rho_{i,j} P_{i,j} + P_C}, \tag{15.3}$$

where $\rho_{i,j}$ is a binary variable, with value $\rho_{i,j} = 1$ denotes that the channel of the $i$th CU is shared with the $j$th D2D user, and $\rho_{i,j} = 0$ otherwise; $R(P_{i,j})$ denotes the sum-rate achieved with transmit power $P_{i,j}$; $P_C$ and $P_{i,j}$ denote the maximum transmit power transmitted by the CU and the D2D user, respectively. Numerical results confirm that the proposed model significantly increases the EE even for larger number of D2D pairs in the cell. A list of such works is provided in Table Table 15.4.

**Table 15.4** Resource allocation schemes for D2D communications that are addressing energy efficiency maximization, $N_c$ is numbers of CUs, and $N_D$ is number of D2D users

| References | Spectrum share | Channel sharing direction | Complexity | Analytical tool |
|---|---|---|---|---|
| Wu et al. (2015) | Underlay | UL | $\log_2(\frac{\overline{M}-\underline{M}}{\epsilon_2}) \times 2K \left[\log_2(\frac{\overline{Q}-\underline{Q}}{\epsilon_1})\right]^2$, $\overline{M}-\underline{M}$, $\overline{Q}-\underline{Q}$ are feasible power region for CU and D2D user, $\epsilon_1$ and $\epsilon_2$ are error tolerance | Heuristic approach (Iterative algorithm) |
| Della Penda et al. (2016) | Underlay | UL/DL | Non-trivial to determine | Heuristic approach |
| Zhao et al. (2016) | Underlay | DL | $(O(B+R)N)$, $B$ = no. of BS, $R$ = no. of relays, $N$ = no. of time slots | Graph theory |
| Khazali et al. (2018) | Overlay | UL | Iterative algorithm with $I_{\max}$ iterations | Game theory |
| Jiang et al. (2015) | Underlay | UL | $O(N_D N_C)$ | Fractional programming |

### 15.3.4 Energy Efficiency and Spectral Efficiency Trade-Off

In general, the performance of a communication system is measured by three performance parameters: energy efficiency (EE), spectral efficiency (SE), and QoS. Ideally, it is desired that all the parameters are maximized; however, in practical networks, these parameters are not independent to each other, sometimes improvement of one parameters may have negative impact on another one. Previously, many resource allocation schemes have been proposed which find a good trade-off between SE and EE in D2D-enabled cellular networks.

Authors in Zhou et al. (2014) formulate an energy efficiency maximization problem with constraint on the SE requirement and maximum transmission power in a D2D-enabled underlay cellular networks. To closely relate to the practical networks, the circuit power consumption is also considered. The objective and constraints of the formulated general EE-SE trade-off problem are as follows:

$$\max U^d_{i,\mathrm{EE}}, = \max \left( \frac{\sum_{k=1}^{K} \log_2 \left( 1 + \frac{p_i^k g_i^k}{p_c^k g_{c,i}^k + \sum_{j=1, j=i}^{N} p_j^k g_{j,i}^k + N_0} \right)}{\sum_{k=1}^{K} \frac{1}{\eta} p_i^k + 2 p_{cir}} \right) \tag{15.4}$$

$$C_1 : U^d_{i,\mathrm{SE}} \geq R^d_{i,\min}, \tag{15.5}$$

$$C_2 : 0 \leq \sum_{k=1}^{K} p_i^k \leq p^d_{i,\max}, \tag{15.6}$$

where $U^d_{i,\mathrm{SE}}$ is the SE (bits/s/Hz) of $i$th D2D user; $g^k_{c,i}$ is the interference channel gain between the $k$th CU to the $i$th D2D receiver; $\eta$ denotes the efficiency of power amplifier; $g^k_{j,i}$ denotes the interference link from the $j$th D2D pair transmitter to the $i$th D2D pair receivers; $K$ denotes the total number of channels. To solve the problem efficiently, a resource allocation scheme based on the noncooperative game is proposed, where each user behaves in a greedy manner and tries to maximize the individual EE. By using numerical simulations, it is shown that for high values of transmission power, there is little gain in SE, while significant loss in EE. The limitation of this work is that the computational complexity to find trade-off increases significantly with increase in number of D2D users and CUs. Through numerical simulations, it is shown that for increasing value of SE, EE has typical bell-shaped characteristic, an initial increase in EE, then rapid decrease. Indeed this trend is expected for EE, because to fulfill high requirement of SE, more power is allocated to D2D transmitters, thus high power consumption, and more interference to CUs. Thus, less increase in sum-throughput, and more increase in power consumption, so overall decrease in EE.

The work of Hao et al. (2018) formulates a multi-objective optimization problem to investigate the trade-off between EE and SE in a D2D-enabled underlay cellular networks. It addresses two scenarios: first, channel state information (CSI) is available at the BS, second, CSI is not available. By exploiting the $\epsilon$-constraint and

strict robustness, the formulated problem is transformed into a deterministic single-objective optimization problem. However, finding an optimal solution requires exponential computational power. To decrease it, a two-stage iterative algorithm is proposed which solves the channel and power allocation problem separately. Numerical results demonstrate that there exist an intrinsic relationship between EE-SE performance of CUs and the minimum rate requirement of D2D pairs. A major limitation of these approaches is that they are iterative in nature; therefore, convergence of the proposed algorithm is not guaranteed. It highly depends upon the number of D2D pairs per channel.

The work of Zhang et al. (2018b) investigates the EE-SE trade-off for dynamic D2D communications in underlay cellular networks. The mobile device is assumed to be moving, and to capture the dynamic channel characteristics vehicle-to-vehicle channel model is utilized. Depending upon the vehicular traffic density (VTD), whether high or low, it formulates two optimization problems. Using numerical simulations, it is shown that, for high VTD, a small decrease in EE around its peak value leads to significant gain in SE, while in low VTD scenarios, a slight decrease in SE results in higher gain of EE. Moreover, to improve the practicability of the proposed solution, a more general performance metric "economic efficiency" is employed, which measures profitability in terms of monetary unit per second. The limitation of this work is that it only considers that a CU channel is shared by only one D2D pair, which is underutilization of spectrum.

In Wei et al. (2015), authors analyze the EE-SE trade-off for the multihop D2D communications in the cellular networks. They specifically addressed the scenario where a D2D user may help other two D2D users to exchange information. The proposed work compares the EE and SE performance of the possible three communication modes: (1) when there is direct D2D communication between the transmitting and receiving node, (2) multihop D2D communication through another user, and (3) conventional communication through centralized BS. The presented work also finds the optimal transmit power of D2D users that maximize the EE of each mode. By using numerical simulations, it is shown that for larger transmission distance, the EE of multihop D2D communication is better, while for smaller transmission distance, the SE of multihop is better in comparison with direct and conventional communications. However, orthogonal channel allocation is assumed for D2D communications which may lead to an inefficient utilization of available spectrum. Table 15.5 provides a brief summary of such works.

### *15.3.5 Delay Minimization*

To support the real-time applications, many research efforts have been done to exploit D2D communications for relay communications. Researchers try to formulate optimization problems with objective of minimizing the delay or number of hops in relay communications, while ensuring a certain level of QoS to CUs and D2D users.

**Table 15.5** Resource allocation schemes for D2D communications that are addressing energy-spectral efficiency trade-off, $N_c$ is numbers of CUs, and $N_D$ is number of D2D users

| References | Spectrum share | Channel sharing direction | Complexity | Analytical tool |
|---|---|---|---|---|
| Zhou et al. (2014) | Underlay | UL | $O(I^d_{i,\text{dual}} I^d_{i,\text{loop}} K)$, $I^d_{i,\text{dual}}$ and $I^d_{i,\text{loop}}$ are the required number of iterations required for convergence, and for solving dual problem, $K$ is number of CUs | Game theory |
| Hao et al. (2018) | Underlay | UL | $O(K\,Q\,(N_D - 1)\,(N_C + K)^{\mu}(N_C N_D + K)^{\nu})$ | Heuristic (Pareto optimality) |
| Zhang et al. (2018b) | Underlay | UL | Predefined $I_{\max}$ number of iterations for convergence | Heuristic approach |
| Wei et al. (2015) | Overlay | UL | Non-trivial to determine | Network coding and heuristic approach |

The recent work by Huang et al. (2019) formulates a joint channel and power allocation problem for delay-aware D2D communication. The system model consists of D2D users sharing uplink channels with CUs, with a CU channel is allowed to share by maximum one D2D pair. By utilizing Lyapunov optimization framework, an optimal delay-aware traffic admission, mode selection and resource allocation (DTMR) scheme is proposed. The results obtained show that the proposed scheme reduces the queue length of every user, and increases average per-user data rate.

Mi et al. (2015), formulate an optimization problem with an objective of minimizing the average delay and average drop rate in a D2D-enabled underlay cellular network. The formulated problem is remodeled into an infinite horizon average cost Markov decision process (MDP). To solve the problem efficiently, a two-stage channel and power allocation scheme is proposed. In the first stage, a heuristic approach is proposed which considers the imbalance of queue lengths, and in the second stage, the transition probabilities in the system states are derived. To allocate the optimal power to D2D transmitters, the *Bellmen equation* is solved. By using numerical simulations, it is shown that the proposed scheme decreases the average packet delay and average drop rate in comparison with baseline models, such as fixed power and round robin. It is further extended in Lei et al. (2015) where authors consider dynamic data arrival, and reformulate the resource allocation problem into an infinite horizon average reward constrained MDP. It is shown that the delay can be minimized either by a deterministic policy or a simple mixed policy which randomize between two deterministic policy.

Kazeminia et al. (2019) formulate a dynamic spectrum sharing and power control problem in a D2D-enabled overlay cellular networks. To ensure the average queue

length should be below certain threshold, a constraint on maximum delay is considered. Again by utilizing Lyapunov optimization technique, the formulated problem is transformed into a maximum weighted sum-rate problem with the average power consumption as a penalty term. To solve the optimization problem efficiently and priorly to the next scheduling frame, a low-complexity heuristic scheme is proposed. By using numerical simulations, it is shown that the dynamic spectrum sharing approach significantly reduces the queuing delay with the same power consumption of the static approach. However, overlay communication is assumed, which may lead to an underutilization of available spectrum. A list of previous works addressing delay minimization is provided in Table 15.6.

## 15.4 Resource Allocation for Sum-Throughput Maximization

In this section, we formulate a sum-throughput maximization problem in a D2D-enabled underlay cellular networks and proposed a channel and power allocation problem to efficiently solve the problem.

### *15.4.1 System Model*

We envisioned a D2D-enabled underlay cellular network of cell radius $R$, as depicted in Fig. 15.3, in which the locations of the CUs are assumed to be uniformly distributed. The D2D users are assumed to follow the Poisson point process (PPP) $\Pi_d$, and the node density is $\lambda_d$. The CUs and D2D users are assumed to share the same uplink channels. The total number of uplink channels are denoted as $\mathcal{K} = \{1, 2, \ldots, N_c\}$. To generalize, we use the common index for CU and orthogonal channels as $1, 2, \ldots, N_c$. To characterize the channel between the $i$th transmitter and the $j$th receiver, i.e., $h_{ij}$, Nakagami fading model is assumed. In uplink channel sharing mode, let $k$th channel is used by the $k$th CU and $\mathcal{D}_k \subseteq \mathcal{D}$ D2D transmitters. It is assumed that all the CUs transmit at orthogonal channels, and this helps in avoiding severe co-channel interference among CUs.

### *15.4.2 Problem Formulation*

The achievable rate of the $k$th CU sharing channel with other D2D pairs which are transmitting on the $k$th channel is expressed as follows:

**Table 15.6** Resource allocation schemes for D2D communications that are addressing delay minimization, $N_c$ is numbers of CUs, and $N_D$ is number of D2D users

| References | Spectrum share | Channel sharing direction | Complexity | Analytical tool |
|---|---|---|---|---|
| Huang et al. (2019) | Overlay | UL | $O((N_c + 3N_D)K)$, K denotes number of resource blocks | Lyapunov optimization |
| Mi et al. (2015) | Underlay | UL | $O(N_c^2(N_c + p))$, $p$ denotes number of possible control inputs | Heuristic scheme |
| Kazeminia et al. (2019) | Overlay | UL | $O(N_{\text{RB}}^2 L_D^6 L \log N_{\text{RB}})$, $N_{\text{RB}}$ is number of RBs; $L_C$ and $L_D$ denote the number of CU links and D2D links, respectively; $L = L_C + L_D$ and $L \log N_{RB}$ is a function of rate of increase of iteration | Heuristic approach |
| Lei et al. (2015) | Underlay | UL/DL | $O(\lvert\mathcal{S}\rvert^2 \lvert\mathcal{A}_x\rvert)$, $\mathcal{S}$ and $\mathcal{A}_x$ denote state space and auction space, respectively | Dynamic optimization |

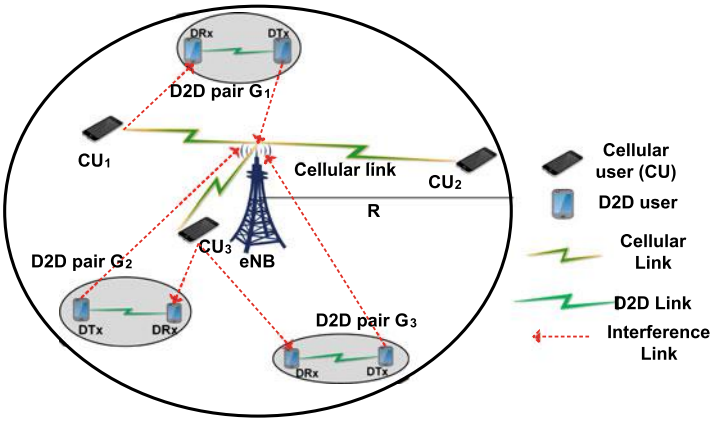



**Fig. 15.3** Depiction of a D2D-enabled underlay cellular networks, where D2D pair $G_1$ reuses channel with $\text{CU}_1$, and D2D pair $G_2$ and D2D pair $G_3$ reuse uplink channel with $\text{CU}_3$

$$R_{k,g}^{c} = B_k \log_2 \left(1 + \frac{p_{c,k} h_{c,b}^{k} d_{cb}^{-\alpha}}{\sum_{g=1}^{|\mathcal{D}_k|} p_{g,k} h_{g,b}^{k} d_{gb}^{-\alpha} + N_0}\right), \tag{15.7}$$

where $h_{c,b}^{k}$ and $h_{g,b}^{k}$ represent the channel gain from the $k$th CU and $g$th D2D transmitter to the BS, respectively; $p_{c,k}$ and $p_{g,k}$ denote the transmit power of the $k$th CU and the $g$th D2D pair transmitter; $d_{c,b}$ and $d_{g,b}$ represent the distance between the CU and the D2D transmitter to the base station, respectively; $\alpha$ is path-loss exponent; $N_0$ denotes the variance of additive white Gaussian noise (AWGN) noise. Similarly, the achievable SINR of the $g$th D2D pair sharing channel with other D2D pairs and the CU which are transmitting on the $k$th channel is as follows:

$$\Gamma_{r,g}^{k} = \frac{p_{d,k} h_{g,r}^{k} d_{gr}^{-\alpha}}{\sum_{j=1, j\neq g}^{|\mathcal{D}_k|} p_{j,k} h_{j,r}^{k} d_{jr}^{-\alpha} + p_{c,k} h_{c,r}^{k} d_{cr}^{-\alpha} + N_0}. \tag{15.8}$$

In (15.8), $h_{j,r}^{k}$ is channel gain from the co-channel D2D transmitter to the receiver of other co-channel D2D pairs. The achievable rate of the $g$th D2D pair is

$$R_{k,g}^{D} = B_k \log_2 \left(1 + \Gamma_{r,g}^{k}\right). \tag{15.9}$$

To determine the sum-throughput gain achieved by sharing the channels, we need to find the sum-throughput when channels are not shared. Therefore, the achievable data rate of the $k$th CU in absence of interference created by D2D links is determined as:

$$\widehat{R_k^c} = B_k \log_2 \left(1 + \frac{p_{c,k} h_{c,b}^{k} d_{cb}^{-\alpha}}{N_0}\right). \tag{15.10}$$

From (15.7) and (15.10), it can be deduced that the data rate of the $k$th CU decreases, and the reason for this decrement is interference created by the co-channel $|\mathcal{D}_k|$ D2D transmitters. The value of decrement is calculated by (15.11).

$$\Delta R_{k,g}^{c} = \widehat{R_k^c} - R_{k,g}^{c}. \tag{15.11}$$

Similarly, the gain in achievable rate after channel sharing can be determined as follows:

$$\Delta R_{k,g} = \sum_{g\in\mathcal{D}_k} R_{k,g}^{D} + R_{k,g}^{c} - \widehat{R_k^c} = \sum_{g\in\mathcal{D}_k} R_{k,g}^{D} - \Delta R_{k,g}^{c} \tag{15.12}$$

The lowest bound on the achievable throughput gain can be calculated by assuming all the D2D transmitters transmit at maximum power, because this created maximum co-channel interference.

Let $R_c$ and $R_d$ denote the achievable sum-throughput of all the CUs and D2D pairs, respectively, and can be derived using following expressions:

$$R_c = \sum_{k=1}^{K} R_{k,g}^{c} \tag{15.13}$$

$$R_d = \sum_{k=1}^{K} a_{g,k} R_{k,g}^{D}, \tag{15.14}$$

where $a_{g,k}$ denotes the binary variable which determines whether a channel is shared or not. $a_{g,k} = 1$ denotes that the $k$th is shared, otherwise not. In order to increase the sum-throughput which is summation of the achievable rate of CUs and D2D users, our objective is to maximize the sum-throughput of all users while maintaining SNR's above thresholds. Thus, the optimization problem can be defined as

$$\begin{aligned} \mathbf{P}_1 : \max_{p_{d,k}, p_{c,k}, a_{g,k}} & \left(R_c + R_g\right) \\ \text{subject to} \quad & C_1 : 0 \le p_{c,k} \le P_c^{\max} \quad \forall c \in C \\ & C_2 : 0 \le p_{d,k} \le P_d^{\max} \quad \forall k \in \mathcal{K}, g \in \mathcal{D}_k \\ & C_3 : \Gamma_c^k \ge \gamma_{k,\text{th}} \\ & C_4 : \Gamma_{r,g}^k \ge \gamma_{g,\text{th}} \\ & C_5 : a_{g,k} \in \{0, 1\}, g \in \mathcal{D}, k \in C. \end{aligned} \tag{15.15}$$

Constraint $C_1$ denotes the maximum power that can be allowed to a CU; constraint $C_2$ indicates the maximum power that can be allowed to a D2D transmitter. Constraints $C_3$ and $C_4$ determine the minimum SINR level should be maintained above certain threshold, and constraint $C_5$ denotes the binary channel indication variable. By seeing the structure of the formulated problem, it is deduced that it is an MINLP, and solving an MINLP problem is computationally hard; therefore, we provide a heuristic approach which solves the problem $\mathbf{P}_1$ in polynomial time.

### *15.4.3 Outage Probability Analysis*

This subsection derives the closed-foam expressions for the outage probability (OPs) experienced at a typical receiver of a D2D pair. Outage occurs when the SINR at the receiver is less than a decoding threshold Lin et al. (2014b). Here, we consider a D2D pair is under outage when the SINR at its receiver falls below the decoding threshold. As per our system model, the SINR at the $r$th receiver of a D2D pair on the $k$th channel is given in (15.8). For initial outage calculation, we utilize the full channel inversion for uplink communication ElSawy et al. (2014); therefore, the transmit power by D2D-Tx and CU is $\rho_d d_{g,r}^{\alpha}$ and $\rho_{bs} d_{cb}^{\alpha}$, respectively, where $\rho_d$ and $\rho_{bs}$ denote the receiver sensitivity of D2D receiver and BS, respectively. We focus on spectrum sharing which means the hybrid network is interference limited and the impact of thermal noise ($N_0$) can be neglected. Hence, SIR is assumed in place of

SINR, and (15.8) is simplified to

$$\mathrm{SIR}_k^r = \frac{h_{g,r}^k \rho_d}{I_k^r}, \tag{15.16}$$

where $I_k^r$ ($I_k^r = \sum_{j=1, j\neq g}^{|\mathcal{D}_k|} p_{j,k} h_{j,r}^k d_{jr}^{-\alpha} + p_{c,k} h_{c,r}^k d_{cr}^{-\alpha}$) denotes the total interference at the $r$th D2D receiver on the $k$th channel. The probability that SIR of a typical D2D receiver is less than a threshold $\gamma_{\mathrm{th}}$, i.e, outage probability can be expressed as follows:

$$P(\mathrm{SIR}_k^r < \gamma_{\mathrm{th}}) = E_{I_k^r, h_{g,r}^k} \left\{ P\left( \frac{h_{g,r}^k \rho_d}{I_k^r} \right) < \gamma_{\mathrm{th}} \right\}, \tag{15.17}$$

where $E_{I_k^r, h_{g,r}^k}\{.\}$ stands for the expectation with respect to $I_k^r$ and $h_{g,r}^k$. By following the power gain-based reference link and assuming the reference link suffering from Nakagami fading model with shape parameter $m$ as in Guo et al. (2014, 2016), (15.17) can be written as:

$$P_{\mathrm{out}}^k(\gamma_{th}) = 1 - \sum_{l=0}^{m-1} \frac{-s^l}{l!} \frac{\mathrm{d}^l}{\mathrm{d}s^l} \mathbb{M}_{I_k^r(s)}|_{s=m\frac{\gamma_{th}}{\rho_d}}, \tag{15.18}$$

where $\mathbb{M}_{I_k^r(s)} = E_{I_k^r}\left[\exp\left(-sI_k^j\right)\right]$ is the moment generating function (MGF) of $I_k^r$. An advantage of considering Nakagami-m model is that it covers all the fading models; therefore, the expression for OP in Rayleigh fading can also be derived (by placing $m = 1$). Next, we find the MGF of interference created at the $r$th receiver.

We condition on the $r$th D2D-Rx which is located at a distance $d$ from the base station. The isotropic network region and PPPs rotation-invariant property says that OP derived at the $r$th receiver is the same for those receivers which are at the same distance $d$ from the base station. Therefore, according to Silvnyak's theorem, Daley and Vere-Jones (2007), the MGF of the $I_k^r$ can be derived as follows:

$$\begin{aligned}
\mathbb{M}_{I_k^r}(s, d) &= E_{I_c^r}\{\exp\left(-sI_c^r\right)\} \\
&\quad \times E_{g_k \in G_k\, g}\left\{ \exp\left( -s \sum_{G_k\, g} I_k^r(r) \right) \right\} \\
&= \mathbb{M}_{I_c}^r(s, d)\, E_{G_k}\left\{ \exp\left( -s \sum_{g_k \in G_k} I_k^r(r) \right) \right\} \\
&= \mathbb{M}_{I_c}^r(s, d) \exp\left(\lambda_d |\mathcal{A}| \left(\mathbb{M}_{I^r}(s, d) - 1\right)\right),
\end{aligned} \tag{15.19}$$

where $\mathbb{M}_{I^r}(s,d)$ denotes the MGF of interference from other D2D pairs which communicate using the $k$th channel. $\mathbb{M}_{I_c^r}(s,d)$ denotes MGF of interference created from shared CU.

### 15.4.4 Channel Allocation Algorithm

Based on the expression of OPs, the BS may easily make a decision whether two D2D pairs are allowed to share resource with a CU or not. If they are allowed to share resources with a CU, we may have more options for selecting a CU such as:

Objective 1: Choose that CU that minimizes the outage probability of a particular D2D pair, either $G_1$ or $G_2$.

$$i^* = \underset{i \in C}{\operatorname{argmin}} P_{\text{out}}^{G_1}, \text{ or } i^* = \underset{i \in C}{\arg\min} P_{\text{out}}^{G_2} \tag{15.20}$$

Objective 2: Choose that $CU_i$ that minimizes the maximum of OPs of two D2D pairs.

$$i^* = \underset{i \in C}{\operatorname{argmin}} \max \left( P_{\text{out}}^{G_1}, P_{\text{out}}^{G_2} \right) \tag{15.21}$$

Objective 3: Choose that CU that minimizes the sum of OP of two D2D pairs.

$$i^* = \underset{i \in C}{\operatorname{argmin}} \left( P_{\text{out}}^{G_1} + P_{\text{out}}^{G_2} \right) \tag{15.22}$$

Depending on the requirement of service on each D2D pair, the base station would use the optimal objective for selecting the CU. For example, if D2D pair 1 needs higher QoS than D2D pair 2, the base station opts for Objective 1 to guarantee performance for D2D pair 1. Objective 2 tries to minimize the maximum of outage probabilities of the two D2D pairs, thus tries to maintain the SINR of both pairs above some threshold value. Objective 3 tries to minimize the sum of outage probabilities of both D2D pairs.

### 15.4.5 Power Allocation

After the channel allocation step, the binary variable $a_{g,k}$ vanishes. Therefore, we need to only allocate the transit power to the D2D transmitters such that they maximizes the sum-throughput. For this step, we utilized "Generalized Distributed Constrained Power Control Algorithm (GDCPC) (Im et al. 2008)." The algorithm fulfills the SIR requirement of every D2D user while limiting the interference to other receiving nodes upto a tolerable level. The algorithm ensures the interference constraint and concomitantly QoS constraint by distributing the total interference to CU into

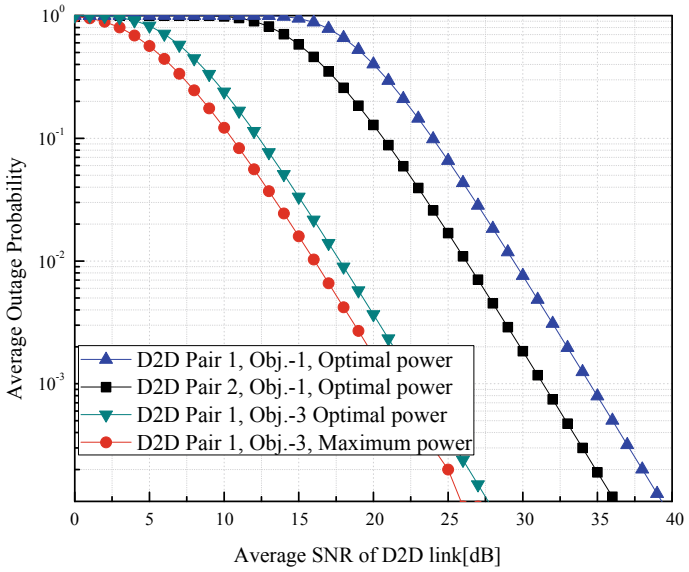



**Fig. 15.4** Outage probability variation with average SNR of the D2D multicast group, CU = 5, $M = 10$

the individual constraint of D2D transmitter, and it can be written as

$$h_{g,b}^{k} p_{g,k} \leq \frac{I_{th}^{k}}{G_k} \quad \forall g \in G_k \tag{15.23}$$

By rearranging (15.23), we derive the upper bound on the maximum power of a D2D transmitter.

$$p_{g,k} = \frac{I_{th}^{k}}{h_{g,b}^{k} G_k} \tag{15.24}$$

It may be noted that to calculate the $p_{g,k}$, we need to know the other co-channel D2D pair, and it can be obtained from the channel allocation step. Also, this may be obtained from the beacon signals that are transmitted by the BS. Therefore, maximum transmission power of a D2D transmitter is derived as

$$p_{g,k}^{\max} = \min\{P_{g,D}^{\max}, \frac{I_{\text{th}}}{h_{g,b}^{k} G_k}\} \tag{15.25}$$

where $P_{g,D}^{\max}$ is the maximum transmission power that a mobile device may radiate, that is 30 dBm.

#### 15.4.5.1 Effect of SNR on the Outage Probabilities of D2D Pairs

Figure 15.4 depicts the outage probability variation with average SNR of D2D links. It is assumed that a cellular channel is shared by maximum of two D2D links. Two observations can be drawn from this figure. First, for objective 1 where the CU for resource sharing is chosen such that it minimizes the OP of D2D pair 1 instead of D2D pair 2, with optimal power allocation, the gap between their respective achievable SNR is nearly 2.5 dB. Second observation is impact of power allocation on OPs. When selection criteria is objective 3, the gap between their achievable SNR with the proposed and maximum power allocation is nearly 3 dB. This shows the efficacy of proposed algorithm in decreasing the outage probability of D2D pairs.

## 15.5 Conclusion and Future Works

This chapter begins by a brief introduction to the potential usage and different modes of operation of D2D communication in next generation of wireless networks. Then, to seamlessly integrate the D2D communication in cellular networks, it presents the challenges in designing the optimal resource allocation schemes. Next, it provides the classification of the existing resource allocation schemes based on the challenges they are addressing. Then, a sum-throughput maximization problem is formulated, and to solve that a heuristic scheme is proposed. The results obtained with presented resource allocation schemes show the potential of D2D communication to provide higher data rate, lower end-to-end latency, and improved spectral and energy efficiencies. These results may further lead to integration of this technology in the next generation of wireless networks by helping with the design and analysis of relevant D2D resource allocation schemes. Further, most of the existing schemes deal with the sum-throughput maximization; however, an interesting extension would be to provide resource allocation scheme considering various utility maximization problem for D2D-enabled cellular networks. Also, the insights so obtained with centralized resource allocation schemes will be useful for designing fast and more pragmatic distributed resource allocation schemes for D2D-enabled cellular networks.

## References

Asadi A, Wang Q, Mancuso V (2014) A survey on device-to-device communication in cellular networks. IEEE Commun Surv Tutorials 16(4):1801–1819

Cai X, Zheng J, Zhang Y (2015) A graph-coloring based resource allocation algorithm for D2D communication in cellular networks. In: Proceedings of IEEE ICC, London, pp 5429–5434

Daley DJ, Vere-Jones D (2007) An introduction to the theory of point processes: volume II: general theory and structure. Springer Science & Business Media

Della Penda D, Fu L, Johansson M (2016) Energy efficient D2D communications in dynamic TDD systems. IEEE Trans Commun 65(3):1260–1273

ElSawy H, Hossain E, Alouini MS (2014) Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks. IEEE Trans Commun 62(11):4147–4161
Feng D, Lu L, Yuan-Wu Y, Li GY, Feng G, Li S (2013) Device-to-device communications underlaying cellular networks. IEEE Trans commun 61(8):3541–3551
Fettweis GP (2014) The tactile internet: applications and challenges. IEEE Veh Technol Mag 9
Gamage AT, Shen XS (2018) Resource allocation for D2D communication underlaying cellular/WLAN interworking. In: Resource management for heterogeneous wireless networks. Springer, pp 47–64
Gui J, Zhou K (2018) Cellular throughput optimization by game-based power adjustment and outband D2D communication. EURASIP J Wirel Commun Network 1:254
Guo J, Durrani S, Zhou X (2014) Outage probability in arbitrarily-shaped finite wireless networks. IEEE Trans Commun 62(2):699–712
Guo J, Durrani S, Zhou X, Yanikomeroglu H (2016) Device-to-device communication underlaying a finite cellular network region. IEEE Trans Wirel Commun 16(1):332–347
Hao Y, Ni Q, Li H, Hou S (2018) Robust multi-objective optimization for ee-se tradeoff in D2D communications underlaying heterogeneous networks. IEEE Trans Commun 66(10):4936–4949
Huang S, Li J, Huang J (2019) DTMR: delay-aware traffic admission, mode selection, and resource allocation for D2D communication. Wirel Netw 25(1):439–453
Im S, Jeon H, Lee H (2008) Autonomous distributed power control for cognitive radio networks. In: Proceedings of IEEE VTC, Canada
Index CVN (2019) Forecast and trends, 2017–2022. CISCO
Janis P, Koivunen V, Ribeiro C, Korhonen J, Doppler K, Hugl K (2009) Interference-aware resource allocation for device-to-device radio underlaying cellular networks. In: Proceedings of IEEE VTC. Barcelona, Spain, pp 1–5
Jiang Y, Liu Q, Zheng F, Gao X, You X (2015) Energy-efficient joint resource allocation and power control for D2D communications. IEEE Trans Veh Technol 65(8):6119–6127
Kai C, Li H, Xu L, Li Y, Jiang T (2019) Joint subcarrier assignment with power allocation for sum rate maximization of D2D communications in wireless cellular networks. IEEE Trans Veh Technol 68(5):4748–4759
Kazeminia M, Mehrjoo M, Tomasin S (2019) Delay-aware spectrum sharing solutions for mixed cellular and D2D links. Comput Commun 139:58–66
Khazali A, Sobhi-Givi S, Kalbkhani H, Shayesteh MG (2018) Energy-spectral efficient resource allocation and power control in heterogeneous networks with D2D communication. Wirel Netw, pp 1–15
Kim HM, Nguyen HV, Kang GM, Shin Y, Shin OS (2018) Spectral efficiency maximization for D2D communications underlaying a cellular system. In: Proceedings of IEEE PIMRC. Italy, pp 588–589
Lee J, Lee JH (2019) Performance analysis and resource allocation for cooperative D2D communication in cellular networks with multiple D2D pairs. IEEE Commun Lett 23(5):909–912
Lei L, Kuang Y, Cheng N, Shen XS, Zhong Z, Lin C (2015) Delay-optimal dynamic mode selection and resource allocation in device-to-device communications-part I: optimal policy. IEEE Trans Veh Technol 65(5):3474–3490
Li X, Shankaran R, Orgun MA, Fang G, Xu Y (2018) Resource allocation for underlay D2D communication with proportional fairness. IEEE Trans Veh Technol 67(7):6244–6258
Lien SY, Chien CC, Tseng FM, Ho TC (2016) 3GPP device-to-device communications for beyond 4G cellular networks. IEEE Commun Mag 54(3):29–35
Lin X, Andrews JG, Ghosh A, Ratasuk R (2014a) An overview of 3GPP device-to-device proximity services. IEEE Commun Mag 52(4):40–48
Lin X, Ratasuk R, Ghosh A, Andrews J (2014b) Modeling, analysis, and optimization of multicast device-to-device transmissions. IEEE Trans Wirel Commun 13(8):4346–4359
Mach P, Becvar Z, Vanek T (2015) In-band device-to-device communication in OFDMA cellular networks: a survey and challenges. IEEE Commun Surv Tutorials 17(4):1885–1922

Mi X, Zhao M, Xiao L, Zhou S, Wang J (2015) Delay-aware resource allocation and power control for device-to-device communications. In: IEEE WCNCW, pp 311–316

Mustafa HA, Shakir MZ, Sambo YA, Qaraqe KA, Imran MA, Serpedin E (2014) Spectral efficiency improvements in Hetnets by exploiting device-to-device communications. In: Proceedings of IEEE Globecom Workshops, Texas, pp 857–862

Sobhi-Givi S, Khazali A, Kalbkhani H, Shayesteh MG, Solouk V (2018) Joint mode selection and resource allocation in D2D communication based underlaying cellular networks. Telecommun Syst 67(1):47–62

Trigui I, Affes S (2018) Unified analysis and optimization of D2D communications in cellular networks over fading channels. IEEE Trans Commun 67(1):724–736

Wei L, Hu RQ, Qian Y, Wu G (2015) Energy efficiency and spectrum efficiency of multihop device-to-device communications underlaying cellular networks. IEEE Trans Veh Technol 65(1):367–380

Wu X, Tavildar S, Shakkottai S, Richardson T, Li J, Laroia R, Jovicic A (2013) FlashLinQ: a synchronous distributed scheduler for peer-to-peer ad hoc networks. IEEE/ACM Trans Network 21(4):1215–1228

Wu Y, Wang J, Qian L, Schober R (2015) Optimal power control for energy efficient D2D communication and its distributed implementation. IEEE Commun Lett 19(5):815–818

Ye Q, Al-Shalash M, Caramanis C, Andrews JG (2014) Distributed resource allocation in device-to-device enhanced cellular networks. IEEE Trans Commun 63(2):441–454

Yin R, Zhong C, Yu G, Zhang Z, Wong KK, Chen X (2015) Joint spectrum and power allocation for D2D communications underlaying cellular networks. IEEE Trans Veh Technol 65(4):2182–2195

Zhang H, Song L, Han Z, Zhang Y (2018a) Radio resource allocation for device-to-device underlay communications. In: Hypergraph theory in wireless communication networks. Springer, pp 21–39

Zhang R, Li Y, Wang CX, Ruan Y, Fu Y, Zhang H (2018b) Energy-spectral efficiency trade-off in underlaying mobile D2D communications: an economic efficiency perspective. IEEE Trans Wirel Commun 17(7):4288–4301

Zhao X, Yuan P, Chen Y (2017) Chen P (2017) Femtocaching assisted multi-source D2D content delivery in cellular networks. EURASIP J Wirel Commun Network 1:125

Zhao Y, Li Y, Zhang H, Ge N, Lu J (2016) Fundamental tradeoffs on energy-aware D2D communication underlaying cellular networks: a dynamic graph approach. IEEE JSAC 34(4)

Zhou Z, Dong M, Ota K, Wu J, Sato T (2014) Energy efficiency and spectral efficiency tradeoff in device-to-device (D2D) communications. IEEE Wirel Commun Lett 3(5)

Zulhasnine M, Huang C, Srinivasan A (2010) Efficient resource allocation for device-to-device communication underlaying lTE network. In: Proceedings of IEEE WiMob, China, pp 368–375

# Chapter 16
# V2X Communications: Recent Advancements and Performance Analysis


**Shubham Dwivedi, Devendra Singh Gurjar, Prabina Pattanayak, and Tripti Goel**

**Abstract** Vehicular communication is a key technology for realizing automatic vehicle systems, improving the safety of drivers and passengers, and providing support for smart traffic handling. In addition, it can help the drivers to take appropriate decisions and support the modernization of different operations of vehicles along with some useful applications for the passengers. In this chapter, we discuss various forms of vehicular communications, e.g., vehicle-to-vehicle, vehicle-to-infrastructure, and vehicle-to-pedestrian communications. We also highlight the recent advancements done for different modes of vehicular communications, which can be summarized as vehicle-to-everything (V2X) communications. Moreover, we identify the potential research directions to address the challenges for bringing V2X communications into operation. To illustrate the impact of mobility, we provide mathematical modeling of the node's mobility and analyze the system performance in terms of outage probability and system throughput under Rayleigh fading environment.

**Keywords** V2X communications · V2V communications · V2I communications · Time-varying channels

## 16.1 Introduction

The intelligent transportation systems (ITS) are expected to be an integral part of the upcoming developed societies. Vehicle-to-everything (V2X) communication has flair of becoming the foundation for such a futuristic transportation system. This is

S. Dwivedi · D. S. Gurjar (✉) · P. Pattanayak · T. Goel
National Institute of Technology Silchar, Silchar, Assam, India
e-mail: devendra.gurjar@ieee.org

S. Dwivedi
e-mail: shubhamdwivedi96@gmail.com

P. Pattanayak
e-mail: prabina.pattanayak@ieee.org

T. Goel
e-mail: triptigoel83@gmail.com



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_16

due to the fact that it significantly improves the efficiency of traffic in several aspects, e.g., traffic management, monitoring of the traffic and its control. Further, it can also provide a robust network connectivity to transfer a large amount of data among vehicles and infrastructures that are related to the traffic management. Moreover, the vehicles can also interact with the infrastructure (centralized base station or relay node). Herein, the protocols employed in cellular systems for route mapping and localization can be of particular interest. Also, mobile vehicles can share the information related to the current location, speed, and direction through the secure transmissions that can be realized to typically 300 m in any direction under some extreme environment barriers such as heavy snowfall or fog. Further, pedestrians and bicycles can be connected to the common platform using portable devices such as smartphones, and thus, drivers can access their information in well advance.

According to the guidelines of the Department of Transportation (DOT), USA, vehicle-to-vehicle (V2V) communication should provide 360-degree situational awareness on the roads to enhance the safety of passengers and vehicles. For personal privacy, DOT suggested that no personal information about the vehicle or driver should be broadcasted using V2V technology, except the information required for general safety. The national safety administrations of different countries are also focusing on ensuring the security of transmitted data for V2X communications against any organized cyber-attacks. In general, three barriers are stopping the deployment of fully autonomous vehicles on the roads; they are

- The price of sophisticated technological components.
- Consumer trust in the automation technology.
- Lack of well-defined regulations.

For fulfilling the communication requirements of the vehicles, vehicular communications can be realized in the cellular networks to minimize the cost of transmission. Due to the mobile nature of vehicles, the topology for vehicular networks turns out to be highly dynamic and therefore poses several challenges for its implementation. As vehicular communications can be accomplished in different ways, this chapter aims to give a brief overview of different forms of V2X communications and discusses the relevance of fifth-generation cellular technologies. Also, it encompasses different types of communications that are pertaining to V2X communication, such as V2V communications, vehicle-to-infrastructure (V2I) communications, vehicle-to-pedestrian (V2P) communications, and vehicle-to-network (V2N) communications.

### *16.1.1 Need for Vehicular Communication*

Every year, we come across many reports which suggest that traffic mishaps are taking dozens of lives, and astonishing fact is that it even surpasses deaths caused by an epidemic or natural disaster. If we take the scenario of India, it is shocking that even though India is having less than one percent of the world's automobiles, the nation has a share of around six percent of total road mishaps and around ten

percent of all the traffic injuries over the globe (Singh 2017). In 2016, a report by the Indian government marked that around 0.5 million people were injured in road accidents. Similarly, the World Health Organization (WHO) report stated that around 1.35 million people are killed every year due to traffic mishaps on roads over the globe (Soleimani and Boukerche 2018). Notably, the road traffic injuries are at eighth position in the list of causes of death for all ages globally (WHO 2018). It is also the leading cause of death of young people having age between 5 and 29 years.

One can list numerous factors that are behind these accidents, such as component malfunctioning, sudden natural mishap, severe climatic conditions, and many more. Some of these factors are out of human control, and the noteworthy fact is that about 90% of cases are caused due to driver's fault. Also, if vehicles are following a lane, then there are certain chances that the vehicles can bump into each other if the crash occurs between any two vehicles. Also, about 60% of these traffic accidents could be avoided if some prior warning is available to the drivers (Wang and Thompson 1997). In crisis circumstances, a driver typically depends on the tail brake light of the vehicle ahead to choose their braking activity. Under normal street circumstances, this generally is not the best impact evasion procedure for diverse reasons. As human eyes have a limitation that it cannot see past the vehicle before them, it limits the drivers to foresee the emergency event. Human drivers also experience the ill effects of observation limitations on roadway crisis occasions, bringing about a considerable amount of delays in communicating crisis alerts.

Chain crashes can be possibly averted, or their seriousness can be reduced by lessening the slowdown between the time of a crisis occasion and the time at which the vehicles behind are trained about it (Xu et al. 2003). One approach to give more opportunity for the drivers to respond in crises is to create ITS applications by utilizing evolving wireless communication technology. Driver help frameworks are intended to assist drivers with the driving procedure to evade from mishaps, also enhancing the speed and control based on the traffic. Many frameworks offer help to the drivers, for example, traffic sign recognition, automatic parking, adaptive cruise control, and so on. These methods utilize the capacity of the vehicles to share the critical data between them, along with the infrastructure around them. With the approaching updates of these systems, they are turning into an appealing answer to extend support with the sharing techniques between vehicles.

### *16.1.2 Challenges in Vehicular Communications*

The V2V and V2I communications are characterized by high mobility of vehicles, dynamic surroundings, and relatively lower antenna height corresponding to the communicating nodes that can be vehicles or infrastructure. These physical phenomenons are responsible for making wireless channel modeling for such dynamic systems significantly challenging. The important aspects that distinct vehicular communications from other wireless communication scenarios are perceived as (Viriyasitavat et al. 2015)

- Vehicular communications are realized in diverse communication environments.
- Mixtures of different communication forms, e.g., V2V, V2I, and V2P communications make compatibility issues.
- The communicating entities can be either static or mobile that significantly affect the achievable performance.

In recent developments, several channel models have been studied so far mainly for the cellular networks, which may not be suitable for the vehicular networks, because of the different vehicular statistics. In particular, differences in the height of antennas at the transceivers can exhibit a considerably different signal propagation behavior. Moreover, communication distance and operating frequency considered for vehicular communications are different from those in the cellular networks. In particular, vehicular communication systems are developed to function at 5.9 GHz applicable for short distances (10–500 m). In contrast, the cellular networks realize communications at 700–2100 MHz for relatively longer distances up to tens of kilometers (Viriyasitavat et al. 2015).

## 16.2 Vehicle-to-Vehicle (V2V) Communications

In V2V communications, vehicles are capable of interacting with other vehicles of their close proximity. Primarily, V2V is introduced as an accident evasion innovation, which depends on the correspondence of data between close-by vehicles to possibly alert drivers about crisis circumstances that could prompt an accident. For instance, V2V can help in sending an alert message to the drivers that a vehicle ahead is braking, and they have to back off. Or let a driver realize that it is not safe to continue through convergence as another vehicle (yet not in the vision of the driver) is rapidly drawing closer. V2V frameworks are made out of devices, introduced in vehicles that utilize dedicated short-range radio communication (DSRC) to trade messages having vehicle data (e.g., heading, braking status, vehicle's speed). V2V devices utilize this data from different vehicles and decide whether a warning to the vehicle's driver is required, which could forestall a vehicle crash.

To give a better idea about V2V communications, a V2V scenario is considered in Fig. 16.1 where three cars, namely A, B, and C, are running on a roadway. These cars are moving in the left direction. When any of the two cars come under the radio communication range, they establish a network that enables them to share information regarding direction, speed, and position of the car. The cars can also be seen as routers as the information can be sent to the distant car by multiple hops. The algorithm should be such that it can handle the fast transitions in the network topology.

The V2V utilizes a decentralized framework that can be arranged with fully connected or partially connected mesh topology (Arena and Pau 2019). In earlier days, this topology was quite costly and hard to implement, as every node must be physically associated with the others using wired connections. Nowadays, these concerns

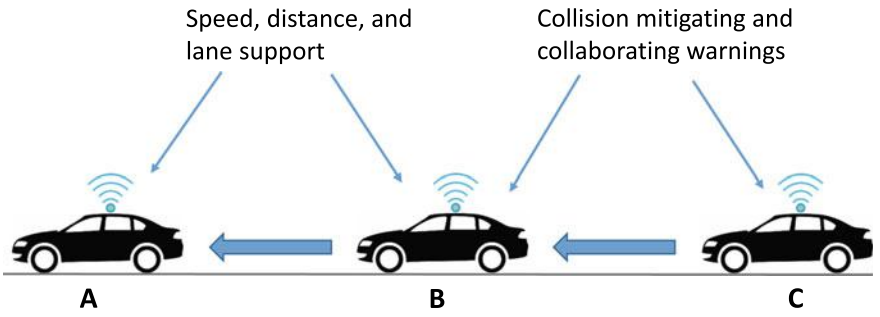



**Fig. 16.1** Vehicle-to-vehicle communications

have been fixed by employing the virtue of wireless communications along, especially the advent of wireless personal area networks. In such systems, all vehicular nodes (mobile or static) need to form partial or full mesh connections. Instead of depending on the base station, the autonomous nodes can transfer the data packets to other intended node(s). The wireless topology changes fast and unpredictably as the nodes make random movements. Moreover, these systems can work alone or be associated with the Internet to give more services.

Based on the the deployed system configuration, the driver can get a warning signal in case of a mishap to happen, or the vehicle itself can freely take preventive activities that include emergency braking, cruise control, and cautionary intercessions. As such, V2V communications are expected to be more compelling than current embedded frameworks where vehicle safety is altogether subject to the functionality of different locally available onboard sensors, super-fast cameras, and radars (Arena and Pau 2019). The vehicle can respond to any critical circumstances depending on parameters identified by these components set on it. Generally, the fundamental parameters that can be monitored are the drive speed, the range from an impediment, or the proximity of vulnerable blindside of vehicle. These existing innovations are reliable. However, calculation mistakes cannot be omitted. Albeit, V2V communication protocols can improvise the safety domain by offering collaboration from all vehicles that are near to one another. They can help the vehicle at serious risk (e.g., driver's sleep, a segment breakdown, obstruction in the path, etc.) to make effective decisions in taking care of the issue. Thus, the primary role of every vehicle forming mesh network is information collection to ensure fair security to itself and its neighbors. This framework can be referred to as cooperative awareness.

The emerging V2V technology, along with the other currently available systems, will enhance the management of traffic problems. The integration of V2V with the V2I can change the scenario of the transportation systems. It can be considered that soon, self-driving reliable cars will run on roadways, and that will be possible because of the ITSs. However, the implementation of these technologies has three major impediments: the cost of deployment of this venture; the privacy and security of data that is exchanged; the willingness from automobile companies to incorporate such

features along with the consideration for standard security protocols. At present, the picture is not clear how the fund will be raised for technology deployment, whether it will be public or private. It is encouraging that automobile giants such as BMW, Audi, Daimler, General Motors, Volvo are working on the development and deployment of these technologies in their vehicles.

### *16.2.1 Prior Arts on V2V Communications*

In recent years, one can see the surge in the use of novel machine learning algorithms for different applications. Likewise, for V2V communications, the researchers in Ye et al. (2019) have developed a novel decentralized resource allocation technique that utilizes deep reinforcement learning. The proposed method can satisfy the latency constraints along with the minimization in the interference to V2I communication. Sensors are used to collect information from nearby entities to get more information about the surrounding environment. One of the sensors in this regard can be windshield-installed cameras. Researchers of Gomes et al. (2012) have proposed a driver assistance system that converts the vision barring vehicles in the form of transparent tubular targets with the help of V2V communication and wrap-around cameras. By doing so, the maneuvers become safer as this method enhances the visibility of drivers. In the coming years, automobiles are expected to serve a variety of services. Technically, these services will be based initially on the transmission of beacon messages that gives an insight into the presence of the vehicle. In this regard, the investigators of Bazzi et al. (2015) have proposed that the long-term evolution (LTE), a 4G mobile communication standard, can be used as the enabling technology for V2V beaconing. This can be achieved by exploiting device-to-device (D2D) communication and full-duplex (FD) radio provided by LTE.

For designing an automatic driving system, preview control is an integral concept that needs to be incorporated into the system. Motivated by this, a preview controller for vehicle stability using V2V communication is proposed in Yim (2017). The idea is that the preceding vehicle driver's steering input that is obtained on the following automobile acts as a reference for preview control. As discussed in the earlier section, link reliability is a crucial performance metric, and researches are going on to enhance reliability. One of the modern methods, as proposed in Park et al. (2019), is based on resource size control (RSC). Based on the message size, communication range, and vehicle density, the offered RSC alters the resource size. This technique results in enhancement of link reliability as compared to the fixed resource size technique. Further, interference can be seen as a major impairment and becomes the limitation for high-rate transmissions. With the advancement in V2V communication, interference mitigation techniques are investigated. Recently, researchers have proposed schemes based on the Doppler-shift-based frequency-domain interference alignment scheme and frequency-domain pre-coding scheme. These techniques help in dealing with interference (Wu et al. 2017). Security and privacy are the primary requirements of any wireless communication. This motivates the development of

secure communication links. To maximize the secrecy performance, the joint relay selection and sub-carrier allocation schemes are devised in Han et al. (2017), along with no compromise in fairness. The problem can be modeled in a random bipartite graph (RBG) and solved by an algorithm with low complexity.

### *16.2.2 Future Research Problems on V2V Communications*

1. Different machine learning algorithms and their combinations can be incorporated to develop novel resource allocation techniques that can minimize the latency in V2V communications.
2. Safe maneuvers are an important metric in V2V communications. Therefore, efficient schemes to improve maneuvering can be considered as future problems. Moreover, designing robust sensors or cameras to assist V2V communication can also serve this purpose.
3. Future problem can be based on beacon signaling as it helps in localizing vehicles.
4. One can come up with new interference mitigation techniques to maximize the achievable data rates in V2V communication.
5. Enhancing security and privacy are other interesting and crucial areas of research for the current scenario in the era of supercomputers.
6. Novel computational algorithms can be devised to reduce computational complexities.

## 16.3 Vehicle-to-Infrastructure (V2I) Communications

V2I communication enables vehicles to communicate with the roadside infrastructure, unlike the V2V communication (concerned only with the transfer of data among vehicles). The components used for supporting V2I communications include RFID readers, traffic lights, roadside cameras, lane markers, street lights, and parking notes. In a technical way, V2I is closely related to V2V, as it also uses DSRC for interfacing. The protocols of the ad hoc network allow the infrastructure to interact with the vehicle or vice versa. The strength of the V2I is evident by the fact that real-time advice concerning traffic conditions can be made. Scenarios like traffic congestion, the availability of nearby parking, accident-prone areas can also be communicated by acquiring data by robust sensors. Similarly, variable speed limits and adjustment of the signal phase and timing can be achieved, which in turn help in effectively saving fuel and maintaining a steady flow of traffic. These advisories can be communicated to operators of vehicles by road displays or by wireless connections. To better comprehend the concept, we take the example of having two different traffic scenarios. On the right-hand side of the highway in Fig. 16.2, the V2I focuses on improving fuel efficiency and reduction of emission from vehicles by controlling through central infrastructure when the traffic is low. On the left-hand side of the

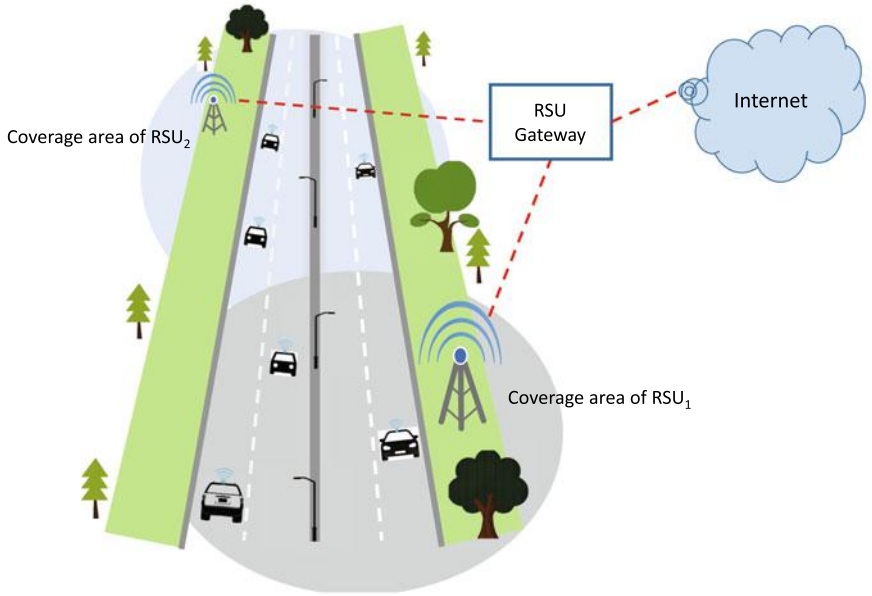



**Fig. 16.2** Vehicle-to-infrastructure communications

highway, the main concern is depleting queues when there is greater congestion. This example emphasizes that the motto of V2I infrastructures changes with the changing traffic conditions. Roadside units (RSU) play an important role in providing network coverage to the vehicles of highways. A group of nearby RSUs is connected to the Internet and other control units through RSU gateways. For providing smooth network connectivity even at higher speeds, soft hand-off techniques are employed.

To accomplish the vision of autonomously driven cars, the primary need is the development and implementation of the hardware, software, and firmware that collectively establish the reliable link between infrastructure and vehicles. Many countries are planning to improve safety and mobility by implementing smart road infrastructures. Governments face the grave challenge of funding in such projects. It is predicted that by 2020–2025, the V2I systems will be devised and implemented.

### *16.3.1 Prior Arts on V2I Communications*

In V2I communications, the required amount of real-time data rates are exceptionally high to take prompt decisions. Therefore, it is important to focus on paradigms that can fulfill such demand of high data rates. From the recent studies, it has been shown that massive multiple-input multiple-output (MIMO) can be a key technology to satisfy the needs of an intelligent transport system, which are low latency, high

data rate along with high energy efficiency (Gurjar and Upadhyay 2016b). Adaptive tracking of vehicles is also possible by using multiple antenna elements because of viability of multiple beams formation in the elevation domain. It thereby enhances spatial multiplexing efficiency along with the surge in the throughput. Recently, researchers in Zhang et al. (2016) have developed a packet-level channel model for V2I elevation multiplexing links that suffices the interference issue of MIMO-based systems. Estimation of channel state information for effective beamforming is also crucial to maintain the performance of communication systems (Gurjar and Upadhyay 2016a). Further, to deal with high computational complexities, fast processing algorithms have been developed in Jiang et al. (2018). On the other hand, adaptive directional beamforming comes in the limelight to make MIMO communications more effective. This approach ultimately improves the capacity of V2I communications. Researchers of Pyun et al. (2016) have proposed an iterative method that allocates the resources to maximize the capacity, whereas the interference has been mitigated by leveraging the beam properties.

Another potential solution to address the needs of V2I communications is to exploit millimeter-wave (mm-Wave) communications. However, channel estimation is a cumbersome task in millimeter-wave communication systems as it consumes resources such as hardware, bandwidth, and time. These channel tracking schemes can play a role in V2I communication. Recently, data-aided iterative channel tracking has been proposed that uses the spatial diversity of mm-Wave channels employing a lens antenna array (Gong et al. 2019). To detect the current data, previous time slots are being used for channel estimation that, in turn, reduces the overhead of channel estimation. As any imperfections in channel state information (CSI) impair vehicular communication such as V2V and V2I. Therefore, to obtain accurate CSI, the researchers in Yang et al. (2017) have proposed the inter-vehicle cooperation channel estimation (IVC-CE) method. Single or multiple vehicle measurements, such as the speed and position of automobiles, are used in this optimal channel estimation.

In urban vehicular environments, positioning accuracy is an issue of paramount importance. The existing GPS cannot provide that level of precision. So, localizing the vehicle is an active area of research. In this regard, authors in Fascista et al. (2017) have proposed a GPS-free localization scheme that leverages the architecture of the V2I framework. The beacon packets sent by RSU play a major role in this method. The weighted least square algorithm incorporates estimating parameters such as the angle of vehicle arrival along with the position information of RSU, thus resulting in effectively localizing the vehicles. The information provided by V2V and V2I communication channels can help to make these automobiles more energy-efficient. In this direction, the researchers in Zhang et al. (2017) have utilized the chaining neural network (CNN) for velocity predictions based on the information rendered by V2V and V2I frameworks, whereas the resulting velocity predictions play a major role in the energy management of these vehicles.

### *16.3.2 Research Directions on V2I Communications*

1. Assimilating MIMO with V2I architecture is a crucial area to work. In particular, dealing with the challenges that arise in MIMO systems like interference, beamforming, computational complexities can be considered as worth doing problems.
2. Channel estimation is another potential area as it is a cumbersome task to reduce the overhead involved in channel tracking.
3. Higher accuracy in the localization of vehicles is the need of the hour, and therefore, effective localization methods can be devised.
4. Utilizing V2I frameworks in energy or fuel conserving scenarios can be a new area of research.

## 16.4 Vehicle-to-Pedestrian (V2P) Communication

One of the most vulnerable beings is pedestrians in vehicle collision scenarios. Also, it is reported that a fair percentage of accidents that lead to injuries and fatalities account from vulnerable road users (VRUs) such as bicyclists and pedestrians. With the help of data provided by various national traffic safety agencies, one can find that notably affected beings are VRUs. This alarming surge in fatality results point of attraction for researchers around the world. The dominant direction of driving research is collision avoidance approaches. The idea is to stop vehicles before possible clashes by providing collision warnings.

Conventional systems rely on noncooperative perception sensors such as IR sensors, RADARs, and LASER scanners for getting the warning. However, these technologies suffered from the limitation of the line-of-sight. For instance, the pedestrian who is obstructed by a nearby tree can suffer from a fatal accident. In recent years, the idea of using wireless communication to exchange the collected data among vehicles or between pedestrians and vehicles became popular. There are several benefits of implementing a communication-based safety system in respect of promptness, and the capability to act in the presence of obstacles. In 2003, the Federal Communication Commissions (FCC) established DSRC service due to which pedestrian's road safety got enhanced. With the sensor-based approach, wireless communication technology has been clubbed in recent years. Also, the concept of the vehicular ad hoc network (VANETs) is quite popular nowadays, wherein the entities can exchange data wirelessly without the requirement of the central hub or server. This technology can be seen as an emerging advanced technology that offers desired outcomes. Using a number of these vehicular ad hoc networks, many VRUs can connect or interact with nearby vehicles. The combination of roadside sensors and VANETs can help in averting the possible mishaps by giving warnings.

Nowadays, excessive use of smartphones during driving is acting as a major distraction and contributing aspect in numerous vehicle accidents. There are many reports which are suggesting that the use of cell phones has impaired driver attention.

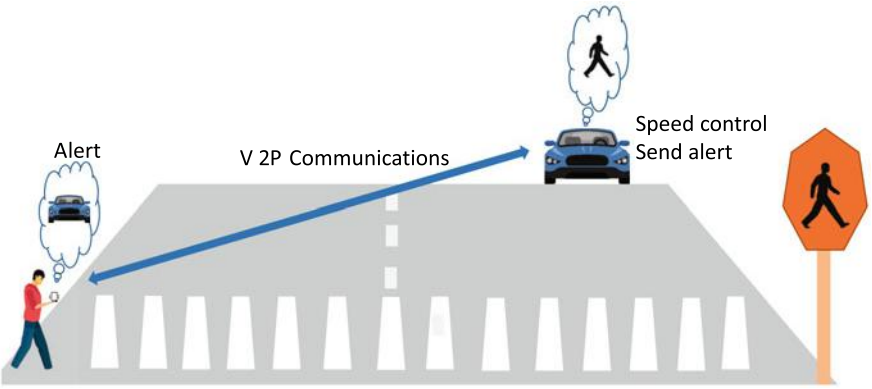



**Fig. 16.3** Vehicle-to-pedestrian communications

Also, the pedestrians can be at grave risk when texting or talking on the cell while crossing the road. Some researchers suggest the solution of this problem through cellular technology. The concept can be perceived as the replacement of a conventional handheld communication device by DSRC-enabled smartphones. WLAN and Wi-Fi technology can be used to establish a communication link between the pedestrian and incoming vehicles. In Fig. 16.3, pedestrian, and vehicle share information such as speed, direction, and location, which in turn help avoid mishaps and to manage the traffic. A pedestrian using smartphone or walking sticks or bicycles installed with a wireless module can communicate with on-road vehicles, thus giving alert indications to the drivers. Although the incorporation of existing technology to provide a new approach is commendable, there are certain limitations that one should keep in mind. These impairments include varying transmission quality based on the surrounding conditions, smartphones having short battery lives, the inability of positioning systems to pinpoint the exact location. It is important to note that most of the experiments are done under some predefined conditions, which are quite different from the real traffic scenario. Therefore, conducting on-field trials is one of the major concerns before enabling such technologies in practical vehicular systems.

### *16.4.1 Recent Research on V2P Communications*

In recent years, researchers have focused on modeling of accurate channel models for V2P communication. In one of the research scenarios, the researchers have proposed a wideband measurement-based stochastic channel modeling scheme along with parameterization (Makhoul et al. 2019). It was comprised of the first path component and the secondary multi-path component originated due to detection and tracking. Further, it has been suggested that the culmination of path loss, large and small-scale fading help in the modeling of these paths. Another group of researchers

incorporated non-line-of-sight (LoS) scenarios by designing a 3D scenario model that stimulates the diffraction at a row of parked automobiles (Rashdan et al. 2019). This loss caused by diffraction has been analyzed and eventually modeled by combining the two-ray path loss model, along with the multiple knife-edge diffraction model. The fusion of perception and communication technology can be an effective step in enhancing the safety of VRUs. This novel idea paved a new domain for research. Integration of embedded sensors-based perception systems and V2P communication can collectively provide better safety for VRUs. Using multi-hypothesis tracking, researchers developed a probabilistic association between perception and V2P communication information. The results showed that the benefits of both technologies enhance the safety of VRU (Merdrignac et al. 2017). The above scenario motivates to enhance the classification performance of the sensors involved and to improve the accuracy of GPS positioning data sent by the communication system. Further, the usage of the 2.5-D multi-layer laser scanner has been proposed as the best option for the primary sensor in Kim et al. (2016). They also suggested new features for this scanner that enhances the classification performance and employed support vector machine to lessen the computational time involved. This method has shown a good response in dealing with real-life situations.

On the same track of enhancing the perception systems, researchers have proposed a density enhancement method for long-range pedestrian detection (Li et al. 2016). The limitation of 3D LIDAR used here can show the sparse behavior of the object point when the vehicle is away. By using a new evaluation metric and incorporating the interpolation and finally using a resampling algorithm, a new point cloud is generated. During the daytime, vision systems find no difficulties in detecting a pedestrian. But the case is quite different in the night-time, as now the pedestrian does not act as an apparent object. To deal with these problems, the researchers have proposed a night-time part-based recognition scheme (Lee et al. 2015). In this approach, the pedestrian is divided into parts with a camera and a near-infrared lighting projector. Selecting useful parts is a significant aspect of this scheme. Spatial relationships have been analyzed to choose these parts.

### *16.4.2 Research Directions on V2P Communications*

1. Channel modeling for the V2P scenario is an active area of research to design a reliable V2P communication framework.
2. Perception-based systems installed on vehicles can be improved, or novel systems should be devised to enhance the safety of VRUs.
3. Algorithms that play a key role in perception systems for different operations like segmentation, classification, and feature extraction can be made more robust to make systems more capable of real-life scenarios.
4. Effective methods should be devised to deal with the different changing environments around pedestrians such as day-night, sparse or dense traffic scenarios.

## 16.5 Self-driving Vehicles

Self-driving vehicle (SDV) can be taken as a promising example of the ITS that is introduced to transform the roadways. After its deployment, the traffic jams and number of accidents are expected to be reduced significantly. This technological advancement will offer a great opportunity for the people who cannot drive vehicle. The driverless vehicles could essentially change the way of thinking about the transportation system. Indeed, the SDV technology has already made its appearance on the roads of American streets. For instance, Uber has introduced self-driving cabs in Pittsburgh and is planning to test self-driving trucks for long-range commercial deliveries. On the other hand, Google's prototype cars are also running on the roads along with human supervisors for others' safety. Other automakers like Tesla, Toyota, and Subaru are also planning to include features such as guided steering and automatic braking on new cars. Fully automated/autonomous or "self-driving" vehicles are defined by the NHTSA of USA as those vehicles in which execution of different operations transpires without the direct input of the handlers to control the steering, acceleration, and braking. These vehicles are devised in such a manner that the driver is not needed to observe the roadway while running in self-driving mode continually. The NHTSA has further defined vehicle automation process into five levels. In this list, the higher level shows that the vehicle is equipped with more automation features. The NHTSA's five levels of automation can be listed as follows:

1. **No-automation (level zero)**: For this level, the driver is the complete and sole controller of the primary vehicle controls such as brake, throttle, and steering at all times.
2. **Function-specific automation (level one)**: Automation involves one or more specific control functions. For example, pre-charged brakes or electronic stability control, where the vehicle automatically applies brakes to enable the driver to recapture control of the vehicle.
3. **Combined function automation (level two)**: This level enables automation of at least two primary control functions. An example of combined functions of this level system is lane centering in combination with adaptive cruise control.
4. **Limited self-driving automation (level three)**: At this level of automation, driver can give up full control of all safety-critical functions under specific traffic or environmental conditions and rely heavily on the vehicle to monitor for changes. However, the driver must be available with a sufficiently comfortable transition time for occasional control. Google's car of the second generation is an example of the limited self-driving vehicle of this level.
5. **Full self-driving automation (level four)**: Different from level three, all the monitoring of roadway conditions and safety-critical driving functions are performed by the vehicle for an entire trip. Herein, the role of the driver is to provide destination or navigation input only. Google's car of the third generation is based on full self-driving automation technology. Vehicles equipped with level four automation technologies may be referred to as autonomous vehicles.

From these five levels of automation, only up to level two is accessible to the public. Albeit, the manufacturers are now developing, and testing vehicles of level four on the public roads.

### *16.5.1 Challenges for SDV*

Technical and scientific experts have listed the challenges and roadblocks that need to resolve before the handover of SDV in the public domain.

#### 16.5.1.1 Unexpected Encounters

While testing, it has been observed that SDV struggles to interpret unfamiliar situations, such as traffic police official gesticulate vehicles using a red light. For these unexpected encounters, simple rule-based programming will not give practical solutions since it is impossible to code for every possible situation in advance. In practice, body language and other contextual signs benefit people to navigate these situations profoundly, but it is a challenging task for a computer to interpret such situations on the road. For tackling these encounters, automatic vehicles have to be able to learn roadside conditions through artificial intelligence.

#### 16.5.1.2 Sensing the Surroundings

To a cyber-system, a highway with fog or dust surroundings seems completely different than on a bright day. To adjust with the current conditions, SDV has to sense road and traffic characteristics in all possible situations regardless of lighting and weather conditions. Recent tests in MIT show that SDV can perform efficiently in rainy conditions but the performance in snow environment is still not as per expectation. Moreover, all the embedded sensors need to be compact, reliable, and reasonably priced and should be compatible with the local maps so that vehicle can take a smart decision during the ride.

#### 16.5.1.3 Human–Robot Interaction

Making human robot interaction with time efficiency is another challenging task for SDV systems. Even with level four vehicles, most SDV will be semi-autonomous for at least the coming future. But figuring out the responsibilities for vehicles and human can make the vehicular system more prone to accidents. The key questions are how does a vehicle inform a passenger who has been taking a nap or reading something that to take control of a task, and how do the vehicles know that the passenger is ready to take over?

#### 16.5.1.4 Cyber-Security

The recent demonstration by a group of hackers proved that even conventional vehicles (level one or level two) are vulnerable and can be wirelessly accessed to occupy controls of steering and braking via the onboard entertainment system that could lead to an accident. After deployment, SDV are expected to get updates and maps through the cloud, which can introduce even greater risk.

## 16.6 Vehicle-to-Network (V2N) Communications

In this form of communication, mobile vehicles are capable of making communications through centralized network for different V2X application services. For instance, in Fig. 16.4, mobile relay node (roadside unit) establishes a communication link between mobile vehicle and vehicular base station using a cellular network.

### *16.6.1 Modeling of Node's Mobility*

A three node architecture is considered, consist of a vehicular base station (VBS), a mobile vehicle (MV), and a roadside relay node (RN) as shown in Fig. 16.4. All the vehicular nodes are equipped with single antenna systems and operate on half-duplex mode. In this model, due to the relative velocity between the two vehicular nodes, all the channel gains are to follow time-varying Rayleigh fading. These channel

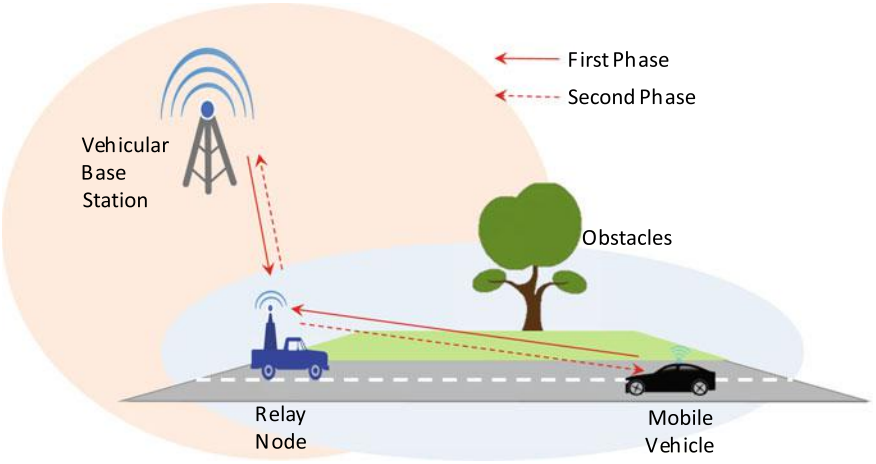



**Fig. 16.4** System model for two-way communications in V2N

coefficients are mathematically modeled as the first-order Markovian process. The channel coefficients between VBS to RN links and RN to MV link are denoted as $h_{B,R}$ and $h_{R,M}$, respectively. Further, it is assumed that all the nodes suffer from the additive white Gaussian noises (AWGN), which are defined as $n_j \sim \mathcal{CN}(0, \sigma^2)$, for $j \in \{B, R, M\}$. Unit-energy symbols $s_B$ and $s_M$ from VBS and MV are transmitted in the first phase. Based on the definition of first-order Markovian process, the channel gains for $l \in \{B, M\}$ can be given as

$$h_{l,R} = \rho_{lR}\hat{h}_{l,R} + \sqrt{1 - \rho_{lR}^2}\tilde{h}_{l,R}, \tag{16.1}$$

where $h_{l,R}$ is the channel coefficient in the first phase. In particular, both $h_{l,R}$ and $\hat{h}_{l,R}$ are assumed to be jointly distributed as Gaussian random variable having a correlation coefficient $\rho_{lR}$. Further, $\tilde{h}_{l,R} \sim \mathcal{CN}(0, \Omega_{lR})$ denotes the error due to the relative velocity between two communicating nodes. Mathematically, the correlation coefficient $\rho_{lR}$ is formulated according to Jakes' model as $\rho_{lR} = \mathcal{J}_0\left(\frac{2\pi f_c v_{lR} \tau_{lR}}{c}\right)$. Here, $v_{lR}$ denotes the relative velocity between two vehicular nodes and $\mathcal{J}_0$ is used to denote zeroth-order Bessel function of first kind (Gradshteyn and Ryzhik 2007, Eq. 8.411). Moreover, $\tau_{lR}$ denotes time lag between estimation of channel and actual decoding of signals, whereas $c$ represents the speed of light. Note that, when the the communicating nodes are stationary, the corresponding correlation coefficient turns out to be unity.

### *16.6.2 End-to-End Signal-to-Noise Ratio (SNR)*

In this section, instantaneous end-to-end SNR expressions at VBS and MV are obtained for two decode-and-forward (DF) relaying techniques, i.e., XOR-based DF and superposition-based DF.

In the first phase, the received signals at RN can be expressed as

$$y_R = \sqrt{P_B}h_{B,R}s_B + \sqrt{P_M}h_{M,R}s_M + n_R, \tag{16.2}$$

where $P_B$ and $P_M$ are the transmitted power at VBS and MV. Further, $n_R$ is AWGN at RN. To provide relay cooperation, RN first decode the signals $s_B$ and $s_M$ from the combined received signals. After that, it re-encodes the decoded signals by utilizing any of the two following strategies.

#### 16.6.2.1 XOR-Based DF Strategy

In this strategy, RN performs bit-wise XOR to get the re-encoded signal ($s_B \oplus s_M$). If the bit sequence lengths of signals ($s_B$ and $s_M$) are different, zero padding technique can be utilized to ensure the same length of shorter one (Gurjar and Upadhyay 2018).

In the second phase, RN broadcasts the re-encoded signal. The received signal at the VBS and MV can be given as

$$y_{R,l}^{\text{xor}} = \sqrt{P_R} h_{R,l}(s_B \oplus s_M) + n_l, \tag{16.3}$$

where $n_l$ for $l \in \{B, M\}$ is AWGN at respective node and $P_R$ denotes the transmitted power at RN. On adopting time-varying channel model as defined in (16.1) into (16.3), $y_{R,l}^{\text{xor}}$ can be expressed as

$$y_{R,l}^{\text{xor}} = \underbrace{\rho_{Rl}\sqrt{P_R}\hat{h}_{R,l}(s_R \oplus s_M)}_{\text{Desired signal}} + \underbrace{\sqrt{(1-\rho_{Rl}^2)P_R}\tilde{h}_{R,l}(s_R \oplus s_M)}_{\text{Residual self-interference}} + n_l, \tag{16.4}$$

Since both nodes (VBS and MV) know their corresponding transmitted signals, they can cancel the partial self-interference. As a result, the residual self-interference is considered to be additional noise. Hereby, the instantaneous SNR at the nodes in the second phase can be expressed as

$$\Upsilon_{R,l}^{\text{xor}} = \frac{\rho_{Rl}^2 \eta_R |\hat{h}_{R,l}|^2}{(1-\rho_{Rl}^2)\eta_R \Omega_{Rl} + 1}, \tag{16.5}$$

where $\eta_R = P_R/\sigma^2$.

#### 16.6.2.2 Superposition-Based DF

In this strategy, RN applies superposition coding for re-encoding of the signals. Likewise to XOR-DF, RN broadcasts the re-encoded signal in the second phase. Hereby, the received signal at the MV and VBS in the second phase can be given as

$$y_{R,l}^{\text{sup}} = \sqrt{\frac{P_R}{2}} h_{R,l}(s_B + s_M) + n_l. \tag{16.6}$$

Considering the time-varying channel model as defined in (16.1) and incorporating into (16.6), we can have the expression of instantaneous SNR at VBS and MV

$$\Upsilon_{R,l}^{\text{sup}} = \frac{\rho_{Rl}^2 \eta_R |\hat{h}_{R,l}|^2}{(1-\rho_{Rl}^2)\eta_R \Omega_{Rl} + 2}. \tag{16.7}$$

To investigate the systems performance in terms of outage probability and achievable sum rate, we will utilize these SNR expression in the next section.

### *16.6.3 Performance Analysis*

In this section, the outage probability (OP) expressions are derived for signal transmission through both links VBS to MV and MV to VBS under time-varying Rayleigh fading. Further, the system throughput expression is also provided after considering a delay limited transmission scenario.

#### 16.6.3.1 Outage Probability

Outage event occurs if the RN fails to decode any of the signals ($s_B$ or $s_M$) in the first phase or the corresponding link in the second phase fails to achieve a predefined target rate.

#### 16.6.3.2 OP for XOR-DF

For XOR-based DF, the OP can be defined as

$$\mathcal{P}_{\text{out},j}^{\text{xor}} = 1 - \Pr[\mathcal{Q}_R]\ \Pr[\mathcal{R}_{R,j}^{\text{xor}} > r_j], \tag{16.8}$$

where $j \in \{B, M\}$, $\Pr[\mathcal{Q}_R]$ defines the probability that RN successfully decodes both ($s_B$ and $s_M$) in the first phase and $r_j$ denotes the target rate. Here, the instantaneous achievable rates can be defined as

$$\mathcal{R}_{R,j}^{\text{XOR}} = \frac{1}{2}\log_2\left(1 + \Upsilon_{R,j}^{\text{XOR}}\right). \tag{16.9}$$

The factor $1/2$ comes to show that the transmission is being completed in two transmission phases.

Let $Y_R \triangleq |\hat{h}_{j,R}|^2$ and $Z_R \triangleq |\hat{h}_{\acute{j},R}|^2$ $j, \acute{j} \in \{B, M\}$ with $j \neq \acute{j}$ are exponentially distributed random variables with parameters $\Omega_{jR}$ and $\Omega_{\acute{j}R}$, respectively. Now, the following conditions should satisfy for successful decoding of in the first phase

$$\begin{cases} \frac{1}{2}\log_2(1 + \phi_{j,R} Y_R) \geq r_{\acute{j}} \\ \frac{1}{2}\log_2(1 + \phi_{\acute{j},R} Z_R) \geq r_j \\ \frac{1}{2}\log_2(1 + \phi_{j,R} Y_R + \phi_{\acute{j},R} Z_R) \geq r_{\acute{j}} + r_j, \end{cases} \tag{16.10}$$

where $\phi_{j,R}$ and $\phi_{\acute{j},R}$ are defined after (16.12). Thereafter, based on the conditions given in (16.10), we can formulate $\Pr[Q_R]$ as

$$\Pr[Q_R] = \int\limits_{\frac{2^{3r_{\acute{j}}}-1}{\phi_{j,R}}}^{\infty} f_{Y_R}(y_R) \int\limits_{\frac{2^{3r_j}-1}{\phi_{\acute{j},R}}}^{\infty} f_{Z_R}(z_R)\mathrm{d}z_R\mathrm{d}y_R - \int\limits_{\frac{2^{3r_{\acute{j}}}-1}{\phi_{j,R}}}^{\mathcal{A}_R} f_{Y_R}(y_R) \int\limits_{\frac{2^{3r_j}-1}{\phi_{\acute{j},R}}}^{\mathcal{B}_R-\frac{\phi_{j,R}}{\phi_{\acute{j},R}}y_R} f_{Z_R}(z_R)\mathrm{d}z_R\mathrm{d}y_R, \tag{16.11}$$

The expression of $\Pr[Q_R]$ can be derived by following the similar steps as in Gurjar and Upadhyay (2017)

$$\Pr[Q_R] = \begin{cases} \mathcal{P}_{Q_R} & \text{for } \phi_{j,R}\Omega_{jR} \neq \phi_{\acute{j},R}\Omega_{\acute{j}R}, \\ \breve{\mathcal{P}}_{Q_R} & \text{for } \phi_{j,R}\Omega_{jR} = \phi_{\acute{j},R}\Omega_{\acute{j}R}, \end{cases} \tag{16.12}$$

where

$$\mathcal{P}_{Q_R} = \mathrm{e}^{-\left(\frac{2^{3r_j}-1}{\phi_{\acute{j},R}\Omega_{\acute{j}R}}+\frac{\mathcal{A}_R}{\Omega_{jR}}\right)} - \frac{\phi_{\acute{j},R}\Omega_{\acute{j}R}\mathrm{e}^{\frac{-\mathcal{B}_R}{\Omega_{\acute{j}R}}}}{\phi_{j,R}\Omega_{jR}-\phi_{\acute{j},R}\Omega_{\acute{j}R}} \times \left(\mathrm{e}^{\mathcal{A}_R\left(\frac{\phi_{j,R}\Omega_{jR}-\phi_{\acute{j},R}\Omega_{\acute{j}R}}{\Omega_{\acute{j}R}\Omega_{jR}\phi_{\acute{j},R}}\right)} - \mathrm{e}^{\left(\frac{\phi_{j,R}\Omega_{jR}-\phi_{\acute{j},R}\Omega_{\acute{j}R}}{\Omega_{\acute{j}R}\Omega_{jR}\phi_{j,R}}\right)\left(\frac{2^{3r_{\acute{j}}}-1}{\phi_{j,R}}\right)}\right), \tag{16.13}$$

and

$$\breve{\mathcal{P}}_{Q_R} = \mathrm{e}^{-\left(\frac{2^{3r_j}-1}{\phi_{\acute{j},R}\Omega_{\acute{j}R}}+\frac{\mathcal{A}_R}{\Omega_{jR}}\right)} + \frac{\mathrm{e}^{-\frac{\mathcal{B}_R}{\Omega_{\acute{j}R}}}}{\Omega_{jR}}\left(\mathcal{A}_R - \frac{2^{3r_{\acute{j}}}-1}{\phi_{j,R}}\right). \tag{16.14}$$

for $j, \acute{j} \in \{B, M\}$ with $j \neq \acute{j}$. Further $\mathcal{A}_R = (2^{3(r_{\acute{j}}+r_j)} - 2^{3r_j})/\phi_{j,R}$ and $\mathcal{B}_R = (2^{3(r_{\acute{j}}+r_j)} - 1)/\phi_{\acute{j},R}$. Further, we define $\phi_{j,R} = \eta_c\rho_{jR}^2/(\eta_c(1-\rho_{jR}^2)+1)$ and $\phi_{\acute{j},R} = \eta_c\rho_{\acute{j}R}^2/(\eta_c(1-\rho_{\acute{j}R}^2)+1)$ with $P_B = P_M = P$ and $\eta_c = P/\sigma^2$.

Further, $\mathcal{P}_{R,j}^{\mathrm{xor}} \triangleq \Pr[\mathcal{R}_{R,j}^{\mathrm{xor}} > r_j]$ can be formulated as

$$\begin{aligned} \mathcal{P}_{R,j}^{\mathrm{xor}} &= 1 - \Pr[\Upsilon_{R,j}^{\mathrm{xor}} < \gamma_j] \\ &= 1 - F_{\Upsilon_{R,j}^{\mathrm{xor}}}(\gamma_j), \end{aligned} \tag{16.15}$$

where $\gamma_j = 2^{2r_j} - 1$. The CDF $F_{\Upsilon^{\text{xor}}_{R,j}}(\gamma_j)$ can be expressed as

$$F_{\Upsilon^{\text{xor}}_{R,j}}(\gamma_j) = 1 - \mathrm{e}^{\frac{-\left(1+(1-\rho^2_{Rj})\eta_R\Omega_{Rj}\right)\gamma_j}{\rho^2_{Rj}\eta_R\Omega_{Rj}}}. \tag{16.16}$$

Hereafter, on invoking (16.12)–(16.16) in (16.8), one can get the desired OP for XOR-based DF strategy.

#### 16.6.3.3 OP for SUP-DF

The OP for SUP-DF can be defined as

$$\mathcal{P}^{\text{SUP}}_{\text{out},j} = 1 - \Pr[Q_R]\,\Pr[\mathcal{R}^{\text{SUP}}_{R,j} > r_j] \tag{16.17}$$

Here, the instantaneous achievable rates can be defined as

$$\mathcal{R}^{\text{SUP}}_{R,j} = \frac{1}{2}\log_2\left(1 + \Upsilon^{\text{SUP}}_{R,j}\right). \tag{16.18}$$

Further, $\mathcal{P}^{\text{SUP}}_{R,j} \triangleq \Pr[\mathcal{R}^{\text{SUP}}_{R,j} > r_j]$ can be formulated as

$$\begin{aligned}\mathcal{P}^{\text{SUP}}_{R,j} &= 1 - \Pr[\Upsilon^{\text{SUP}}_{R,j} < \gamma_j] \\ &= 1 - F_{\Upsilon^{\text{SUP}}_{R,j}}(\gamma_j),\end{aligned} \tag{16.19}$$

where $\gamma_j = 2^{2r_j} - 1$. The CDF $F_{\Upsilon^{\text{SUP}}_{R,j}}(\gamma_j)$ can be expressed as

$$F_{\Upsilon^{\text{SUP}}_{R,j}}(\gamma_j) = 1 - \mathrm{e}^{\frac{-2\left(1+(1-\rho^2_{Rj})\eta_R\Omega_{Rj}\right)\gamma_j}{\rho^2_{Rj}\eta_R\Omega_{Rj}}}. \tag{16.20}$$

On invoking (16.12)–(16.14) and (16.20) in (16.17), one can obtain the desired OP for SUP-based DF strategy.

#### 16.6.3.4 System Throughput

System throughput is another important performance metric for delay limited transmission. It can be defined as the average target rate achieved successfully at the communicating nodes. Hereby, system throughput can be formulated as

$$\mathfrak{R}_{st} = \frac{r_B}{2}\left(1 - \mathcal{P}^k_{\text{out},B}\right) + \frac{r_M}{2}\left(1 - \mathcal{P}^k_{\text{out},M}\right), \tag{16.21}$$

for $k \in \{\text{XOR}, \text{SUP}\}$.

### 16.6.4 Numerical Results

For numerical investigation, we define $\eta_c = \eta_R$ as the SNR and $\Omega_{BR} = 1, \Omega_{MR} = 2$. We set $f_c = 0.8$ GHz and the values of $\tau$ are assumed to be 1 ms. Figure 16.5 depicts the outage probability versus SNR curves for different relative velocities between communicating nodes. In this figure, the target rates are set as $r_B = r_M = 1/4$ bps/Hz. One can clearly observe that when nodes are stationary, i.e., $v_{BR} = v_{MR} = 0$ kmph, system shows the best outage performance among all the considered cases. When the relative velocities increase from 0 to 20 kmph, the outage probability gets saturated for higher values of SNR and diversity order reduces to zero. This is because of residual self-interference by which the receiving node suffers at the time of decoding the intended signal in the second phase. As we have considered two DF strategies in this paper, the performance of these schemes needs to be analyzed. From Fig. 16.5, it is evident that XOR-based DF outperforms the SUP-based DF. This is due to the fact that the symbol power is divided in case of SUP-DF to half as compared to XOR-DF. However, the receiver designing is complex for XOR-DF as compared to SUP-DF.

Figure 16.6 shows system throughput versus SNR curves for different target rates. In this figure, the relative velocities between communicating nodes are set as 20 kmph. From this figure, one can manifestly observe that system throughput is minimum at 0 dB and increases sharply for relatively higher values. After a certain value of SNR, it gets saturated. For lower SNR values, the system throughput is dominated by the

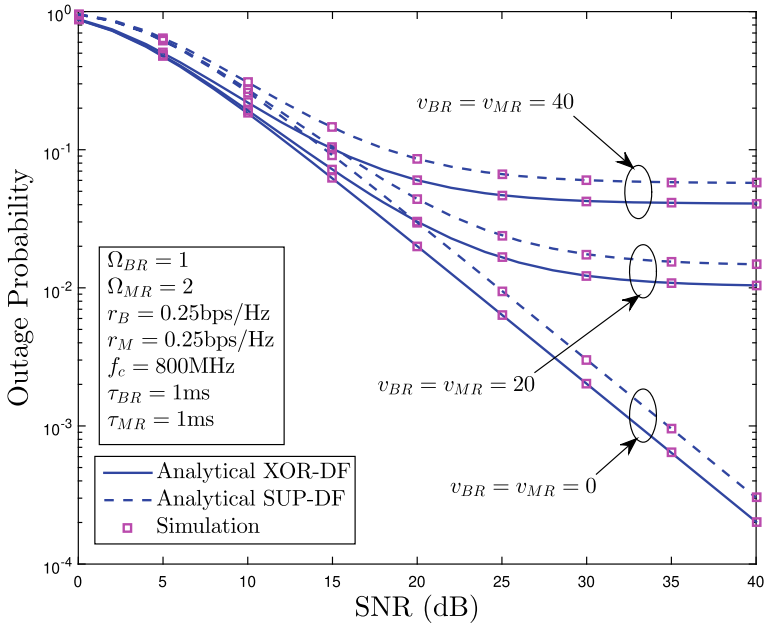



**Fig. 16.5** Outage probability versus SNR curves for different velocities

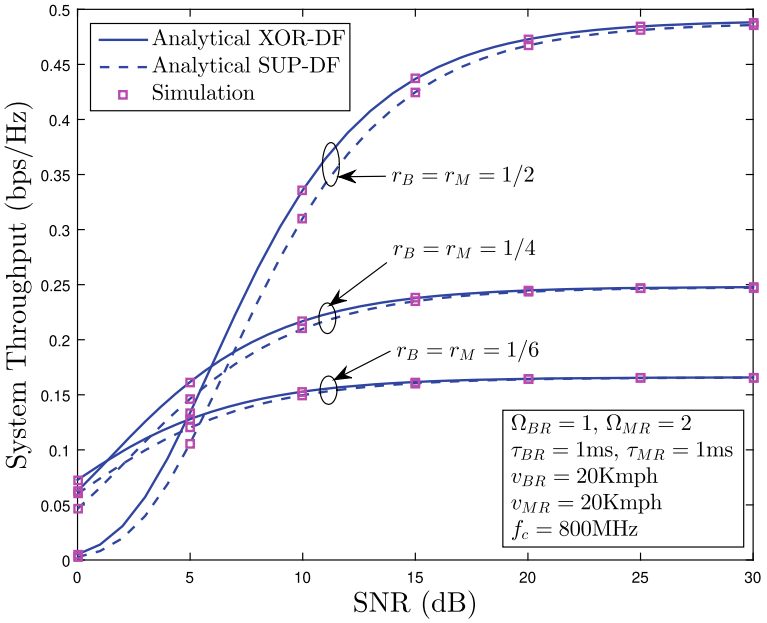



**Fig. 16.6** System throughput versus SNR curves for different target rates

outage probability of corresponding link, however, when the SNR value is sufficiently higher to make outage probability negligible as compared to unity, the target rate start playing an important role. This is the reason that after a certain SNR value, the improvement in system throughput is negligible. Moreover, the maximum achievable system throughput corresponds to the curve plotted for the rate $1/2$ bps/Hz, which is the highest among all.

## 16.7 Conclusion

Various modes of vehicular communications, V2V, V2I, V2P, and V2N, are discussed extensively. This chapter covered the recent state of art on various forms of V2X communications and identified possible research directions. Moreover, a typical V2N model is considered to analyze the system performance in terms of outage probability and system throughput by invoking nodes' mobility under Rayleigh fading environment. In this model, a relay node helps to realize communications between VBS and MV by following two strategies that are XOR-DF and SUP-DF for providing relay assistance. Among these two schemes, XOR-based DF outperformed SUP-based DF in terms of link reliability. Further, this chapter illustrated that the nodes-mobility deteriorate the system performance significantly. Numerical and simulation results confirm the accuracy of our mathematical findings.

## References

Arena F, Pau G (2019) An overview of vehicular communications. Future Internet 11(2):1–12
Bazzi A, Masini BM, Zanella A (2015) Performance analysis of V2V beaconing using LTE in direct mode with full duplex radios. IEEE Wirel Commun Lett 4(6):685–688
Fascista A, Ciccarese G, Coluccia A, Ricci G (2017) A localization algorithm based on V2I communications and AOA estimation. IEEE Sig Process Lett 24(1):126–130
Gomes P, Olaverri-Monreal C, Ferreira M (2012) Making vehicles transparent through V2V video streaming. IEEE Trans Intell Transp Syst 13(2):930–938
Gong Z, Jiang F, Li C (2019) Angle domain channel tracking with large antenna array for high mobility V2I millimeter wave communications. IEEE J Sel Top Sig Process 13(5):1077–1089
Gradshteyn I, Ryzhik I (2007) Table of integrals, series, and products, seventh, edition edn. Academic Press, San Diego
Gurjar DS, Upadhyay PK (2017) Overlay spectrum sharing for device-to-device communications in two-way cellular networks with nodes mobility. Trans Emerg Telecommun Technol 28(10):e3164–e3164 ett.3164
Gurjar DS, Upadhyay PK (2016a) Impact of channel estimation error on zero-forcing-based multiple-input-multiple-output two-way relaying. IET Sig Proc 10(3):210–217
Gurjar DS, Upadhyay PK (2016b) Performance analysis of zero-forcing-based multiple-input multiple-output two-way relaying in overlay device-to-device communications. IET Commun 10(6):699–708
Gurjar DS, Upadhyay PK (2018) Overlay device-to-device communications in asymmetric two-way cellular systems with hybrid relaying. IEEE Syst J 12(4):3713–3724
Han D, Bai B, Chen W (2017) Secure V2V communications via relays: resource allocation and performance analysis. IEEE Wirel Commun Lett 6(3):342–345
Jiang F, Li C, Gong Z (2018) Low complexity and fast processing algorithms for V2I massive mimo uplink detection. IEEE Trans Veh Technol 67(6):5054–5068
Kim B, Choi B, Park S, Kim H, Kim E (2016) Pedestrian/vehicle detection using a 2.5-D multi-layer laser scanner. IEEE Sens J 16(2):400–408
Lee Y, Chan Y, Fu L, Hsiao P (2015) Near-infrared-based nighttime pedestrian detection using grouped part models. IEEE Trans Intell Transp Syst 16(4):1929–1940
Li K, Wang X, Xu Y, Wang J (2016) Density enhancement-based long-range pedestrian detection using 3-D range data. IEEE Trans Intell Transp Syst 17(5):1368–1380
Makhoul G, D'Errico R, Oestges C (2019) Wideband measurement-based vehicle-to-pedestrian channel models. IEEE Trans Veh Technol 68(10):9339–9347
Merdrignac P, Shagdar O, Nashashibi F (2017) Fusion of perception and V2P communication systems for the safety of vulnerable road users. IEEE Trans Intell Transp Syst 18(7):1740–1751
Park Y, Kim T, Hong D (2019) Resource size control for reliability improvement in cellular-based V2V communication. IEEE Trans Veh Technol 68(1):379–392
Pyun S, Lee W, Cho D (2016) Resource allocation for vehicle-to-infrastructure communication using directional transmission. IEEE Trans Intell Transp Syst 17(4):1183–1188
Rashdan I, De Ponte Müller F, Jost T, Sand S, Caire G (2019) Large-scale fading characteristics and models for vehicle-to-pedestrian channel at 5-GHz. IEEE Access 7:107648–107658
Singh SK (2017) Road traffic accidents in India: issues and challenges. Transp Res Procedia 25:4708–4719
Soleimani H, Boukerche A (2018) On vehicular safety message transmissions through LTE-advanced networks. Ad Hoc Netw 79:133–145
Viriyasitavat W, Boban M, Tsai H, Vasilakos A (2015) Vehicular communications: survey and challenges of channel and propagation models. IEEE Veh Technol Mag 10(2):55–66
Wang CD, Thompson JP (U.S. Patent 5613039, 1997) Apparatus and method for motion detection and tracking of objects in a region for collision avoidance utilizing a real-time adaptive probabilistic neural network
WHO (2018) Global status report on road safety

Wu L, Zhang Z, Dang J, Wu Y (2017) Frequency-domain intergroup interference coordination for V2V communications. IEEE Sig Process Lett 24(11):1739–1743

Xu Q, Segupta R, Jiang D, Chrysler D (2003) Design and analysis of highway safety communication protocol in 5.9 GHz dedicated short range communication spectrum. In: 57th IEEE semiannual vehicular technology conference, 2003. VTC 2003-Spring., vol 4, pp 2451–2455

Yang Y, Fei D, Dang S (2017) Inter-vehicle cooperation channel estimation for IEEE 802.11p V2I communications. J Commun Netw 19(3):227–238

Ye H, Li GY, Juang BF (2019) Deep reinforcement learning based resource allocation for V2V communications. IEEE Trans Veh Technol 68(4):3163–3173

Yim S (2017) Preview controller design for vehicle stability with V2V communication. IEEE Trans Intell Transp Syst 18(6):1497–1506

Zhang R, Zhong Z, Zhao J, Li B, Wang K (2016) Channel measurement and packet-level modeling for V2I spatial multiplexing uplinks using massive mimo. IEEE Trans Veh Technol 65(10):7831–7843

Zhang F, Xi J, Langari R (2017) Real-time energy management strategy based on velocity forecasts using V2V and V2I communications. IEEE Trans Intell Transp Syst 18(2):416–430

# Chapter 17
# Physical Layer Security in Cooperative Vehicular Relay Networks


**Anshul Pandey and Suneel Yadav**

**Abstract** This chapter presents an integrated framework that draws on previous research, evaluates, and exposes the performance of physical layer security (PHY-security) to ensure secure communication in the cooperative vehicular relay networks. In particular, we consider a cooperative vehicular network scenario wherein a moving source vehicle wants to share confidential information with a fixed destination infrastructure terminal via a moving amplify-and-forward relay vehicle in the presence of a moving passive eavesdropper vehicle. Consequently, we model the vehicle-to-vehicle links by a $N$-Rayleigh fading channel, whereas the vehicle-to-fixed link undergoes Rayleigh fading. Thereafter, under such realistic channel modeling, we evaluate the system secrecy performance in terms of secrecy outage probability, non-zero secrecy capacity, intercept probability, secrecy diversity order, and ergodic secrecy capacity. We show that the eavesdropper has a detrimental effect on the secrecy performance and it reduces the secrecy diversity order to zero. Finally, numerical and simulation results are provided to corroborate our analytical findings.

**Keywords** Physical layer security · $N$-Rayleigh fading channels · Cooperative vehicular networks

## 17.1 Introduction

The rapid development and proliferation of autonomous intelligent vehicles with wireless communication capabilities resulted in the emergence of a plethora of novel intelligent transportation systems (ITS) applications, such as collision warning, traffic monitoring and management, and infotainment. (Hartenstein and Laberteaux 2008;

A. Pandey · S. Yadav (✉)
Department of Electronics and Communication Engineering,
Indian Institute of Information Technology Allahabad, Prayagraj 211015, India
e-mail: suneel@iiita.ac.in

A. Pandey
e-mail: rse2016503@iiita.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_17

Martinez et al. 2010; Yan and Rawat 2017). Thus, vehicular communication networks provide new venues for the accumulation and dissemination of the critical real-time data concerned with the drivers, vehicle functioning, environment, etc., accessed with the help of various installed sensory units. The dissemination of the information over such networks involves broadly four modes of communication, namely, vehicle-to-user, vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-Internet. Due to the versatility, heterogeneous nature, and seamless supporting applications, vehicular communication networks are envisioned to play a crucial part in the fifth generation (5G) and beyond communication networks to provide ubiquitous connectivity, ultra-reliable, and low latency transmissions (Shah et al. 2018). For instance, several leading economies in the world are gearing up themselves for these future vehicular networks, and according to a recent news South Korea has set up an entire city for the 5G autonomous vehicles (5gworldpro.com 2019). Moreover, cooperation via relaying terminals in such networks has emerged as a viable answer to improve the efficiency, reliability, and coverage of high mobility applications (Akin et al. 2015; Molisch et al. 2016). However, owing to the decentralized and ad-hoc nature, cooperativeness, and the broadcast wireless medium, these networks are exposed to several security threats (e.g., passive eavesdropping and jamming) and attacks (Parkinson et al. 2017; Sakiz and Sen 2017). Customarily, key-based cryptographic techniques are utilized at the upper layers to give information security. However, such key-based arrangements may not be reliable because of the presence of the eavesdroppers with infinite computing capabilities, lack of the trusted center for secret key distribution and management, and low bandwidth allocations in such networks (Zou et al. 2016). To this end, Physical layer security (PHY-security) has emerged as an attractive option to provide information security to secure wireless communication at the physical layer (ElHalawany et al. 2019; Yang et al. 2015; Yener and Ulukus 2015). PHY-security uses the inherent characteristics of the physical channel such as noise and interference to realize key-less secure transmissions through various coding and signal design and processing techniques. Moreover, PHY-security techniques operate independently of the existing key-based security solutions and can easily be integrated into the prevailing security infrastructure of the vehicular networks.

### *17.1.1 Literature Review*

PHY-security approaches have gained significant interest in the research community (Akhtar et al. 2014; Fan et al. 2014, 2016; Hoang et al. 2015; Kuhestani et al. 2018; Liu et al. 2015; Lv et al. 2017; Nguyen and Kim 2015; Niu et al. 2016; Sun et al. 2015; Wang et al. 2016; Yang et al. 2017; Zhao et al. 2018; Zou et al. 2013) owing to their associated advantages and their ability to operate independently with the available security infrastructure. In particular, the reported literature in Liu et al. (2015), Akhtar et al. (2014) has examined the security aspects in cooperative relaying networks under PHY-security. Furthermore, the authors have demonstrated the

benefits of exploiting multi-relay diversity in Zou et al. (2013), Nguyen and Kim (2015), multi-user diversity in Niu et al. (2016), Hoang et al. (2015), and the combined benefits of multi-user and multi-relay diversity in Fan et al. (2014), Fan et al. (2016) for secure communications under the passive eavesdropping scenarios. Further, the authors in Wang et al. (2016), Sun et al. (2015), Kuhestani et al. (2018) have evaluated the secrecy performance of cooperative relaying networks for the situation where the relaying terminal is untrusted. The authors in Lv et al. (2017) have used some external jamming signal techniques to enhance the system secrecy performance. Moreover, the authors in Yang et al. (2017) have exploited the advantages of multi-relay diversity for secure communications in the presence of an active eavesdropper. Further, PHY-security aspects have been extensively evaluated for the multiple-input-multiple-output systems in Kuhestani et al. (2018), Zhao et al. (2018). From the above discussion, we can infer that the field of wireless PHY-security has become rather quiet mature and numerous work have been addressed in the literature to characterize the PHY-security performance of cooperative relaying networks. However, all the above-mentioned works are suited for the situations where the communicating terminals are stationary, and hence the associated channels can be well characterized by the classical fading distributions, such as Rayleigh, Nakagami-$m$ fading. Moreover, the vehicular communication networks are characterized by both V2V (where both the transmitter and receiver terminals are moving) and V2I (where either of the communicating terminal is in motion) links. Both theoretically and experimentally, it has been established that the propagation environment for the V2V channels undergoes multiple scattering phenomenon (Andersen 2002; Salo et al. 2006). Such an effect of multiple scattering in the propagation environment can be realized by the cascaded channel fading models.

The performance evaluation of cooperative vehicular relay networks (CVRNs) considering cascaded fading models, such as double-Rayleigh, double-Nakagami-$m$ fading, have been extensively evaluated in the literature [see Seyfi et al. (2011), Ilhan and Akin (2012), Alghorani et al. (2016), Nguyen and Kong (2016); Ilhan et al. (2009) and the references therein]. However, none of the reported literature exploited the benefits of PHY-security in such networks. Moreover, the literature exploiting PHY-security in CVRNs is sparse. In particular, $K$th best relay selection strategy was proposed in Zhang and Pan (2017) to improve the secrecy performance of the cooperative vehicular networks under decode-and-forward relays with double-Rayleigh distributed V2V channels. Further, the authors in Pandey and Yadav (2018a, b) have evaluated the security performance of the CVRNs with the V2V and V2I channels modeled as Rayleigh and double-Rayleigh fading, respectively. Recently, the work in Pandey and Yadav (2020) has analyzed and optimized the secrecy performance of the CVRNs with the V2V and V2I channels modeled as double-Nakagami-$m$ fading and Nakagami-$m$ fading, respectively. The previous works although investigated the PHY-security aspects in CVRNs under double-Rayleigh and double-Nakagami-$m$ fading channels. However, for the V2V links, the propagation environment is dynamic and is not always necessary to undergo only two independent scattering. Therefore, it is interesting to investigate the PHY-security aspects in CVRNs by

modeling the V2V channels with a generalized cascaded fading environment, such as $N$-Rayleigh fading.

Motivated by the above works, this chapter considers a CVRN wherein a moving source vehicle wishes to send the confidential messages to some fixed destination infrastructure terminal (e.g., base station/roadside access point) via a moving amplify-and-forward relay vehicle. An eavesdropper vehicle exists in the network and tries to overhear the legitimate transmissions, thereby causing privacy and security threats. We model the V2V links by $N$-Rayleigh fading channel, whereas the vehicle-to-fixed link undergoes Rayleigh fading. Such a scenario can find its relevancy on the motor highways in the vehicular communication networks. Under such a realistic channel modeling, we evaluate the system secrecy performance by evaluating various key PHY-security metrics, viz., secrecy outage probability, intercept probability, and ergodic secrecy capacity. An asymptotic secrecy outage probability expression in the high signal-to-noise ratio (SNR) regime is also derived to gain key insights into the system achievable secrecy diversity order.

### *17.1.2 Organization of the Chapter*

The aim of this chapter is to shed a light upon the achievable performance of PHY-security in CVRNs. In particular, the remainder of the chapter is organized as follows. Sect. 17.2 describes the system and channel models under consideration. Section 17.3 evaluates the system performance in terms of various secrecy metrics, viz., secrecy outage probability, probability of non-zero secrecy capacity, intercept probability, and the ergodic secrecy capacity. We also deduce the high SNR approximation for the secrecy outage probability to gain more insights on the system secrecy performance. In Sect. 17.4, we present and discuss the key results on the system secrecy performance. Finally, we discuss the conclusions in Sect. 17.5.

## 17.2 System and Channel Models

We consider a system of terminals operating in a secure cooperative vehicular relaying network as shown in Fig. 17.1, wherein a source vehicle $S$ sends the confidential messages to a fixed infrastructure unit $D$ (say, base station/roadside access point) with the help of a relay vehicle $R$. Owing to simplicity and ease in designing, we employ amplify-and-forward relaying protocol at $R$. We assume an eavesdropper vehicle $E$ to exist in the close vicinity of $R$ with a malicious intent to intercept the secret information. All the communicating entities in the network are assumed to be embedded with a single antenna and the communication over the channel occurs in the half-duplex mode. The communication from $S$ to $D$ occurs in the two transmission phases. We assume $S \rightarrow D$ link is absent because of the large distance between the terminals and extreme shadowing environment. Moreover, $S \rightarrow E$ link is also assumed to be

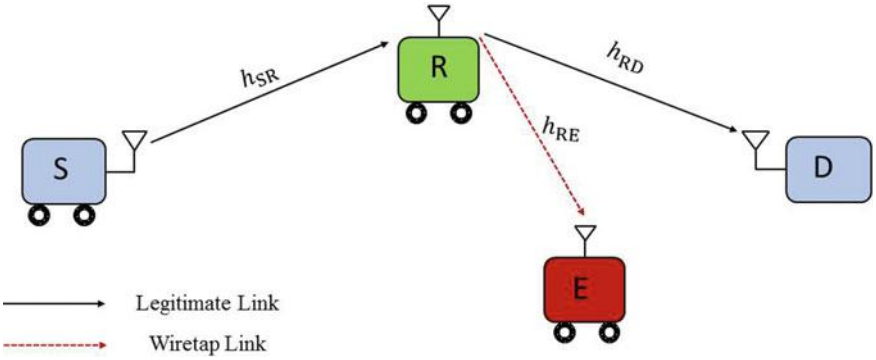



**Fig. 17.1** System model of the secure CVRN under consideration

unavailable as $E$ is outside the transmission range of first-hop transmission occurring from $S$, and hence making the first-hop information secure. However, $E$ is capable of overhearing the relayed transmissions, and thus causes security concerns to emerge. Furthermore, all the channels $S \rightarrow R$, $R \rightarrow D$, and $R \rightarrow E$ are assumed to be quasi-static, reciprocal, and subject to independent and non-identical fading distribution (i.ni.d.). Note that, we did not consider the effect of relative velocity between the two terminals, and assumed the channel to be time invariant for one fading block. Such an assumption is practically suited in rush hour traffic situations (Seyfi et al. 2011).

In particular, the channel between the two vehicular terminals $S \rightarrow R$ and $R \rightarrow E$ is modeled as $N$-Rayleigh fading, where $N$ represents the degree of cascading for the V2V channels. On the contrary, the channel for $R \rightarrow D$ link undergoes the classical Rayleigh fading. Moreover, the complex channel gains $h_{\iota\jmath}$ for $\{\iota\jmath\} \in \{SR, RE\}$ can be modeled as the product of $h_{\iota\jmath,1}, h_{\iota\jmath,2}, \cdots, h_{\iota\jmath,N}$, i.e., $h_{\iota\jmath} = h_{\iota\jmath,1} h_{\iota\jmath,2} h_{\iota\jmath,3} \times \cdots \times h_{\iota\jmath,N}$, where $h_{\iota\jmath,1}$ $h_{\iota\jmath,2}$, $\cdots$, and $h_{\iota\jmath,N}$ are independent complex Gaussian random variables (RVs) with zero mean and variance $\Omega_{\iota\jmath,1}$, $\Omega_{\iota\jmath,2}$, …, $\Omega_{\iota\jmath,N}$, respectively, whereas $h_{RD}$ is modeled as $\mathcal{CN}(0, \Omega_{RD})$. We also assume additive white Gaussian noise (AWGN) for each link which is modeled as $\mathcal{CN}(0, N_0)$.

In particular, $S$ requires two transmission phases to transmit the secret information to $D$. In the first transmission phase, $S$ uses a power $P_S$ to broadcasts the unit energy information signal $x_S$ to $R$, and thus the signal received at $R$ is given by

$$y_R = \sqrt{P_S} h_{SR} x_s + n_R, \tag{17.1}$$

where $n_R$ denotes the AWGN at the receiver $R$. The vehicle $R$ then with a variable gain $\mathcal{A} = \sqrt{\frac{P_R}{P_S |h_{SR}|^2 + N_0}}$, amplifies the received signal $y_R$ and further, broadcasts this amplified signal to both $D$ and $E$. The $P_R$ denotes the relay power. The signals arrived at $D$ and $E$ from $R$ are given, respectively, as

$$y_D = \mathcal{A} h_{RD} y_R + n_D, \tag{17.2}$$

$$y_E = \mathcal{A} h_{RE} y_R + n_E, \tag{17.3}$$

where $n_D$ and $n_E$ denote the AWGNs at the receivers $D$ and $E$, respectively. With the aid of (17.1)–(17.3), the received end-to-end instantaneous SNRs at the destination and eavesdropper can be expressed, respectively, as

$$\digamma_D = \frac{\Lambda_{SR}\Lambda_{RD}}{\Lambda_{SR} + \Lambda_{RD} + 1}, \tag{17.4}$$

$$\digamma_E = \frac{\Lambda_{SR}\Lambda_{RE}}{\Lambda_{SR} + \Lambda_{RE} + 1}, \tag{17.5}$$

where $\Lambda_{SR} = \frac{P_S}{N_0}|h_{SR}|^2$, $\Lambda_{RD} = \frac{P_R}{N_0}|h_{RD}|^2$, and $\Lambda_{RE} = \frac{P_R}{N_0}|h_{RE}|^2$. Consequently, the instantaneous capacities associated with the main and the eavesdropping channels can be written as $C_D = \frac{1}{2}\log_2(1 + \digamma_D)$ and $C_E = \frac{1}{2}\log_2(1 + \digamma_E)$, respectively.

## 17.3 Performance Analysis

In this section, we evaluate the various key secrecy metrics, viz., secrecy outage probability, probability of non-zero secrecy capacity, intercept probability, and ergodic secrecy capacity to evaluate the secrecy performance of the considered CVRNs. We also present and derive the asymptotic secrecy outage expression to gain some useful insights on the system secrecy performance.

### *17.3.1 Preliminaries*

It should be noted that the channel coefficients $h_{\iota\jmath}$ for $\{\iota\jmath\} \in \{SR, RE\}$ for the V2V links $S \to R$ and $R \to E$ are assumed to experience the $N$-Rayleigh fading, therefore the cumulative distribution function (CDF) and the probability distribution function (PDF) for the channel gain $|h_{\iota\jmath}|^2$ can be expressed as

$$F_{|h_{\iota\jmath}|^2}(y) = \mathcal{G}_{1,N+1}^{N,1}\left(\frac{y}{\Omega_{\iota\jmath}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right), \; y > 0, \tag{17.6}$$

$$f_{|h_{\iota\jmath}|^2}(y) = \frac{1}{y}\mathcal{G}_{0,N}^{N,0}\left(\frac{y}{\Omega_{\iota\jmath}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right), \; y > 0, \tag{17.7}$$

where $\Omega_{\iota_J} = \prod_{\kappa=1}^{N} \Omega_{\iota_J,\kappa}$, and $\mathcal{G}_{p,q}^{m,n}\left(x\middle|_{b_1,\ldots,b_q}^{a_1,\ldots,a_p}\right)$ is the Meijer-$\mathcal{G}$ function (Gradshteyn and Ryzhik 2000, Eq. (9.301)). In addition, the channel coefficient $h_{RD}$ for $R \to D$ link undergoes Rayleigh fading, therefore the channel gain $|h_{RD}|^2$ is exponentially distributed, whose CDF and PDF can be represented, respectively, as

$$F_{|h_{RD}|^2}(y) = 1 - \mathrm{e}^{-\frac{y}{\Omega_{RD}}}, \quad y > 0. \tag{17.8}$$

$$f_{|h_{RD}|^2}(y) = \frac{1}{\Omega_{RD}} \mathrm{e}^{-\frac{y}{\Omega_{RD}}}, \quad y > 0. \tag{17.9}$$

Now, the above-mentioned CDFs and PDFs will be used to evaluate the security performance of the considered secure CVRN in the sequel.

### *17.3.2 Secrecy Outage Probability (SOP)*

The SOP provides information regarding the legitimate transmissions compromised for security in the probabilistic sense by quantizing the secrecy outage event. In other words, it tells us what is the probability that the instantaneous secrecy capacity does not satisfy the desired secrecy threshold $\mathcal{R}_s$ (in bps/Hz) and falls below it. Consequently, it can be mathematically expressed as

$$\mathcal{P}_{\text{out}}^{\text{sec}} = \Pr\Big\{[C_D - C_E]^+ < \mathcal{R}_s\Big\}, \tag{17.10}$$

where $[x]^+$ represents $\max(0, x)$. Furthermore, from (17.10) we can observe that for the case when $C_D \leq C_E$, the system security is compromised i.e., $\mathcal{P}_{\text{out}}^{\text{sec}} = 1$. Therefore, we resort our SOP analysis for the case when $C_D > C_E$, and consequently (17.10) can be written as

$$\begin{aligned}
\mathcal{P}_{\text{out}}^{\text{sec}} &= \Pr\Big\{C_D - C_E < \mathcal{R}_s\Big\} \\
&= \Pr\left[\frac{1}{2}\log_2(1+\digamma_D) - \frac{1}{2}\log_2(1+\digamma_E) < \mathcal{R}_s\right] \\
&= \Pr\left[\frac{1+\digamma_D}{1+\digamma_E} < \eta\right],
\end{aligned} \tag{17.11}$$

where $\eta = 2^{2\mathcal{R}_s}$. Further, to make the analysis tractable, we use the well-known fact that the harmonic mean of two positive numbers can be upper bounded by the minimum of those two numbers (Fan et al. 2014; Pandey and Yadav 2018a). Therefore, we can obtain an upper bound (UB) for both $\digamma_D$ and $\digamma_E$ as $\digamma_D^{\text{UB}} < \min(\Lambda_{SR}, \Lambda_{RD})$ and $\digamma_E^{\text{UB}} < \min(\Lambda_{SR}, \Lambda_{RE})$, respectively. Consequently, a lower bound (LB) on SOP can be given by

$$\mathcal{P}_{\text{out}}^{\text{sec,LB}}(\eta) = \Pr\left[\frac{1+\widetilde{F}_D}{1+\widetilde{F}_E} < \eta\right], \tag{17.12}$$

where $\widetilde{F}_D = \min(\Lambda_{SR}, \Lambda_{RD})$ and $\widetilde{F}_E = \min(\Lambda_{SR}, \Lambda_{RE})$. Moreover, we can express (17.12) into its corresponding integral representation as

$$\mathcal{P}_{\text{out}}^{\text{sec,LB}}(\eta) = \int_0^\infty F_{\widetilde{F}_D}(\eta - 1 + \eta x) f_{\widetilde{F}_E}(x)\mathrm{d}x. \tag{17.13}$$

Before proceeding further, we note that the integral in (17.13) can be evaluated only when we have the knowledge of the CDF and the PDF of $\widetilde{F}_D$ and $\widetilde{F}_E$, respectively. Therefore, we firstly evaluate the CDF of $\widetilde{F}_D$ and the PDF of $\widetilde{F}_E$ as per the proposition given below.

**Proposition 1** *The CDF of $\widetilde{F}_D$ and the PDF of $\widetilde{F}_D$ can be expressed, respectively, as*

$$F_{\widetilde{F}_D}(x) = 1 - e^{-\frac{x}{\lambda_{RD}}} + e^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right), \tag{17.14}$$

$$\begin{aligned} f_{\widetilde{F}_E}(x) = {} & \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) + \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) \\ & - \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right) \\ & - \frac{1}{x}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right), \end{aligned} \tag{17.15}$$

*where* $\lambda_{SR} = \frac{P_S \Omega_{SR}}{N_0}$, $\lambda_{RD} = \frac{P_R \Omega_{RD}}{N_0}$, *and* $\lambda_{RE} = \frac{P_R \Omega_{RE}}{N_0}$.

***Proof*** The CDF of $\widetilde{F}_D$ can be represented as $F_{\widetilde{F}_D}(x) = \Pr[\min(\Lambda_{SR}, \Lambda_{RD}) < x] = F_{\Lambda_{SR}}(x) + F_{\Lambda_{RD}}(x) - F_{\Lambda_{SR}}(x)F_{\Lambda_{RD}}(x)$. Further, by invoking the CDFs of $\Lambda_{SR}$ and $\Lambda_{RD}$, and after simplifying, we can obtain the CDF of $\widetilde{F}_D$ as shown in (17.14). Likewise, $F_{\widetilde{F}_E}(x)$ can be expressed as

$$F_{\widetilde{F}_E}(x) = F_{\Lambda_{SR}}(x) + F_{\Lambda_{RE}}(x) - F_{\Lambda_{SR}}(x)F_{\Lambda_{RE}}(x). \tag{17.16}$$

Thenafter, we differentiate (17.16) with respect to $x$ to get the PDF of $\widetilde{F}_E$ as

$$f_{\widetilde{F}_E}(x) = f_{\Lambda_{SR}}(x) + f_{\Lambda_{RE}}(x) - f_{\Lambda_{SR}}(x)F_{\Lambda_{RE}}(x) - F_{\Lambda_{SR}}(x)f_{\Lambda_{RE}}(x). \tag{17.17}$$

Now, invoking the required CDFs and PDFs into (17.17), the obtained PDF of $\widetilde{F}_E$ is shown in (17.15).

Furthermore, inserting the CDF of $\widetilde{F}_D$ from (17.14) and the PDF of $\widetilde{F}_E$ from (17.15) into (17.13), and by carried out the required integrals, the closed-form SOP expression can be obtained in accordance with the below theorem.

**Theorem 1** *The closed-form expression for SOP can be expressed as*

$$\mathcal{P}_{out}^{sec,LB}(\eta) = \mathcal{I}_1 - \mathcal{I}_2(\eta) + \mathcal{I}_3(\eta), \tag{17.18}$$

*where*

$$\mathcal{I}_1 = 2 - \mathcal{G}_{N+1,N+1}^{N,N+1}\left(\frac{\lambda_{SR}}{\lambda_{RE}}\middle|\begin{matrix}1,\underbrace{0,\ldots,0}_{N}\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right) - \mathcal{G}_{N+1,N+1}^{N,N+1}\left(\frac{\lambda_{RE}}{\lambda_{SR}}\middle|\begin{matrix}1,\underbrace{0,\ldots,0}_{N}\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right), \tag{17.19}$$

$$\begin{aligned}\mathcal{I}_2(\eta) = e^{-\frac{\eta-1}{\lambda_{RD}}}\Bigg[&\mathcal{G}_{1,N}^{N,1}\left(\frac{\lambda_{RD}}{\eta\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) + \mathcal{G}_{1,N}^{N,1}\left(\frac{\lambda_{RD}}{\eta\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\\ &- \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\lambda_{RE}}{\lambda_{RD}},\frac{\lambda_{RE}}{\lambda_{SR}}\right)\\ &- \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\lambda_{SR}}{\lambda_{RD}},\frac{\lambda_{SR}}{\lambda_{RE}}\right)\Bigg],\end{aligned} \tag{17.20}$$

$$\begin{aligned}\mathcal{I}_3(\eta) = e^{-\frac{\eta-1}{\lambda_{RD}}}&\sum_{i=1}^{L} w_i\left(\frac{1}{x_i}\right)e^{-(\frac{\eta}{\lambda_{RD}}-1)x_i}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta x_i+\eta-1}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\\ &\times\Bigg[\mathcal{G}_{0,N}^{N,0}\left(\frac{x_i}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\left(1-\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\right)\\ &+\mathcal{G}_{0,N}^{N,0}\left(\frac{x_i}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\left(1-\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\right)\Bigg],\end{aligned} \tag{17.21}$$

*where* $\mathcal{G}_{p_1,q_1:p_2,q_2:p_3,q_3}^{m_1,n_1:n_2,m_2:n_3,m_3}\left(x,y\middle|\begin{matrix}a_1,\cdots,a_{p_1}\\ b_1,\cdots,b_{q_1}\end{matrix}\middle|\begin{matrix}c_1,\cdots,c_{p_2}\\ d_1,\cdots,d_{q_2}\end{matrix}\middle|\begin{matrix}e_1,\cdots,e_{p_3}\\ f_1,\cdots,f_{q_3}\end{matrix}\right)$ *is the extended generalized bivariate Meijer-*$\mathcal{G}$ *function (EGBMGF) (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)), and* $\mathcal{L}_L(x)$ *denotes the Laguerre polynomial with* $w_i = \frac{x_i}{(L+1)^2[\mathcal{L}_{L+1}(x_i)]^2}$ *and* $x_i$ *as the weight factor and i-th root of Laguerre polynomial (Abromowitz and Stegun 1970). The EGBMGF can be implemented using MATHEMATICA software which can be evaluated as given in Ansari et al. (2011).*

***Proof*** To evaluate the SOP in (17.18), we first invoke $F_{\widetilde{\Gamma}_D}(x)$ of (17.14) and $f_{\widetilde{\Gamma}_D}(x)$ of (17.15) into (17.13), and consequently, we can express $\mathcal{I}_1$, $\mathcal{I}_2(\eta)$, and $\mathcal{I}_3(\eta)$, respectively, as

$$\begin{aligned}\mathcal{I}_1 = &\underbrace{\int_0^\infty \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{11}} + \underbrace{\int_0^\infty \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{12}}\\ &\underbrace{-\int_0^\infty \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1,0}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{13}}\\ &\underbrace{-\int_0^\infty \frac{1}{x}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1,0}_{N}\end{matrix}\right)\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{14}},\end{aligned} \tag{17.22}$$

$$\begin{aligned}\mathcal{I}_2(\eta) = &\underbrace{\int_0^\infty \frac{1}{x}\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{21}(\eta)}\\ &\underbrace{+\int_0^\infty \frac{1}{x}\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{22}(\eta)}\\ &\underbrace{-\int_0^\infty \frac{1}{x}\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1,0}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq\mathcal{I}_{23}(\eta)}\end{aligned}$$

$$\underbrace{-\int_0^{\infty} \frac{1}{x} \mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathrm{d}x}_{\triangleq \mathcal{I}_{24}(\eta)}, \tag{17.23}$$

$$\text{and} \quad \mathcal{I}_3(\eta) = \mathcal{I}_{31}(\eta) + \mathcal{I}_{32}(\eta) - \mathcal{I}_{33}(\eta) - \mathcal{I}_{34}(\eta), \tag{17.24}$$

where

$$\mathcal{I}_{31}(\eta) = \int_0^{\infty} \frac{\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}}{x} \mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta-1+\eta x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathrm{d}x,$$

$$\mathcal{I}_{32}(\eta) = \int_0^{\infty} \frac{\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}}{x} \mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta-1+\eta x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathrm{d}x,$$

$$\mathcal{I}_{33}(\eta) = \int_0^{\infty} \frac{\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}}{x} \mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)$$
$$\times \mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta-1+\eta x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathrm{d}x,$$

$$\mathcal{I}_{34}(\eta) = \int_0^{\infty} \frac{\mathrm{e}^{-\frac{\eta-1+\eta x}{\lambda_{RD}}}}{x} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)$$
$$\times \mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta-1+\eta x}{\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathrm{d}x.$$

The integrals $\mathcal{I}_{11}$ and $\mathcal{I}_{12}$ can be evaluated by using the fact (Gradshteyn and Ryzhik 2000, Eq. (7.811.4)) as $\mathcal{I}_{11} = \mathcal{I}_{12} = 1$. Moreover, with the aid of the identity (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)) and further carrying out some simplifications, the integrals $\mathcal{I}_{13}$ and $\mathcal{I}_{14}$ can be obtained. Consequently, using the results of the integrals $\mathcal{I}_{11}$, $\mathcal{I}_{12}$, $\mathcal{I}_{13}$, and $\mathcal{I}_{14}$, we can obtain the integral term $\mathcal{I}_1$ as shown in (17.19).

To evaluate $\mathcal{I}_2(\eta)$, we need to simplify $\mathcal{I}_{21}(\eta)$, $\mathcal{I}_{22}(\eta)$, $\mathcal{I}_{23}(\eta)$, and $\mathcal{I}_{24}(\eta)$. $\mathcal{I}_{21}(\eta)$ can be simplified by applying the fact (Gradshteyn and Ryzhik 2000, Eq. (7.813.1)) as

$$\mathcal{I}_{21}(\eta) = \mathrm{e}^{-\frac{\eta-1}{\lambda_{RD}}} \mathcal{G}_{1,N}^{N,1}\left(\frac{\lambda_{RD}}{\eta\lambda_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right). \tag{17.25}$$

Likewise, $\mathcal{I}_{22}(\eta)$ can be expressed as

$$\mathcal{I}_{22}(\eta) = \mathrm{e}^{-\frac{\eta-1}{\lambda_{RD}}} \mathcal{G}_{1,N}^{N,1}\left(\frac{\lambda_{RD}}{\eta\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right). \tag{17.26}$$

Further, to obtain $\mathcal{I}_{23}(\eta)$, we first apply the fact $\mathrm{e}^{-x} = \mathcal{G}_{0,1}^{1,0}\left(x\middle|\begin{matrix}0\\ 0\end{matrix}\right)$ (Prudnikov et al. 1990, Eq. (8.4.3.1)), and hence $\mathcal{I}_{23}(\eta)$ can be expressed as

$$\begin{aligned}\mathcal{I}_{23}(\eta) = \mathrm{e}^{-\frac{\eta-1}{\lambda_{RD}}} \int_0^\infty \frac{1}{x} \mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\lambda_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) \mathcal{G}_{0,1}^{1,0}\left(\frac{\eta x}{\lambda_{RD}}\middle|\begin{matrix}0\\ 0\end{matrix}\right) \\ \times \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right) \mathrm{d}x,\end{aligned} \tag{17.27}$$

which can be solved by using the identity (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)) as

$$\mathcal{I}_{23}(\eta) = \mathrm{e}^{-\frac{\eta-1}{\lambda_{RD}}} \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\lambda_{RE}}{\lambda_{RD}}, \frac{\lambda_{RE}}{\lambda_{SR}}\right). \tag{17.28}$$

Similarly, adopting the same approach as used to obtain (17.28), $\mathcal{I}_{24}(\eta)$ can be evaluated and expressed as

$$\mathcal{I}_{24}(\eta) = \mathrm{e}^{-\frac{\eta-1}{\lambda_{RD}}} \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\lambda_{SR}}{\lambda_{RD}}, \frac{\lambda_{SR}}{\lambda_{RE}}\right). \tag{17.29}$$

Now, with the help of (17.25), (17.26), (17.28), and (17.29), we can obtain $\mathcal{I}_2(\eta)$ as depicted in (17.20).

Moreover, it can be seen that the integral terms $\mathcal{I}_{31}(\eta)$ and $\mathcal{I}_{32}(\eta)$ involved in $\mathcal{I}_3(\eta)$ contain the product of an exponential function, algebraic term, a Meijer-$\mathcal{G}$ function, and a Meijer-$\mathcal{G}$ function with shifted arguments. The exact closed-form solution for such an integral is tedious and intractable. Therefore, to make the evaluation of the corresponding integrals tractable, we first divide and multiply $\mathcal{I}_{31}(\eta)$ and $\mathcal{I}_{32}(\eta)$ by $\mathrm{e}^x$ and then use the Gauss-Laguerre quadrature numerical integration method to obtain the solution for both $\mathcal{I}_{31}(\eta)$ and $\mathcal{I}_{32}(\eta)$.

Furthermore, the integrals $\mathcal{I}_{33}(\eta)$ and $\mathcal{I}_{34}(\eta)$ involve the product of an exponential function, algebraic term, two Meijer-$\mathcal{G}$ functions, and a Meijer-$\mathcal{G}$ function with shifted arguments. Again to the best of author's knowledge, the solution for such an integral is tedious and intractable. Therefore, we again use the Gauss-Laguerre quadrature numerical integration method to obtain the solution for both $\mathcal{I}_{33}(\eta)$ and $\mathcal{I}_{34}(\eta)$. Using the solutions obtained for the integrals $\mathcal{I}_{31}(\eta)$, $\mathcal{I}_{32}(\eta)$, $\mathcal{I}_{33}(\eta)$, and $\mathcal{I}_{34}(\eta)$, we can express $\mathcal{I}_3(\eta)$ as shown in (17.21).

Moreover, invoking the results obtained in (17.19)–(17.21) into (17.18), a lower bound on the SOP can be expressed as shown in (17.18).

### 17.3.3 Asymptotic SOP

It is difficult to obtain important insights from the SOP expression presented in (17.18). Therefore, to gain the valuable insights on the impact of the system/channel parameters on the system secrecy diversity order, we present an asymptotic SOP analysis in the high SNR regime. Let $P_S = P_R = P$ and define $\rho = \frac{P}{N_0}$ as the SNR. Consequently, we have $\lambda_{SR} = \rho\Omega_{SR}$, $\lambda_{RD} = \rho\Omega_{RD}$, and $\lambda_{RE} = \rho\Omega_{RE}$. Moreover, we can write $\widetilde{F}_D = \rho\Psi_D$ and $\widetilde{F}_E = \rho\Psi_E$, where $\Psi_D \triangleq \min\left(|h_{SR}|^2, |h_{RD}|^2\right)$ and $\Psi_E \triangleq \min\left(|h_{SR}|^2, |h_{RE}|^2\right)$. Therefore, the asymptotic SOP in the high SNR regime ($\rho \to \infty$) can be given by

$$\begin{aligned}\widetilde{\mathcal{P}}_{\text{out}}^{\text{sec}}(\eta) &= \Pr\left[\frac{1+\Psi_D}{1+\Psi_E} < \eta\right] \approx \Pr\left[\frac{\Psi_D}{\Psi_E} < \eta\right] \\ &= \int_0^\infty F_{\Psi_D}(\eta x) f_{\Psi_E}(x)\mathrm{d}x. \end{aligned} \tag{17.30}$$

Moreover, to carry out the integral in (17.30), we need to know the expressions of the CDF of $\Psi_D$ and the PDF of $\Psi_E$. With the aid of Proposition 1, we can re-write the CDF and the PDF of $\Psi_D$ and $\Psi_E$, respectively, as

$$F_{\Psi_D}(x) = 1 - \mathrm{e}^{-\frac{x}{\Omega_{RD}}} + \mathrm{e}^{-\frac{x}{\Omega_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\Omega_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right). \tag{17.31}$$

$$f_{\Psi_E}(x) = \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\Omega_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) + \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\Omega_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)$$
$$- \frac{1}{x}\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\Omega_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\Omega_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)$$
$$- \frac{1}{x}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\Omega_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathcal{G}_{0,N}^{N,0}\left(\frac{x}{\Omega_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right). \tag{17.32}$$

Now, invoking the CDF of $\Psi_D$ from (17.31) and the PDF of $\Psi_E$ from (17.32) into (17.30), and thereafter by carrying out the required integrals with the help of the identities (Gradshteyn and Ryzhik 2000, Eq. (7.811.4)), (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)), (Gradshteyn and Ryzhik 2000, Eq. (1.211.1)), and Guass-Laguerre quadrature integration method, we obtain

$$\widetilde{\mathcal{P}}_{\text{out}}^{\text{sec}}(\eta) = 2 - \mathcal{G}_{N+1,N+1}^{N,N+1}\left(\frac{\Omega_{SR}}{\Omega_{RE}}\middle|\begin{matrix}1,\underbrace{0,\ldots,0}_{N}\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right) - \mathcal{G}_{N+1,N+1}^{N,N+1}\left(\frac{\Omega_{RE}}{\Omega_{SR}}\middle|\begin{matrix}1,\underbrace{0,\ldots,0}_{N}\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)$$
$$- \mathrm{e}^{-\frac{\eta-1}{\Omega_{RD}}}\left[\mathcal{G}_{1,N}^{N,1}\left(\frac{\Omega_{RD}}{\eta\Omega_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right) + \mathcal{G}_{1,N}^{N,1}\left(\frac{\Omega_{RD}}{\eta\Omega_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\right.$$
$$- \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\Omega_{RE}}{\Omega_{RD}},\frac{\Omega_{RE}}{\Omega_{SR}}\right)$$
$$\left. - \mathcal{G}_{N+1,1:0,1:0,N}^{1,N:1,0:N,0}\left(\begin{matrix}\underbrace{0,\ldots,0}_{N},1\\ 0\end{matrix}\middle|\begin{matrix}-\\ 0\end{matrix}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\middle|\frac{\eta\Omega_{SR}}{\Omega_{RD}},\frac{\Omega_{SR}}{\Omega_{RE}}\right)\right] + \mathrm{e}^{-\frac{\eta-1}{\Omega_{RD}}}\sum_{i=1}^{L} w_i\left(\frac{1}{x_i}\right)$$
$$\times \mathrm{e}^{-\left(\frac{\eta}{\Omega_{RD}}-1\right)x_i}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{\eta x_i+\eta-1}{\Omega_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\left[\mathcal{G}_{0,N}^{N,0}\left(\frac{x_i}{\Omega_{SR}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\right.$$
$$\times\left(1 - \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\Omega_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\right) + \mathcal{G}_{0,N}^{N,0}\left(\frac{x_i}{\Omega_{RE}}\middle|\begin{matrix}-\\ \underbrace{1,\ldots,1}_{N}\end{matrix}\right)\left(1\right.$$
$$\left.\left. - \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\Omega_{SR}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\right)\right]. \tag{17.33}$$

From (17.33), it can be observed that the high SNR approximation for the SOP is independent of the transmit SNR ($\rho$). Thus, there exists a ceiling for the asymptotic SOP in the high SNR regime, and thus the secrecy diversity order of the system becomes zero. This is because of the reason that the SNR of the wiretap link is also improved simultaneously along with the SNR of legitimated link.

### *17.3.4 Probability of Non-zero Secrecy Capacity*

The probability of non-zero secrecy capacity is a key metric used to quantify the secure transmission as long as the main channel is better than the eavesdropper channel of cooperative relaying networks. We can represent the probability of non-zero secrecy capacity mathematically as

$$\begin{aligned}\mathcal{P}_{\mathrm{NZ}}^{\mathrm{SEC}} &= \Pr\{C_D > C_E\} \\ &= \Pr\left[\frac{1}{2}\log_2(1+\Gamma_D) > \frac{1}{2}\log_2(1+\Gamma_E)\right] \\ &= \Pr\left[\frac{\Lambda_{SR}\Lambda_{RD}}{\Lambda_{SR}+\Lambda_{RD}+1} > \frac{\Lambda_{SR}\Lambda_{RE}}{\Lambda_{SR}+\Lambda_{RE}+1}\right]. \end{aligned} \tag{17.34}$$

Moreover, using the widely used approximation $\frac{XY}{X+Y+1} \simeq \frac{XY}{X+Y}$ (Pandey and Yadav 2018a, b), and after some manipulations, the expression in (17.34) can be given as

$$\mathcal{P}_{\mathrm{NZ}}^{\mathrm{SEC}} = \Pr\left[\Lambda_{RE} < \Lambda_{RD}\right]. \tag{17.35}$$

Now, expressing (17.35) into its equivalent integral form and then after inserting the CDF of $\Lambda_{RE}$ and the PDF of $\Lambda_{RD}$, we obtain

$$\mathcal{P}_{\mathrm{NZ}}^{\mathrm{SEC}} = \frac{1}{\lambda_{RD}}\int_0^{\infty} \mathrm{e}^{-\frac{x}{\lambda_{RD}}}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|\begin{matrix}1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right)\mathrm{d}x. \tag{17.36}$$

Furthermore, the integral in (17.36) can be simplified with the aid of the identity (Gradshteyn and Ryzhik 2000, Eq. (7.813.1)), and hence the closed-form expression for the probability of non-zero secrecy capacity can be expressed as

$$\mathcal{P}_{\mathrm{NZ}}^{\mathrm{SEC}} = \mathcal{G}_{2,N+1}^{N,2}\left(\frac{\lambda_{RD}}{\lambda_{RE}}\middle|\begin{matrix}0,1\\ \underbrace{1,\ldots,1}_{N},0\end{matrix}\right). \tag{17.37}$$

### 17.3.5 Intercept Probability

When the legitimated link capacity falls below the wiretap link capacity, the eavesdropper can intercept the source signal and an intercept event occurs. Therefore, the probability associated with the intercept event is called as intercept probability, which is mathematically represented as

$$\begin{aligned}\mathcal{P}_{\text{int}} &= \mathbf{Pr}[C_D - C_E < 0] \\ &= 1 - \mathbf{Pr}[\digamma_E < \digamma_D].\end{aligned} \tag{17.38}$$

Moreover, applying the widely used and tight approximation $\frac{XY}{X+Y+1} \simeq \frac{XY}{X+Y}$ (Pandey and Yadav 2018a, b), and after some simplifications, the expression in (17.38) can be represented as

$$\mathcal{P}^{\text{int}} = 1 - \frac{1}{\lambda_{RD}} \int_0^{\infty} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{RE}} \middle|_{\underbrace{1,\ldots,1}_{N},0}^{1} \right) \mathrm{d}x. \tag{17.39}$$

Consequently, after solving the integral in (17.39) with the help of the fact (Gradshteyn and Ryzhik 2000, Eq. (7.813.1)), we can obtain the expression for the intercept probability as

$$\mathcal{P}_{\text{NZ}}^{\text{SEC}} = 1 - \mathcal{G}_{2,N+1}^{N,2}\left( \frac{\lambda_{RD}}{\lambda_{RE}} \middle|_{\underbrace{1,\ldots,1}_{N},0}^{0,1} \right). \tag{17.40}$$

### 17.3.6 Ergodic Secrecy Capacity (ESC)

The instantaneous secrecy capacity of the considered cooperative vehicular relay networks can be given by

$$\begin{aligned}C_s &= C_D - C_E \\ &= \frac{1}{2}\log_2(1 + \widetilde{\digamma}_D) - \frac{1}{2}\log_2(1 + \widetilde{\digamma}_E).\end{aligned} \tag{17.41}$$

The ESC gives the information about the rate below which any average secure communication (i.e., the maximum secrecy rate at which the message can be transmitted by source node and recovered reliably at legitimated destination, while keeping it useless and unrecoverable at malicious terminal) is feasible. In particular, the ESC can be obtained by averaging the instantaneous secrecy capacity expression over the

distributions of the end-to-end SNRs $\widetilde{F}_D$ and $\widetilde{F}_E$. Consequently, the ESC can be mathematically represented as (Ai et al. 2018)

$$\overline{C}_s = \frac{1}{2\ln(2)} \int_0^{\infty} \frac{F_{\widetilde{F}_E(x)}}{1+x} [1 - F_{\widetilde{F}_D(x)}] \mathrm{d}x. \tag{17.42}$$

Now, invoking the CDF of $\widetilde{F}_D$ into (17.42) and after some simplifications, we have

$$\begin{aligned}
\overline{C}_s =& \underbrace{\frac{1}{2\ln(2)} \int_0^{\infty} \frac{F_{\widetilde{F}_E(x)}}{1+x} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathrm{d}x}_{\triangleq \overline{C}_1} \\
& \underbrace{- \frac{1}{2\ln(2)} \int_0^{\infty} \frac{F_{\widetilde{F}_E(x)}}{1+x} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{SR}} \middle| \begin{matrix} 1 \\ \underbrace{1, \ldots, 1}_{N}, 0 \end{matrix} \right) \mathrm{d}x}_{\triangleq \overline{C}_2} \\
=& \frac{1}{2\ln(2)} [\overline{C}_1 - \overline{C}_2].
\end{aligned} \tag{17.43}$$

Thereafter, from (17.43) we observe that to obtain the solution for $\overline{C}_s$, we need to evaluate the integrals involved with both $\overline{C}_1$ and $\overline{C}_2$. To this end, we first expand the integrals involved in $\overline{C}_1$ and express it as

$$\begin{aligned}
\overline{C}_1 =& \underbrace{\int_0^{\infty} \frac{1}{1+x} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{SR}} \middle| \begin{matrix} 1 \\ \underbrace{1, \ldots, 1}_{N}, 0 \end{matrix} \right) \mathrm{d}x}_{\triangleq \overline{C}_{11}} \\
& + \underbrace{\int_0^{\infty} \frac{1}{1+x} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{RE}} \middle| \begin{matrix} 1 \\ \underbrace{1, \ldots, 1}_{N}, 0 \end{matrix} \right) \mathrm{d}x}_{\triangleq \overline{C}_{12}} \\
& \underbrace{- \int_0^{\infty} \frac{1}{1+x} \mathrm{e}^{-\frac{x}{\lambda_{RD}}} \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{SR}} \middle| \begin{matrix} 1 \\ \underbrace{1, \ldots, 1}_{N}, 0 \end{matrix} \right) \mathcal{G}_{1,N+1}^{N,1}\left( \frac{x}{\lambda_{RE}} \middle| \begin{matrix} 1 \\ \underbrace{1, \ldots, 1}_{N}, 0 \end{matrix} \right) \mathrm{d}x.}_{\triangleq \overline{C}_{13}}
\end{aligned} \tag{17.44}$$

To evaluate the integrals $\overline{C}_{11}$ and $\overline{C}_{12}$ in (17.44), we first apply the facts $\frac{1}{(1+x)^{\mu}} = \frac{1}{\Gamma(\mu)}\mathcal{G}_{1,1}^{1,1}\left(x\middle|\begin{matrix}1-\mu\\0\end{matrix}\right)$ (Prudnikov et al. 1990, Eq. (8.4.2.5)), $\mathrm{e}^{-x} = \mathcal{G}_{0,1}^{1,0}\left(x\middle|\begin{matrix}-\\0\end{matrix}\right)$ (Prudnikov et al. 1990, Eq. (8.4.3.1)), and thereafter solving the resulting integrals with the aid of the identity (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)) to obtain

$$\overline{C}_{11} = \lambda_{SR}\mathcal{G}_{N+1,1:0,1:1,1}^{1,N:1,0:1,1}\left(\begin{matrix}\underbrace{-1,\ldots,-1}_{N},0\\-1\end{matrix}\middle|\begin{matrix}-\\0\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\frac{\lambda_{SR}}{\lambda_{RD}},\lambda_{SR}\right). \tag{17.45}$$

$$\overline{C}_{12} = \lambda_{RE}\mathcal{G}_{N+1,1:0,1:1,1}^{1,N:1,0:1,1}\left(\begin{matrix}\underbrace{-1,\ldots,-1}_{N},0\\-1\end{matrix}\middle|\begin{matrix}-\\0\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\frac{\lambda_{RE}}{\lambda_{RD}},\lambda_{RE}\right). \tag{17.46}$$

Furthermore, to solve the integral $\overline{C}_{13}$ in (17.44), we first use the facts $\frac{1}{1+x} = \mathcal{G}_{1,1}^{1,1}\left(x\middle|\begin{matrix}0\\0\end{matrix}\right)$ (Prudnikov et al. 1990, Eq. (8.4.2.5)) and $\mathrm{e}^{-ax} = \sum_{j=0}^{\infty}\frac{(-ax)^j}{j!}$ (Gradshteyn and Ryzhik 2000, Eq. (1.211.1)), and then with the help of the identity (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)), and therefore $\overline{C}_{13}$ in (17.44) can be given as

$$\begin{aligned}\overline{C}_{13} = &\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\frac{(\lambda_{SR})^{k+1}}{(\lambda_{RD})^k}\\ &\times G_{N+1,1:0,1:1,N+1}^{1,N:1,0:N,1}\left(\begin{matrix}\underbrace{-1-k,\ldots,-1-k}_{N},-k\\-1-k\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\begin{matrix}1\\\underbrace{1,\ldots,1}_{N},0\end{matrix}\middle|\frac{\lambda_{SR}}{\lambda_{RD}},\lambda_{SR}\right).\end{aligned} \tag{17.47}$$

Consequently, invoking (17.45)–(17.47) into (17.44), we can express $\overline{C}_1$ as

$$\begin{aligned}\overline{C}_1 = &\lambda_{SR}\mathcal{G}_{N+1,1:0,1:1,1}^{1,N:1,0:1,1}\left(\begin{matrix}\underbrace{-1,\ldots,-1}_{N},0\\-1\end{matrix}\middle|\begin{matrix}-\\0\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\frac{\lambda_{SR}}{\lambda_{RD}},\lambda_{SR}\right)\\ &+\lambda_{RE}\mathcal{G}_{N+1,1:0,1:1,1}^{1,N:1,0:1,1}\left(\begin{matrix}\underbrace{-1,\ldots,-1}_{N},0\\-1\end{matrix}\middle|\begin{matrix}-\\0\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\frac{\lambda_{RE}}{\lambda_{RD}},\lambda_{RE}\right)-\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\frac{(\lambda_{SR})^{k+1}}{(\lambda_{RD})^k}\\ &\times\mathcal{G}_{N+1,1:0,1:1,N+1}^{1,N:1,0:N,1}\left(\begin{matrix}\underbrace{-1-k,\ldots,-1-k}_{N},-k\\-1-k\end{matrix}\middle|\begin{matrix}0\\0\end{matrix}\middle|\begin{matrix}1\\\underbrace{1,\ldots,1}_{N},0\end{matrix}\middle|\frac{\lambda_{SR}}{\lambda_{RD}},\lambda_{SR}\right).\end{aligned} \tag{17.48}$$

Moreover, we can now express the integrals involved with $\overline{C}_2$ as

$$\overline{C}_2 = \overline{C}_{21} + \overline{C}_{22} - \overline{C}_{23}, \tag{17.49}$$

where

$$\overline{C}_{21} = \int_0^\infty \frac{\mathrm{e}^{-\frac{x}{\lambda_{RD}}}}{1+x} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathrm{d}x, \tag{17.50}$$

$$\overline{C}_{22} = \int_0^\infty \frac{\mathrm{e}^{-\frac{x}{\lambda_{RD}}}}{1+x} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathrm{d}x, \tag{17.51}$$

$$\begin{aligned}\overline{C}_{23} = \int_0^\infty \frac{\mathrm{e}^{-\frac{x}{\lambda_{RD}}}}{1+x} \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{RE}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \\ \times \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x}{\lambda_{SR}}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\right) \mathrm{d}x.\end{aligned} \tag{17.52}$$

To the solve the integrals $\overline{C}_{21}$ and $\overline{C}_{22}$, we first use the facts that $\frac{1}{1+x} = \mathcal{G}_{1,1}^{1,1}\left(x\middle|_{0}^{0}\right)$ (Prudnikov et al. 1990, Eq. (8.4.2.5)) and $\mathrm{e}^{-ax} = \sum_{j=0}^{\infty} \frac{(-ax)^j}{j!}$ (Gradshteyn and Ryzhik 2000, Eq. (1.211.1)), and thereafter using the identity (The Wolfram Functions Site 2020, Eq. (07.34.21.0081.01)), we can obtain both the integrals $\overline{C}_{21}$ and $\overline{C}_{22}$. Further, we observe that the integral in $\overline{C}_{23}$ involves the product of three Meijer-$\mathcal{G}$ functions, an exponential term, and an algebraic term. However, to the best of author's knowledge, the exact closed-form solution for such an integral does not exist. Therefore, to make the analysis tractable, we first divide and multiply it with an exponential term $\mathrm{e}^x$, and then use the Gauss-Laguerre quadrature integration method to obtain $\overline{C}_{23}$. Consequently, the results obtained for $\overline{C}_{21}$, $\overline{C}_{22}$, and $\overline{C}_{23}$ into (17.49), we can express $\overline{C}_2$ as

$$\begin{aligned}\overline{C}_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\left(\frac{1}{\lambda_{RD}}\right)^k \Bigg[(\lambda_{SR})^{k+1} \\ \times \mathcal{G}_{N+1,1:1,1:1,N+1}^{1,N:1,1:N,1}\left(\underbrace{-1-k,\ldots,-1-k}_{N},-k \atop -1-k\middle|\begin{matrix}0\\0\end{matrix}\middle|_{\underbrace{1,\ldots,1}_{N},0}^{1}\middle|\lambda_{SR},1\right) + (\lambda_{RE})^{k+1}\end{aligned}$$

$$\times \mathcal{G}_{N+1,1:1,1:1,N+1}^{1,N:1,1:N,1}\left(\underbrace{-1-k,\ldots,-1-k}_{N},-k \atop -1-k \middle| \begin{matrix}0\\0\end{matrix}\middle| \underbrace{1,\ldots,1}_{N},1,0 \middle| \lambda_{RE},\frac{\lambda_{RE}}{\lambda_{SR}}\right)\Bigg] - \sum_{i=0}^{L} w_i$$

$$\times \frac{\mathrm{e}^{-(\frac{1}{\lambda_{RD}}-1)x_i}}{1-x_i}\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\lambda_{SR}}\middle| \underbrace{1,\ldots,1}_{N},1,0\right)\mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\lambda_{SR}}\middle| \underbrace{1,\ldots,1}_{N},1,0\right)$$

$$\times \mathcal{G}_{1,N+1}^{N,1}\left(\frac{x_i}{\lambda_{RE}}\middle| \underbrace{1,\ldots,1}_{N},1,0\right). \tag{17.53}$$

Invoking $\overline{C}_1$ from (17.48) and $\overline{C}_2$ from (17.53) into (17.43), we can easily obtain the expression for the ESC.

## 17.4 Numerical Results and Discussions

In this section, we present numerical results (via MATHEMATICA) and simulation results (on MATLAB with simulations averaged over $10^6$ independent trials) to corroborate our analytical findings. To perform the numerical investigations, we consider $\lambda_{SR} = \lambda_{RD} = \lambda$ and $L = 50$, otherwise unless specified.

Figure 17.2 illustrates the SOP performance for different values of cascading degree parameter $N$, when $\mathcal{R}_s = 0.5$ bps/Hz and $\lambda = \lambda_{RE}$. It can be concluded from the plots in Fig. 17.2 that the results obtained from the derived analytical expressions are in agreement with the simulation results. It can also be seen that the asymptotic results immaculate well with the exact results in medium-to-high SNR regime. Furthermore, the SOP plots exhibit a secrecy floor in the higher SNR regime irrespective of $N$, which validates the conclusion drawn in the Sect. 17.3.3 that the system secrecy diversity order reduces to zero. Moreover, it can also be observed that the system secrecy performance improves as $N$ increases.

Figure 17.3 illustrates the effect of $\lambda_{RE}$ and $\mathcal{R}_s$ on the SOP performance when $N = 3$. It can be concluded from the plots that the derived analytical results for the SOP immaculate well with the simulation results, thus confirming the accuracy of the deduced results. Also, from the plots it can be seen that the system SOP performance gets better with $\lambda$. This is due to the reason that the legitimated channel capacity improves as $\lambda$ increases. Moreover, we also observe that the SOP decreases as $\lambda_{RE}$ decreases which highlights the fact that as $\lambda_{RE}$ decreases, the eavesdropper channel condition deteriorates. Furthermore, the SOP performance improves as the secrecy rate threshold $\mathcal{R}_s$ decreases. This is because of the reason that a much higher power is needed to satisfy a higher $\mathcal{R}_s$.

Figure 17.4 depicts the probability of non-zero secrecy capacity performance versus $\lambda$ for various values of $\lambda_{RE}$ and the cascading parameter $N$. From the plots, it

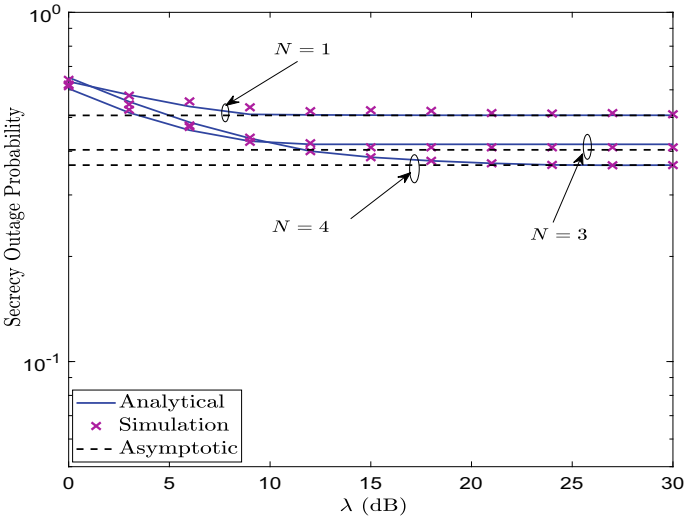



**Fig. 17.2** Secrecy outage probability for various values of $N$

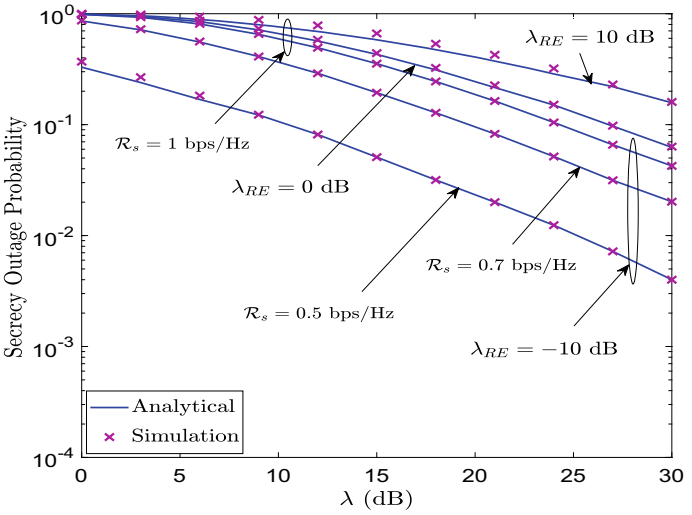



**Fig. 17.3** Secrecy outage probability versus $\lambda$ for different $\lambda_{RE}$ and $\mathcal{R}_s$

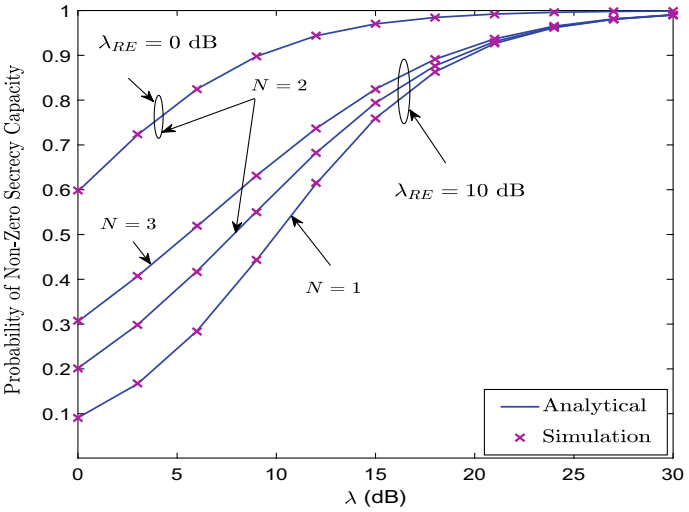



**Fig. 17.4** Probability of non-zero secrecy capacity versus λ

can be seen that the derived analytical results immaculate well with the simulation results. Moreover, it is also seen that the probability of non-zero secrecy capacity increases, as λ increases. This is due to the improvement in legitimated channel capacity with the increase in λ. Furthermore, we can observe that with the increase in the degree of cascading parameter $N$, the secrecy performance improves. Also, it can be seen that the probability of non-zero secrecy capacity is enhanced with the decrease in $\lambda_{RE}$. This is because a decrement in $\lambda_{RE}$ brings a degradation in the eavesdropper channel quality.

Figure 17.5 shows the impact of cascading parameter $N$ and $\lambda_{RE}$ on the intercept probability. From the plots, we can observe that the numerical results deduced from the derived analytical expression of the intercept probability are in an excellent match to the simulation results. This shows that our derived analytical expression is accurate. We can see that the intercept probability decreases with the increase in λ. Moreover, the intercept probability also decreases with the decrement in the value of $\lambda_{RE}$. This is because, the increase in λ and the decrease in $\lambda_{RE}$ leads to better-legitimated channel quality and a poor wiretap channel quality, respectively. Furthermore, the systems secrecy performance improves (intercept probability decreases) as $N$ increases.

Figure 17.6 shows the ESC behavior with respect to λ for different values of $\lambda_{RE}$, when $N = 3$. The ESC analytical curves are obtained by truncating the infinite series expansion over index $k$ to first few terms ($k = 9$) beyond which there is no change in the first five decimal places of the results. Hence, the derived analytical results of ESC in (17.43) immaculate well with the simulation results for the entire SNR regime which confirms the accuracy of the deduced ESC expression. Furthermore, it can also be seen that the ESC performance enhances with the increase in λ. This is is because a higher value of λ improves the legitimated channel quality. Moreover, it is

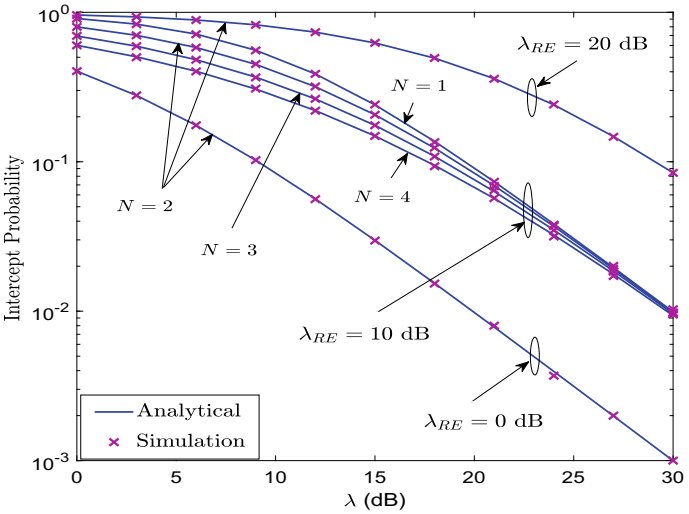



**Fig. 17.5** Intercept probability versus λ for various $\lambda_{RE}$ and $N$

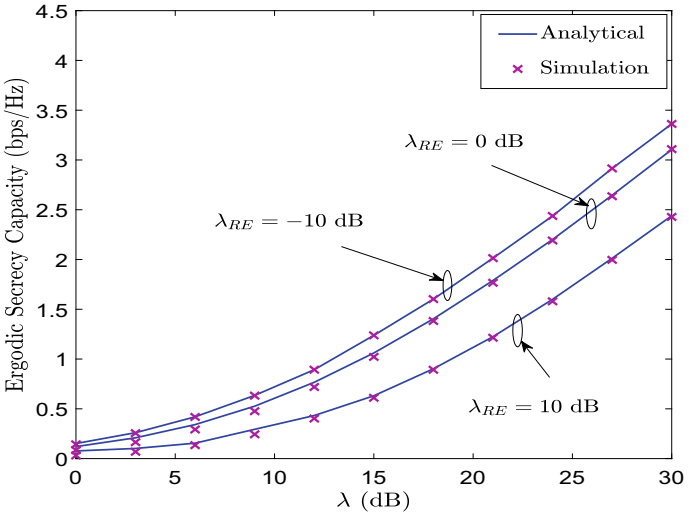



**Fig. 17.6** Ergodic secrecy capacity versus λ

also illustrated that as $\lambda_{RE}$ decreases, the ESC performance shows an improvement. This is due to the reason that a degradation in $\lambda_{RE}$ worsens the eavesdropper channel capacity, and thereby diminishing its ability to eavesdrop the secret transmissions.

## 17.5 Conclusions

In this chapter, we examined the secrecy performance of cooperative vehicular relaying networks, wherein a source vehicle wishes to send the confidential messages to a fixed destination with the aid of an amplify-and-forward relay vehicle in the presence of an eavesdropper vehicle. We model the V2V links by $N$-Rayleigh fading and the V2I links as Rayleigh fading. Under such a scenario, we evaluated the system secrecy performance in terms of key metrics, such as, SOP, intercept probability, probability of non-zero secrecy capacity, and ESC. We also analyzed the system secrecy diversity order by performing asymptotic SOP analysis in the high SNR regime. The asymptotic results revealed that the eavesdropper has negative impact on the SOP performance, and the system secrecy diversity order reduces to zero. Finally, we corroborated our analytical studies through numerical and simulation results.

**Acknowledgements** This research work done in this chapter was supported by the Science & Engineering Research Board (a statutory body of the DST, Govt. of India), under Project ECR/2017/000104.

## References

5gworldpro.com [Online] (2019) https://www.5gworldpro.com/5g-news/145-south-korea-created-an-entire-city-for-5gautonomous-vehicles.html

Abromowitz M, Stegun IA (1970) Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover, New York, NY, USA

Ai Y, Cheffena M, Mathur A, Lei H (2018) On physical layer security of double Rayleigh fading channels for vehicular communications. IEEE Wirel Commun Lett 7(6):1038–1041

Akhtar AM, Behnad A, Wang X (2014) On the secrecy rate achievability in dual-hop amplify-and-forward relay networks. IEEE Wirel Commun Lett 3(5):493–496

Akin AI, Ilhan H, Ozdemir O (2015) Relay selection for DF-based cooperative vehicular systems. EURASIP J Wirel Commun Netw 30:1–9

Alghorani Y, Kaddoum G, Muhaidat S, Pierre S, Al-Dhahir N (2016) On the performance of multihop-intervehicular communications systems over n* Rayleigh fading channels. IEEE Wirel Commun Lett 5(2):116–119

Andersen JB (2002) Statistical distributions in mobile communications using multiple scattering. In: Proceedings of 27th URSI general assembly, pp 0090–6778

Ansari IS, Al-Ahmadi S, Yilmaz F, Alouini MS, Yanikomeroglu H (2011) A new formula for the BER of binary modulations with dual-branch selection over generalized-K composite fading channels. IEEE Trans Commun 59(10):2654–2658

ElHalawany BM, El-Banna AAA, Wu K (2019) Physical-layer security and privacy for vehicle-to-everything. IEEE Communi Mag 57(10):84–90

Fan L, Lei X, Duong TQ, Elkashlan M, Karagiannidis GK (2014) Secure multiuser communications in multiple amplify-and-forward relay networks. IEEE Trans Commun 62(9):3299–3310
Fan L, Yang N, Duong TQ, Elkashlan M, Karagiannidis GK (2016) Exploiting direct links for physical layer security in multiuser multirelay networks. IEEE Trans Wirel Commun 15(6):3856–3867
Gradshteyn IS, Ryzhik IM (2000) Tables of integrals, series, and products. Academic Press, New York
Hartenstein H, Laberteaux KP (2008) A tutorial survey on vehicular ad hoc networks. IEEE Commun Mag 46(6):164–171
He R, Molisch AF, Tufvesson F, Wang R, Zhang T, Li Z, Ai B (2016) Measurement-based analysis of relaying performance for vehicle-to-vehicle communications with large vehicle obstructions. In: 2016 IEEE 84th vehicular technology conference (VTC-Fall), pp 1–6
Hoang TM, Duong TQ, Suraweera HA, Tellambura C, Poor HV (2015) Cooperative beamforming and user selection for improving the security of relay-aided systems. IEEE Trans Commun 63(12):5039–5051
Ilhan H, Uysal M, Altunbas I (2009) Cooperative diversity for intervehicular communication: performance analysis and optimization. IEEE Trans Veh Technol 58(7):3301–3310
Ilhan H, Akin AI (2012) Performance analysis of AF relaying cooperative systems with relay selection over double Rayleigh fading channels. In: 2012 6th International conference on signal processing and communication systems (ICSPCS). IEEE, pp. 1–6
Kuhestani A, Mohammadi A, Mohammadi M (2018) Joint relay selection and power allocation in large-scale MIMO systems with untrusted relays and passive eavesdroppers. IEEE Trans Inf Forens Secur 13(2):341–355
Liu C, Yang N, Yuan J, Malaney R (2015) Location-based secure transmission for wiretap channels. IEEE J Sel Areas Commun 33(7):1458–1470
Lv L, Chen J, Yang L, Kuo Y (2017) Improving physical layer security in untrusted relay networks: cooperative jamming and power allocation. IET Commun 11(3):393–399
Martinez F, Toh C-K, Cano J-C, Calafate C, Manzoni P (2010) Emergency services in future intelligent transportation systems based on vehicular communication networks. IEEE Intell Transp Syst Mag 2(2):6–20
Nguyen BV, Kim K (2015) Secrecy outage probability of optimal relay selection for secure AnF cooperative networks. IEEE Commun Lett 19(12):2086–2089
Nguyen SQ, Kong HY (2016) Outage probability analysis in dual-hop vehicular networks with the assistance of multiple access points and vehicle nodes. Wireless Pers Commun 87(4):1175–1190
Niu H, Zhu N, Sun L, Vasilakos AV, Sezaki K (2016) Security-embedded opportunistic user cooperation with full diversity. Wirel Netw 22(5):1513–1522
Pandey A, Yadav S (2018) Physical layer security in cooperative AF relaying networks with direct links over mixed Rayleigh and double-Rayleigh fading channels. IEEE Trans Veh Technol 67(11):10615–10630
Pandey A, Yadav S (2018) Performance evaluation of amplify-and-forward relaying cooperative vehicular networks under physical layer security. Trans Emerging Telecommun Technol 29(11):e3534
Pandey A, Yadav S (2020) Physical layer security in cooperative AF relay networks over mixed Nakagami-*m* and double Nakagami-*m* fading channels: performance evaluation and optimization. IET Commun 14(1): 95–104. https://doi.org/10.1049/iet-com.2019.0584, www.ietdl.org
Parkinson S, Ward P, Wilson K, Miller J (2017) Cyber threats facing autonomous and connected vehicles: future challenges. IEEE Trans Intell Transp Syst 18(11):2898–2915
Prudnikov AP, Brychkov YA, Marichev OI (1990) Integrals and series vol 3: more special functions. Gordon and Breach Science Publishers
Sakiz F, Sen S (2017) A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV. Ad Hoc Netw 61:33–50
Salo J, El-Sallabi HM, Vainikainen P (2006) Statistical analysis of the multiple scattering radio channel. IEEE Trans Antennas Propag 54(11):3114–3124

Seyfi M, Muhaidat S, Liang J, Uysal M (2011) Relay selection in dual-hop vehicular networks. IEEE Sig Process Lett 18(2):134–137
Shah SAA, Ahmed E, Imran M, Zeadally S (2018) 5G for vehicular communications. IEEE Commun Mag 56(1):11–117
Sun L, Ren P, Du Q, Wang Y, Gao Z (2015) Security-aware relaying scheme for cooperative networks with untrusted relay nodes. IEEE Commun Lett 19(3):463–466
The Wolfram Functions Site. http://functions.wolfram.com/
Wang W, Teh KC, Li KH (2016) Relay selection for secure successive AF relaying networks with untrusted nodes. IEEE Trans Inf Forens Secur 11(11):2466–2476
Yan G, Rawat DB (2017) Vehicle-to-vehicle connectivity analysis for vehicular ad-hoc networks. Ad Hoc Netw 58:25–35
Yang N, Wang L, Geraci G, Elkashlan M, Yuan J, Renzo MD (2015) Safeguarding 5G wireless communication networks using physical layer security. IEEE Commun Mag 53(4):20–27
Yang L, Chen J, Jiang H, Vorobyov SA, Zhang H (2017) Optimal relay selection for secure cooperative communications with an adaptive eavesdropper. IEEE Trans Wirel Commun 16(1):26–42
Yener A, Ulukus S (2015) Wireless physical-layer security: lessons learned from information theory. Proc IEEE 103(10):1814–1825
Zhang J, Pan G (2017) Secrecy outage analysis with Kth best relay selection in dual-hop inter-vehicle communication systems. AEU Int J Electron Commun 71:139–144
Zhao R, Lin H, He YC, Chen DH, Huang Y, Yang Y (2018) Secrecy performance of transmit antenna selection for MIMO relay systems with outdated CSI. IEEE Trans Commun 66(2):546–559
Zou Y, Wang X, Shen W (2013) Optimal relay selection for physical-layer security in cooperative wireless networks. IEEE J Sel Areas Commun 31(10):2099–2111
Zou Y, Zhu J, Wang X, Hanzo L (2016) A survey on wireless security: technical challenges, recent advances, and future trends. Proc IEEE 104(9):1727–1765

# Chapter 18
# Machine Learning in 5G Wireless Networks


**Abhishek and Shekhar Verma**

**Abstract** The surge in demand for faster, reliable, and inter-operable networks leads to the development of 5G networks. The resultant increase in the numbers of services, heterogeneous devices with quality of service assurance requires the infrastructure to be further optimized. A sub-optimal infrastructure can be tuned through manual or rule-based optimization, but to further realize the full potential of the 5G network, the change should be based on the current scenario and traffic pattern in the network. The recent developments in machine learning algorithms where the models are capable of learning from the data itself have proved their potential by minimizing the manual intervention for optimal performance. In this chapter, authors identify the areas where the problems faced in 5G infrastructure can be modeled as machine learning problems, which is followed by the introduction of well-explored machine learning models for solving these problems.

## 18.1 Introduction

The rapid increase in the number of mobile devices, multimedia applications, and wireless data service demands require quick and robust adaptability at network infrastructure end. The next-generation wireless network 5G promises to provide such an adaptive infrastructure that can accommodate 100 times more devices with high-throughput, ultra-low latency, high QoS, low power consumption to have fast and real-time communication services (Agiwal et al. 2016). Additionally, 5G also ensures interoperability between heterogeneous devices or networks by including the Internet of things (IoT), vehicular networks, and device-to-device communication. The major challenges in achieving an ideal 5G network infrastructure are:

Abhishek (✉) · S. Verma
Indian Institute of Information Technology, Allahabad, India
e-mail: rsi2016006@iiita.ac.in

S. Verma
e-mail: sverma@iiita.ac.in



M. Mandloi et al. (eds.), *5G and Beyond Wireless Systems*,
Springer Series in Wireless Technology,
https://doi.org/10.1007/978-981-15-6390-4_18

1. Interference
2. User localization
3. Traffic management
4. Network control
5. Channel optimization
6. Power consumption
7. Security

In an ideal scenario, all the above challenges have been successfully conquered, but the real-time network exhibits completely different properties than that of the ideal. Thus, the standard solutions fail to give satisfactory results and lead to sub-optimal utilization of the infrastructure. Just like the 5G network, the solutions should be adaptive in nature, i.e., the parameters should be tuned according to the current scenario which are difficult to be assured manually. The adaptive solution objective aligns with the aim of all machine learning (ML) models that can exploit the patterns hidden in data by learning from the data itself. In today's scenario, the users generate and consume humongous mobile data regularly which include data generated from device-mounted sensors, consumption of online content, digital media, usage patterns, etc. This data can coarsely tell about the current surrounding of the user and its usage pattern. The network dynamics data can be further used to optimize the resource allocation (Cheng et al. 2017). The availability of big data, high computing, cheap storage, and offline training (De Coninck et al. 2015) has enabled ML solutions to be used widely. In general, ML algorithms extract the patterns hidden in data giving insights of the data. The same task of feature engineering and pattern extraction by domain expert manually has been proved to be very complex, costly, and prone to error (Ahmed et al. 2018; Paolini 2017).

While the simple and effective models such as k-nearest neighbors (kNN) (Bahl and Padmanabhan 2000), support vector machine (SVM) (Wu et al. 2004), linear regression, ridge regression, and logistic regression have already been applied in solving the problems in wireless networks, they become inefficient and prone to error as the volume of data increases. Also, these models require large amount of labeled samples which ultimately increase manual efforts and cost. The recent advancements in deep neural networks (DNN) (Schmidhuber 2015) efficiently exploit the nonlinear relationship present in data by processing it from low level to high level producing optimal semantic information. DNN has already been proved as useful and accurate technique in application areas such as vision, audio, and robotics (Ordóñez and Roggen 2016; Zhang et al. 2016). In this chapter, we explore the efficient and effective deep learning solutions specifically designed to solve wireless network problems. The chapter is further divided into several sections where Sect. 18.2 briefly describes the basics of ML methods followed by their application in 5G wireless networks to solve various problems in Sect. 18.3. The findings and conclusion of this chapter have been illustrated in Sect. 18.4.

## 18.2 Machine Learning Basics

The machine learning algorithms can be broadly classified under three groups: supervised, unsupervised, and semi-supervised. A model built on the basis of labeled data only comes under supervised learning. In unsupervised learning, the aim changes from finding the correct label of the unlabeled data to group data with similar characteristics. Semi-supervised methods include best of both supervised and unsupervised learning models. It enhances the model's accuracy where supervised model fails to perform due to labeled data scarcity. The semi-supervised learning exploits the marginal density information present in abundant unlabeled data to learn an accurate and generic function (Fig. 18.1).

Deep learning is a subset of machine learning models, while most of the machine learning algorithms rely on domain experts to identify the important features present in data, deep learning models do it automatically. It extracts the information hierarchically by passing the raw data through multiple layers of nonlinear functions. The resultant domain-specific semantic information helps in making accurate predictions. Neural networks are the simplest model of deep learning architecture. A neural network essentially consists of an input layer and an output layer with multiple hidden layers sandwiched in between them. The size of input layer depends on the number of features present in the data, and size of the output layer depends on the desired output. Each layer in a neural network is controlled by individual neurons. The neuron takes in data from previous layer's output and gives out a value if the input fulfills the neuron activation criteria. Neurons in one layer are connected to a few or all neurons in the next layer decided by dropouts or fully connected criteria, respectively. The learning of the neural network involves optimizing the edge weight which connects neurons to two adjacent layers. Though any optimization technique can be used, back-propagation is most widely used to learn optimal weights due to its simplicity and efficiency.

The deep learning architecture creates a model through forward and backward propagations as shown in Fig. 18.2. The input layer has been shown with blue color, the three hidden layers with green color and output layer with yellow color. The model consists of a loss function $L(w)$ with learning rate $\lambda$, nonlinear activation function $\sigma$ (e.g., sigmoid, tanh, ReLu), and $W$ is the weight matrix between current and next layer. In mathematical terms, the aim is to minimize the loss function using some optimization (e.g., gradient descent) based on the labeled samples. In the forward

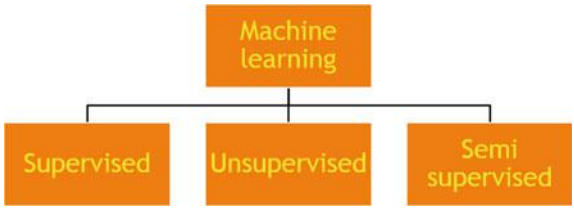



**Fig. 18.1** Machine learning taxonomy

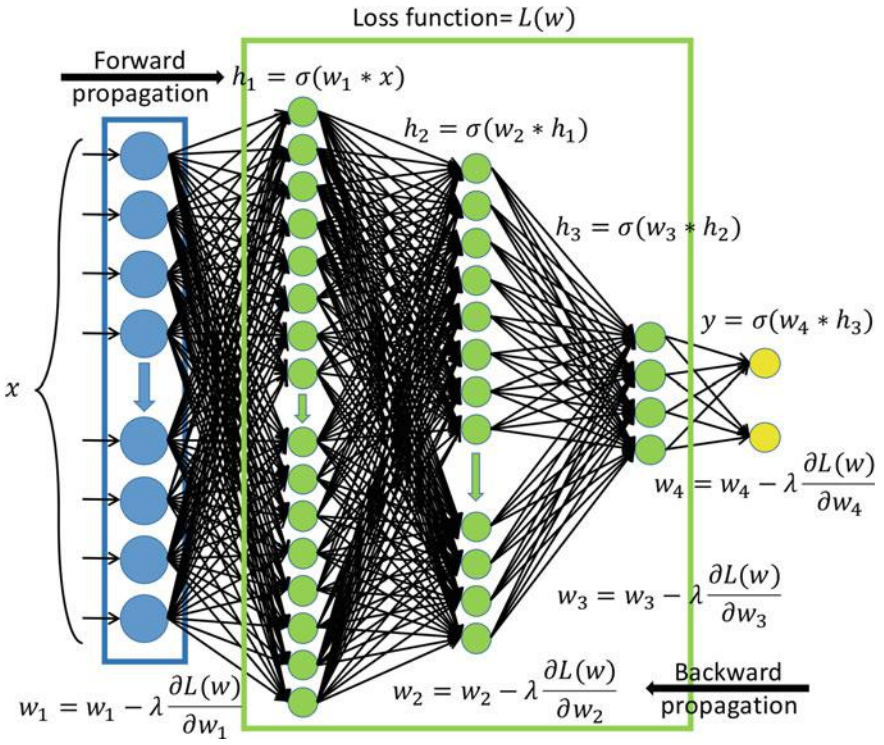



**Fig. 18.2** Neural network

pass, the weights are trained based on the activation of respective neurons, while in back-propagation, the final resultant error is checked for change with respect to each layer's weight using chain rule, and the same value is adjusted to learn appropriate weights. Due to the chain rule, an incompatible activation function can lead to vanishing gradient problem which stops the model from further learning. A machine learning model helps in feature extraction from a high-dimensional data without external input from experts. It is also helpful in handling and exploiting big data.

The artificial neural network is further enhanced by performing convolution operation on input space with different filters or kernels corresponding to different feature extraction. Figure 18.3 shows a $3 \times 3$ kernel represented by green outline on a $6 \times 6$ input instance to give one output shown with dotted lines. The whole input is covered by shifting the kernel in different segments driven by strides (number of intermediate features to be skipped). At first, convolutional neural networks (CNN) were introduced for image and video data, but they have been further expanded to be utilized in different data to exploit the variety of information that can be extracted using different kernels.

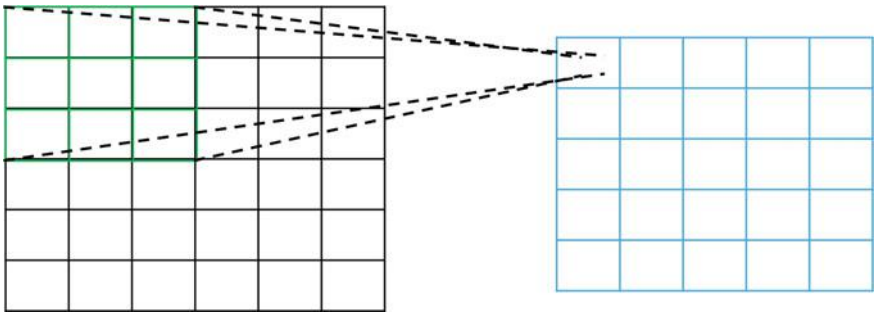


**Fig. 18.3** Convolution operation

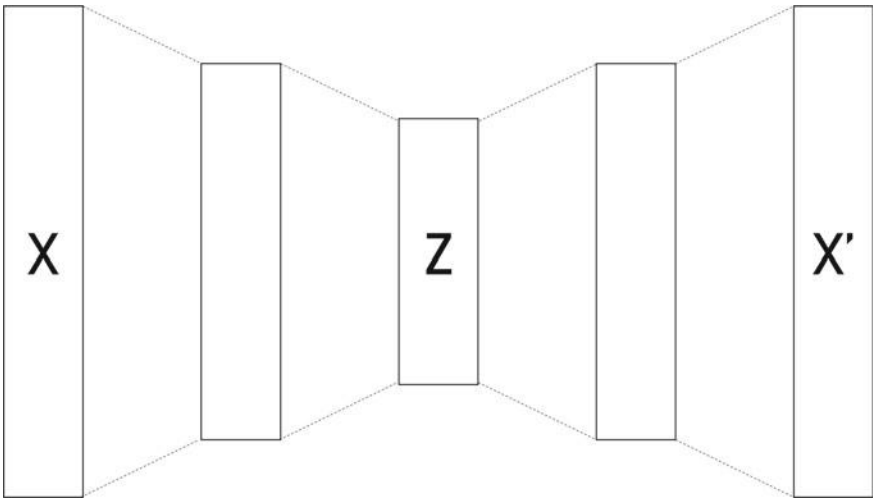



**Fig. 18.4** Auto-encoders

Data is the new fuel. The ML model's accuracy relies on the representation of the given data; better the representation, higher will be the accuracy. As such, most of the data are affected by noise and unknown transformations during different phases of data collection and processing stages. Thus, a better representation of the data is required which can be achieved using auto-encoders. An auto-encoder takes in the input data and searches for optimal intermediate representation by minimizing the reconstruction loss at the other end. As shown in Fig. 18.4, $X$ is the input which is transformed in the optimal intermediate representation $Z$ by minimizing the loss between $X$ and $X'$.

Recurrent neural network (RNN) is the neural network specifically designed for modeling sequential data (time series) (Heck and Salem 2017; Hochreiter and Schmidhuber 1997; Zhou et al. 2016). The 5G network data intrinsically contains sequential relationship which needs to be extracted and exploited to make better

use of underlying infrastructure. In each time step, the RNN utilizes the previous state of the model to predict outcome at current stage. However, the traditional RNN remained affected from vanishing and exploding gradient problem which is solved by introducing input, output, and forget gate in long short-term memory (LSTM)-based RNN. The next section contains a brief description about these machine learning models help in solving various problems in 5G networks.

## 18.3 Machine Learning in 5G

The data in a 5G network can be broadly classified into two categories: network-level data and app-level data. The network-level data contains information about infrastructure and its performance like throughput, end-to-end delay, jitter, etc. The user-level data contains information generated from handheld devices through the sensors and mobile applications. The network-level data also logs the sender–receiver information, communication type, and session time which remain helpful in network management, resource allocation, diagnosis, etc. It exhibits the network dynamics with users' mobility pattern as a spatio-temporal relationship data.

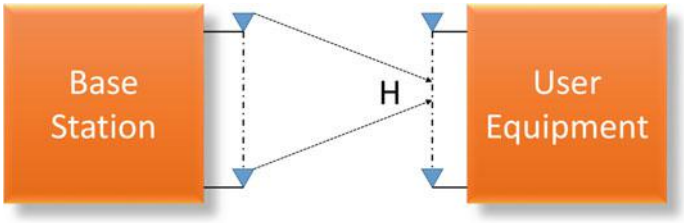
The heterogeneous nature of data in the 5G networks due to different source of collection and varying formats contains complex correlation among them. The aim is to extract this hidden information and utilize them to optimize the underlying network infrastructure and devices. In theory, more data give better insight and analysis of the underlying system, but as the data increases, it becomes impossible for domain experts to extract information manually. This leads to the requirement of some automated system that can analyze and extract the useful patterns from the data. While the existing wireless networks already use traditional machine learning algorithms to solve the problems, they become inefficient as the size and dimension of data increases. In contrast, deep learning methods enable exhaustive analysis on big data efficiently. The hierarchical abstraction of layers to extract the nonlinear correlation gives detailed insight of data through effective semantic representation.

In the core of 5G networks, single-user MIMO system is employed to boost spectral efficiency and reliability in comparison with single antenna systems. Similarly, multiple antennas at the base station efficiently utilize same frequency with spatial multiplexing to serve different users. Fundamentally, MIMO system maximizes the 5G efficiency by increasing the diversity gain, array gain, and spatial multiplexing gain. So, if a symbol is transmitted by multiple antennas ($N_t$) and received using multiple antennas ($N_r$) at receiver end, then due to $N_t \times N_r$ diversity with multiple replica of same symbol, the probability of all the replicas fading at same time drastically reduces. Through focused radio beam and combination of all symbols, the SNR at targeted user can be increased.

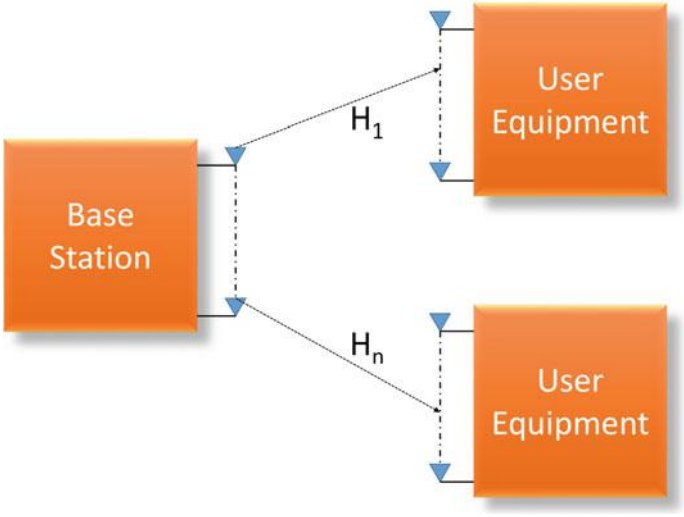
### 18.3.1 Network-Level Data

The four broad areas aimed to be solved using machine learning on network-level data are network state prediction, network traffic classification, call detail records, and channel analysis. As known, the mobile network is not always fully occupied and utilized. It includes peak hours when the network is utilized to its maximum sustainable limit and off-hours when the infrastructure remains idle. A network state prediction refers to estimate the network traffic and its performance based on the historical data available about the mobile infrastructure.

Figure 18.5a, b show single-user and multi-user MIMO systems. While the former involves only one BS and one UE, the later includes single BS with multiple UE using channel $H$. Practically, the MIMO system can achieve its maximum network throughput when there is only one transmitter. Among multiple receivers, a codeword is constructed for each UE such that the intended receiver does not see interference




(a) Single user MIMO.




(b) Multi user MIMO.

**Fig. 18.5** SU-MIMO and MU-MIMO

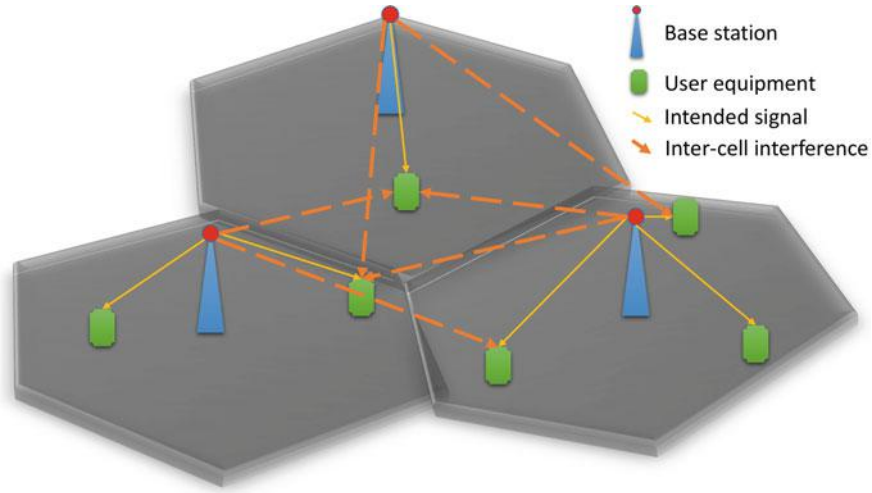



**Fig. 18.6** Inter-cell interference

from signals intended for other receivers. However, knowing the true characteristics of wireless channel and perform successive encoding is not feasible hence, in MIMO systems, the receiver only decodes the desired signal considering others as additive white Gaussian noise. The exponential growth of mobile devices further pushes this limit by enabling multiple BS transmitting at same frequency and time slots. This leads to inter-cell interference between a BS and UE of different BS as shown in Fig. 18.6.

Currently, the problem of inter-cell interference is solved using interference awareness or co-ordination or joint signal processing. In interference aware network, based on feedback from its cell's UE for other BS interference, it controls its transmission power to maximize intra-cell performance. In interference co-ordination, the data exchange between adjacent BS is minimized and tunes antenna gain to form directed beam toward intended UE. Joint processing exploits the spatial information of UE and shares the intended data among BS to simultaneously provide the stream to the same UE. While interference co-ordination is widely used due to its efficiency and simplicity, restricting the data transfer among adjacent BS adversely affects the performance of the whole infrastructure. Along with interference, the 5G network is also affected by sparse connection and shadow regions.

The infrastructure and user-level data available in huge amount can be used to extract the hidden information in them about the quality of experience (QoE) of a user. The same information can also be used to achieve high-throughput network with minimum or no interference with existing infrastructure. The problem of unsatisfactory QoE can also be solved by identifying such regions and altering the infrastructure or services. A multilayer perceptron (MLP) (Pierucci and Micheli 2016) examined the data (average user throughput, number of active users in a cell, average data vol-

ume per user, and channel quality indicators) from UE to identify low QoE regions. Each neuron in MLP used sigmoidal transfer activation function to collectively form one input, one output, and two hidden layers. The sigmoidal transfer function can be expressed as

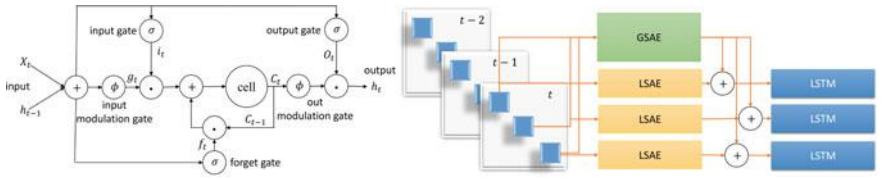
$$S(x) = \frac{e^x}{e^x + 1}$$

where $x$ represents input data instance. The neuron activates when its input value goes above the threshold. Region classification using MLP achieved 97.5% of accuracy as compared to baseline radial basis function network. A wireless network generally remains affected from traffic congestion and bottlenecks which drastically decrease the network performance. Authors in (Gwon and Kung 2014) examined the received packets to predict the network flow which can be further utilized in traffic monitoring, network node scheduling, resource provisioning, and queue management. The model performed in two stages where it extracted the essential features in first stage and then employed a classifier on them. Feature extraction utilized sparse coding through K-SVD and orthogonal matching pursuit then through max pooling operation, and the dimension of the data is reduced to preserve significant features. Further classification task was handled by SVM which reduced the burst size and gap length estimation error to 12.2% and 6.8%, respectively.

Wireless mesh networks to deliver last mile connectivity in 5G to end users. To ensure high QoE, predictive network planning is required. Authors proposed to use a deep belief network (DBN) with Gaussian model which is used to estimate the traffic demand patterns in wireless mesh networks (Nie et al. 2018). The model computed two different wavelet coefficients for both high-pass and low-pass components of the network traffic and employed a DBN on low-pass component to predict the upcoming demand. In terms of traffic prediction accuracy and bias, DBN with logistic regression outperformed the existing state-of-the-art network prediction methods: principal component axis (PCA), tomogravity, and single-range matching filtering (SRMF).

**Algorithm 1** Traffic demand prediction

**Input:** known network traffic $(x^1_{p,q}(t), x^2_{p,q}(t) \ldots x^K_{p,q}(t)) \in X$
**Output:** next traffic prediction $X(T+1)$
**for** $i = 1$ **to** $K$ **do**
  $c^i_{p,q} \leftarrow DWT(x^i_{p,q}(t))$
**end for**
Train DBN $(c^i_{p,q})^K_{i=1}$
$[C, D] \leftarrow DWT(X)$
$\hat{C} \leftarrow DBN(C)$
$\hat{D} \leftarrow SRMF(D)$
$X(T+1) \leftarrow IDWT(\hat{C}, \hat{D})$
**return** $X(T+1)$




(a) LSTM unit. (b) Spatio temporal model.

**Fig. 18.7** Network utilization prediction model

As described in Algorithm 1, it uses the previous traffic information to predict the upcoming requirements. It first finds the discrete wavelet transform (DWT) for both low- and high-pass components and train the DBN on them. Similarly, in prediction phase also, it computes DWT of the data followed by inverse DWT on combined prediction from DBN and SRMF. While the local network prediction analyzes and optimizes the performance locally, the same prediction on a wider scale can benefit all the connected users by optimal resource allocation and mutual co-ordination. Authors in (Wang et al. 2017a) proposed to use auto-encoders with LSTM to exploit the spatio-temporal relationship in the data. A single LSTM unit has been depicted in Fig. 18.7a. It solves the exploding and vanishing gradient problem of recurrent neural network by controlling the amount of information to be passed in successive state through different gates. The model is fed with historical data of average traffic load on each BS for training purpose, and then the same is utilized to perform prediction for the upcoming instances. The model consisted of three components: local-stacked auto-encoders (LSAE), global-stacked auto-encoders (GSAE), and LSTM. It begins with GSAE and LSAE encoding the spatial data of the target cell and its adjacent cells as global and local representation, and then, the concatenated local and global representations are fed to LSTMs for training and future predictions.

The hybrid model as shown in Fig. 18.7b outperformed support vector regression and the auto-regressive integrated moving average model by harnessing the collective processing capability of stacked auto-encoders and LSTM to exploit the spatial and temporal relationship in the network data. The solution still remains short of extracting the complex patterns hidden in the data especially in case of dense network connection resulting in big data. CNNs have the excellent capability to uncover these patterns using different kernels which makes ConvLSTM (Alawe et al. 2018; Chen et al. 2018; Feng et al. 2018; Huang et al. 2017; Zhang and Patras 2018) a better model than stacked AE with LSTM. A spatio-temporal neural network for traffic forecasting includes ConvLSTMs along with 3D convolution layer to encode the network data optimally which is further decoded using fully connected layers to make future traffic predictions (Table 18.1).

The access and mobility management function in 5G generally handles the requests from UEs. On serving large number of demands, it faces a problem due to congestion. A deep neural network (Alawe et al. 2018) trained on historic service requests can predict the upcoming request volume to allocate resources properly. The

**Table 18.1** Normalized RMSE of different methods in network traffic prediction (Zhang and Patras 2018)

| Method | 1-step | 10-step | 20-step | 30-step |
|---|---|---|---|---|
| STN | $0.19 \pm 0.02$ | $0.29 \pm 0.05$ | $0.51 \pm 0.12$ | $0.83 \pm 0.14$ |
| D-STN | $0.19 \pm 0.02$ | $0.28 \pm 0.04$ | $0.48 \pm 0.09$ | $0.71 \pm 0.18$ |
| HW-ExpS | $0.33 \pm 0.03$ | $0.51 \pm 0.08$ | $0.96 \pm 0.01$ | $1.79 \pm 0.45$ |
| ARIMA | $0.20 \pm 0.04$ | $0.39 \pm 0.15$ | $0.77 \pm 0.31$ | $1.00 \pm 0.27$ |
| MLP | $0.23 \pm 0.02$ | $0.38 \pm 0.03$ | $0.67 \pm 0.13$ | $0.96 \pm 0.22$ |
| ConvLSTM | $0.23 \pm 0.02$ | $0.39 \pm 0.05$ | $0.95 \pm 0.24$ | $1.49 \pm 0.243$ |
| D-ConvNet | $0.20 \pm 0.02$ | $0.37 \pm 0.09$ | $0.95 \pm 0.30$ | $1.64 \pm 0.34$ |
| SVM | $0.39 \pm 0.16$ | $0.46 \pm 0.11$ | $0.62 \pm 0.14$ | $0.95 \pm 0.19$ |
| AE+LSTM | $0.24 \pm 0.05$ | – | – | – |

network trained on 60% of total data was able to predict rest 40% service request with 80% accuracy. The deep model can be further refined to increase the prediction accuracy by training it with more data or by adding more hidden layers, but that would restrict the model scalability. While on the same scenario, LSTM was able to outperform deep neural network by increasing the prediction accuracy up to 90%. The upcoming request can be easily translated into preventive measures by implementing load balancing or adding more request handlers.

Infrastructure planning and deployment remain key steps in providing high QoS to the users. Bad services do not always require to install new devices or add resources to the existing infrastructure, but they need to be optimized to increase the throughput. In different regions such as business area and residential area, load on the network does not overlap, i.e., the peak hours are different in both areas. Hence, having dedicated resources for data processing increases the operational overhead and does not fully utilize its potential. By clustering the base stations in different regions with a common data processing unit can further optimize the network usage. A multivariate LSTM (Mu-LSTM) is proposed in (Chen et al. 2018) to predict the upcoming traffic pattern by tapping on the temporal relationship hidden in the network-level data available from multiple BS. In the next step, the complimentary traffic pattern base stations, i.e., the aggregated data usage should not go beyond their capacity, are clustered together using $\epsilon$-NN algorithm on the weighted graph. The Mu-LSTM with $\epsilon$-NN clustering was able to achieve $\approx$93% accuracy in upcoming traffic prediction which increased the capacity utilization of the network to $\approx$83.4%. Table 18.2 shows the comparison between different methods employed to predict upcoming network traffic and their respective utilization ratio.

Deep diving into network-level data also gives insight about the different traffic protocols being utilized. The identification of delay-tolerant and intolerant protocols allow prioritizing them for better QoS. A deep auto-encoder (AE) has been employed for protocol classification using transmission control protocol (TCP) flow dataset in (Wang 2015). The unsupervised AE learning enables the model to extract the most

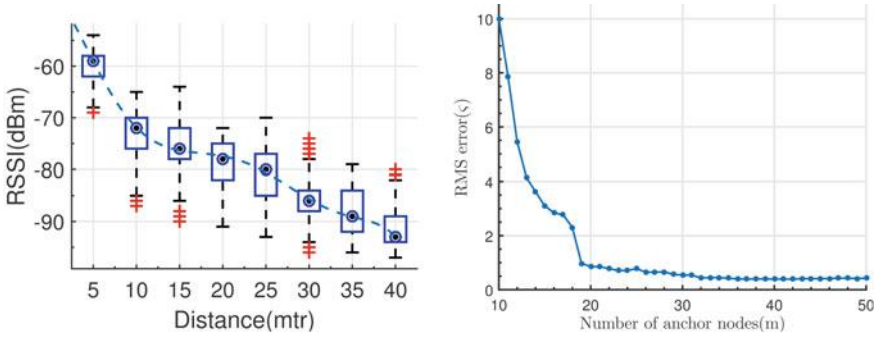
**Table 18.2** Upcoming traffic prediction and BS clustering (Chen et al. 2018)

| Methods | Upcoming traffic pattern error | | Average utility | | Deployment cost | |
|---|---|---|---|---|---|---|
| | Milan (%) | Trentino (%) | Milan (%) | Trentino (%) | Milan (%) | Trentino (%) |
| Traditional | – | – | 38.8 | 29.4 | 182 | 522 |
| ARIMA-DCCA | 20.2 | 23.7 | 65.3 | 45.2 | 112 | 160 |
| WANN-DCCA | 17.5 | 19.8 | 73.4 | 58.8 | 96 | 120 |
| MuLSTM-DC | 7.4 | 8.3 | 58.7 | 39.2 | 120 | 180 |
| MuLSTM-DCCA | 7.4 | 8.3 | 83.4 | 76.7 | 88 | 270 |

relevant features from the data and encode them in its optimal form for classification. The experimental results show that AE-based model was able to classify the known protocols and identify anomalous protocols accurately. Though training a model is easy on plain text dataset, the task becomes difficult when the traffic is encrypted. One solution would be to decrypt the traffic and perform analysis but that would defy the whole purpose of encryption and security, the other is to identify patterns from the encrypted traffic using CNN. 1D CNN (Wang et al. 2017b), deep packet (Lotfollahi et al. 2017), and convLSTM (Aceto et al. 2019) have been proposed which transforms the network traffic into an image which is further used to extract features through CNN. The CNN reduces the amount of hand-crafted features which ultimately reduce the overall models' complexity and give real-time inference for traffic classification. Advantage of utilizing deep learning models helps identifying unusual patterns in the traffic wherein the models can be used to identify and segregate malicious packets from the regular traffic.

### *18.3.2 User-Level Data*

On the other hand, the app-level data contains demographic information and usage pattern of respective users. User localization is an important feature in wireless networks to serve the users better. The current location of all users collaboratively can help identify the overcrowded cell or underutilized cell, the possible transit of mass, predicting next location of users, etc. A smart hand-off mechanism in a 5G network can ensure high QoS by minimizing the call drops and avoiding weak connectivity zones. However, an accurate user localization would be required to ensure such a mechanism. Though most of the devices are nowadays contain GPS (Davidson and Piché 2016; Xia et al. 2017; Xiao et al. 2016) onboard for localization, it remains absent on low-powered devices such as sensor networks which makes it difficult to identify the region of event. Authors in (Singh and Verma 2017) propose iL$^2$PA to use semi-supervised iterative graph Laplacian manifold learning for accurate localization in a situation where only majority of nodes rely on few localized nodes for localization using received signal strength indicator (RSSI). It creates a weighted graph based on the nodes' co-ordinates and performs regression in an iterative fashion to discard




(a) Distance vs RSSI (b) Anchors vs localization error

**Fig. 18.8** User localization using RSSI (Singh and Verma 2017)

the adverse effects of wireless channels. In the final stage, the transformed locations of the anchor nodes are mapped to original space using Procrustes analysis followed by the same operation in each anchor's local region.

Fig. 18.8a shows the relationship between distance and RSSI, and Fig. 18.8b shows the result trend between number of anchor nodes and the overall localization accuracy. While GPS and such devices solve the outdoor localization problem, they fail to deliver the same accuracy when devices are situated indoor. DeepFi (Wang et al. 2015) using channel state information (CSI) solves the problem and increases the indoor localization accuracy by using an auto-encoder. The weights of a trained network for each indoor location are stored as a fingerprint which is compared with the reconstructed CSI weights during testing phase. Experimental results show that DeepFi decreases the localization error by 50% as compared to existing state-of-the-art machine learning methods (FIFS, Horus, etc.) (Brunato and Battiti 2005; Xiao et al. 2012; Youssef and Agrawala 2005).

Mercury (Liang et al. 2016) trains a recurrent neural network (RNN) which is optimal for sequential data on historical call detail records (CDR) containing trajectory of users. Based on the user device connectivity to a series of BSs, the RRN model is trained in combination with current sequence of input from the CDR. The spatio-temporal map thus created ensures the accurate localization of the user or group of users. The usage pattern generated on UE can also be used to approximate the age and gender of the user using CNN (Felbo et al. 2015) which could prove helpful in resource allocation and load balancing. Table 18.3 compares the result of various state-of-the-art methods in user classification.

A device-free localization and activity recognition method (Wang et al. 2016) utilizes sparse AE to extract the important distinguishing features. It is followed by a softmax regression model to predict the user location and other essential parameters. Experimental results proved that automatic feature extraction gave better features than given by domain experts. CNN fed with received signal strength pattern capable enough to discard the multipath propagation effect proved to be 100% accurate

**Table 18.3** User classification accuracy

| Methods | Age (%) | Gender (%) |
|---|---|---|
| SVM | 38.7 | 58.5 |
| CNN | 61.3 | 80.4 |
| CNN-SVM | 65.1 | 84.3 |

in indoor localization (Ibrahim et al. 2018). The input to the model is created by constructing a matrix of $N \times T$, where $N$ is number of nodes in the network and $T$ is the time series containing the RSSI values as shown below:

$$\begin{bmatrix} RSSI_{1,1} & RSSI_{1,2} & \dots & RSSI_{1,T} \\ RSSI_{2,1} & RSSI_{2,2} & \dots & RSSI_{2,T} \\ & & \ddots & \\ RSSI_{N,1} & RSSI_{N,2} & \dots & RSSI_{N,T} \end{bmatrix}$$

Apart from the majorly defined supervised deep learning models, the devices generally provide the position of the unknown locations which can be treated as well-defined semi-supervised model. By leveraging the information hidden in the reported unknown location using deep reinforcement learning (DRL) and variational auto-encoders (VAE) (Mohammadi et al. 2017), the localization accuracy is further increased even with sparse labeled data. The agent here gets labeled target information along with unlabeled target distance calculated using VAE. The agent gets rewarded for correct movement prediction and after training; this helps in accurate user localization using both labeled and unlabeled data. Thus, through user localization, 5G infrastructure can be further optimized on dynamic requirements such as smart hand-off, network load balancing, and dynamic resource allocation.

### *18.3.3 Network Control*

5G network infrastructure runs on packet-based network. Hence, routing also poses an important challenge in network optimization. A quick and smart routing algorithm can reduce the overhead, thereby reducing the turnaround time and also avoiding bottlenecks and congestion by optimally load balancing the network. The network infrastructure can be depicted as a large graph where edges signify the available routes in the network. By exploiting the basic graph properties using a dedicated graph-query neural network, it learns various network topologies already present in the network (Geyer and Carle 2018). In the training phase, few edges of the adjacency matrix are randomly disabled based on Bernoulli distribution; this forces the model to learn alternate optimal path for routing. This significantly optimizes the distributed routing to optimally utilize the network. The deep reinforcement learning

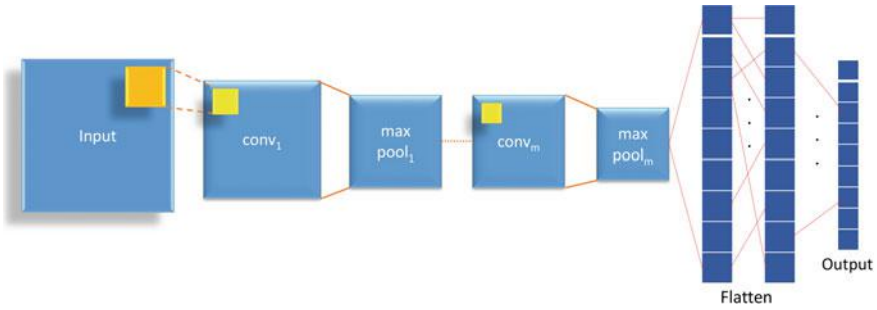



**Fig. 18.9** Convolution neural network

employed to identify the time varying traffic conditions and correlation between traffic flows allow the agent to suggest optimal route ensuring QoS (Pham et al. 2018). The model works with two reward functions based on mean of QoS, i.e., packet latency and drops and mean of flows. Additionally, convolution DRL for routing is also employed which minimizes the number of edge weights in the model based on convolution and pooling layers as shown in Fig. 18.9.

A deep neural network fed with network information establishes virtual temporary routes using the Viterbi algorithm to avoid congestion (Lee 2017). Open shortest path first on top of deep belief network finds the most favorable next routing node to reduce the overhead and delay (Mao et al. 2017a). An optimal network scheduler increases the throughput of the infrastructure thus assuring high QoS. By managing the correct sequence of packets at both sender and receiver ends, the retransmission requests can be minimized. It can also reduce the energy consumption in small powered devices like IoT. A DRL with policy gradient-based scheduler employed to learn the packet scheduling by getting rewards from network throughput increased the network capacity by 14.7% (Chinchali et al. 2018). In other work, DRL used in selective caching of the content by identifying the best base station to share the content reduced the transmission delay, thereby increasing the user speed (Wei et al. 2018). In comparison to conventional scheduling policy, deep Q learning on VANETs learn optimal Q value functions from actions, observations, and rewards modeled as Markov decision process reduce the network latency and busy time to achieve high throughput and longer battery life (Atallah et al. 2017). A multilayer perceptron network to predict the free slots in multiple frequencies based on TDMA enhances the network scheduling which reduces the collisions by 50% (Mennes et al. 2018). As the network infrastructure data contains a sequential pattern, employing an LSTM to predict the upcoming traffic load allows changing resource allocation policy to avoid congestion thus, reducing the packet loss rates.

Radio control is another important area in 5G infrastructure, and by balancing the transmission power and optimizing the dynamic spectrum access, the network throughput can be increased multifold. A deep Q agent in DRL aimed to maximize the signal to noise-plus-interference ration by making transmission power adjustments in cognitive radio enhances the spectrum sharing (Yu et al. 2019). Spectrum

access policy customized for every individual user enables high channel throughput. LSTM with deep Q network learns user specific policies by performing training in a distributed fashion at users' end (Naparstek and Cohen 2017) which doubled the channel throughput. Due to the dynamic nature of wireless infrastructure, most of the times, it goes underutilized due to unpredictable traffic. RNN for forecasting the upcoming traffic enables the network to perform dynamic spectrum assignment which increases the network throughput (Rutagemwa et al. 2018.) In the case of adversarial attack such as network jamming, the whole infrastructure collapses. An anti-jamming solution using a DRL agent which learns about the adversarial situation by analyzing the raw spectrum with additional information about the underlying environment avoids the infrastructure collapse and increases the throughput at the same time.

In other applications, deep learning can be used in predicting bit rate for optimal content transfer without loss. Pensieve system uses DRL algorithm to analyze available bandwidth, buffer size and predicts an adaptive bit rate to enhance user QoE up to 25% (Mao et al. 2017b). As base stations in cellular network exhaust, a large amount of energy to remain operational, identifying the underused base stations and switching them on/off based on demand, can reduce the energy consumption. A deep Q learning agent trained on traffic behavior dynamically controls the usage of base station, thus optimizing the resources in the 5G infrastructure (Liu et al. 2018).

### *18.3.4 Network Security*

Security is one of the most sought-out features of today's network. A user demands fast as well as secured infrastructure to save themselves from various adversarial attacks. Wireless network security includes securing users, equipment, data, and unauthorized access in the infrastructure. The existing security measures include different solutions at both user and network ends such as antivirus software, firewall, intrusion detection system, authentication and authorization protocols. (Buczak and Guven 2015), but lack of adaptability to dynamic attacks leaves both user and network open to threats. The adaptable algorithms using deep learning models can safeguard wireless networks by automatically identifying the threats from past experiences (Kwon et al. 2017). The learning from data reduces the manual efforts to define new signatures and rules and enables the system to be protected real time. The unexpected change in network traffic should be further analyzed for anomalies inside it. Specifically, their characteristics can be well explored using auto-encoders where each AE can distinguish between different types of traffic like legitimate, flooding, injection, impersonation, etc. Based on the correct classification, the threat can be blocked in its initial phase, and thus, the infrastructure can be secured (Thing 2017). The AE can also be utilized for feature extraction which is further utilized by any deep learning model such as MLP to accurately identify anomalies in the network (Aminanto and Kim 2016). MLP with AE further increased the classification accuracy by extracting the required features from the traffic flow. Inclusion of IoT devices in the 5G net-

work also leads to distributed attacks which need to be solved. By training a deep learning model on flooding and distributed denial-of-service attacks, the model was able to reduce the false-positive and false-negative classifications thus, accurately identifying attacks without blocking essential services to legitimate devices (Diro and Chilamkurti 2018; Khan et al. 2016; Saied et al. 2016). Due to low power in IoT devices, set of features remain absent, which are typically found in other networks which further make it difficult to identify attacks. This problem is solved by variational auto-encoders which employ matrix completion task to approximate missing features based on available data (Lopez-Martin et al. 2017).

Similarly, security needs to be assured at users' end also where the mobile device can get infected from malicious apps openly available on the internet. A stacked AE to automatically analyze the modules in the app and their call hierarchy to construct a weighted directed graph (Hou et al. 2016). This helps in detecting apps which have been intentionally repackaged and obfuscated to avoid signature detection. A CNN trained with byte code of the app to discover malicious opcodes present in the app allows the device to block execution of such apps (Martinelli et al. 2017; McLaughlin et al. 2017).

## 18.4 Conclusion

The chapter explores the application of existing state-of-the-art machine learning algorithms in the 5G networks. By exploiting the spatial and temporal information present in both user-level and network-level data, many of the infrastructure planning problems can be solved. Through network-level data analysis, upcoming traffic pattern can be predicted accurately to further allocate the resources in heavy request zones or could cluster the underutilized resources to make substantial utilization of them. The classification of various known and unknown protocols even in encrypted traffic through convolution neural networks allows for prioritizing delay-intolerant traffic and stopping malicious traffic to ensure security. User localization through call detail records further allowed to improvise the network utilization by anticipating the burst traffic. Through machine learning models, the network connectivity graph is exploited to its full extent to discover the best routes available for packet delivery. Further, through intelligent radio control, the trade-off between QoS and power consumption is optimized based on the predictive load. Machine learning in 5G also helps in providing security by automatically identifying the malicious or adversary based on previous signatures and experiences.

## References

Aceto G, Ciuonzo D, Montieri A, Pescapé A (2019) Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges. IEEE Trans Netw Serv Manag

Agiwal M, Roy A, Saxena N (2016) Next generation 5G wireless networks: a comprehensive survey. IEEE Commun Survey Tutorials 18(3):1617–1655

Ahmed E, Yaqoob I, Hashem IAT, Shuja J, Imran M, Guizani N, Bakhsh ST (2018) Recent advances and challenges in mobile big data. IEEE Commun Mag 56(2):102–108

Alawe I, Ksentini A, Hadjadj-Aoul Y, Bertin P (2018) Improving traffic forecasting for 5G core network scalability: a machine learning approach. IEEE Netw 32(6):42–49

Aminanto ME, Kim K (2016) Detecting impersonation attack in WiFi networks using deep learning approach. In International workshop on information security applications. Springer, Berlin, , pp 136–147

Atallah R, Assi C, Khabbaz M (2017) Deep reinforcement learning-based scheduling for roadside communication networks. In: 15th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt). IEEE, pp 1–8

Bahl P, Padmanabhan VN (2000) Radar: an in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on computer communications. Nineteenth annual joint conference of the IEEE computer and communications societies (Cat. No. 00CH37064), vol 2, pp 775–784, Mar 2000. https://doi.org/10.1109/INFCOM.2000.832252

Brunato M, Battiti R (2005) Statistical learning theory for location fingerprinting in wireless lans. Computer Netw 47(6):825–845

Buczak AL, Guven E (2015) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surveys Tutorials 18(2):1153–1176

Chen L, Yang D, Zhang D, Wang C, Li J et al (2018) Deep mobile traffic forecast and complementary base station clustering for c-ran optimization. J Netw Comput Appl 121:59–69

Cheng X, Fang L, Hong X, Yang L (2017) Exploiting mobile big data: Sources, features, and applications. IEEE Netw 31(1):72–79

Chinchali S, Hu P, Chu T, Sharma M, Bansal M, Misra R, Pavone M, Katti S (2018) Cellular network traffic scheduling with deep reinforcement learning. In: Thirty-second AAAI conference on artificial intelligence

Davidson P, Piché R (2016) A survey of selected indoor positioning methods for smartphones. IEEE Commun Surveys Tutorials 19(2):1347–1370

De Coninck E, Verbelen T, Vankeirsbilck B, Bohez S, Simoens P, Demeester P, Dhoedt B (2015) Distributed neural networks for internet of things: the big-little approach. In: International internet of things summit. Springer, Berlin, pp 484–492

Diro AA, Chilamkurti N (2018) Distributed attack detection scheme using deep learning approach for internet of things. Fut Gener Comput Syst 82:761–768

Felbo B, Sundsøy PR, Pentland A, Lehmann S, de Montjoye Y (2015) Using deep learning to predict demographics from mobile phone metadata. CoRR. abs/1511.06660, 2015. http://arxiv.org/abs/1511.06660

Feng J, Chen X, Gao R, Zeng M, Li Y (2018) Deeptp: an end-to-end neural network for mobile cellular traffic prediction. IEEE Netw 32(6):108–115

Geyer F, Carle G (2018) Learning and generating distributed routing protocols using graph-based deep learning. In Proceedings of the 2018 workshop on big data analytics and machine learning for data communication networks. ACM, pp 40–45

Gwon YL, Kung H (2014) Inferring origin flow patterns in Wi-Fi with deep learning. In 11th international conference on autonomic computing (ICAC 14), pp 73–83

Heck JC, Salem FM (2017) Simplified minimal gated unit variations for recurrent neural networks. In: 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE, pp 1593–1596

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Hou S, Saas A, Chen L, Ye Y (2016) Deep4MalDroid: A deep learning framework for android malware detection based on linux kernel system call graphs. In: IEEE/WIC/ACM international conference on web intelligence workshops (WIW). IEEE, pp 104–111
Huang CW, Chiang CT, Li Q (2017) A study of deep learning networks on mobile traffic forecasting. In: 2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC). IEEE, pp 1–6
Ibrahim M, Torki M, ElNainay M (2018) CNN based indoor localization using RSS time-series. In: IEEE symposium on computers and communications (ISCC). IEEE, pp 01044–01049
Khan MA, Khan S, Shams B, Lloret J (2016) Distributed flood attack detection mechanism using artificial neural network in wireless mesh networks. Secur Commun Netw 9(15):2715–2729
Kwon D, Kim H, Kim J, Suh SC, Kim I, Kim KJ (2017) A survey of deep learning-based network anomaly detection. Cluster Comput 1–13
Lee Y (2017) Classification of node degree based on deep learning and routing method applied for virtual route assignment. Ad Hoc Netw 58:70–85
Liang VC, Ma RT, Ng WS, Wang L, Winslett M, Wu H, Ying S, Zhang Z (2016) Mercury: metro density prediction with recurrent neural network on streaming CDR data. In IEEE 32nd international conference on data engineering (ICDE). IEEE, pp 1374–1377
Liu J, Krishnamachari B, Zhou S, Niu Z (2018) Deepnap: Data-driven base station sleeping operations through deep reinforcement learning. IEEE Internet Things J 5(6):4273–4282
Lopez-Martin M, Carro B, Sanchez-Esguevillas A, Lloret J (1967) Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in iot. Sensors 17(9):2017
Lotfollahi M, Siavoshani, MJ. Zade RSH, Saberian M (2017) Deep packet: a novel approach for encrypted traffic classification using deep learning. Soft Comput 1–14
Mao B, Fadlullah ZM, Tang F, Kato N, Akashi O, Inoue T, Mizutani K (2017a) Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning. IEEE Trans Comput 66(11):1946–1960
Mao H, Netravali R, Alizadeh M (2017b) Neural adaptive video streaming with Pensieve. In: Proceedings of the conference of the ACM special interest group on data communication. ACM, pp 197–210
Martinelli F, Marulli F, Mercaldo F (2017) Evaluating convolutional neural network for effective mobile malware detection. Procedia Comput Sci 112:2372–2381
McLaughlin N, Martinez del Rincon J, Kang B, Yerima S, Miller P, Sezer S, Safaei Y, Trickel E, Zhao Z, Doupé A et al (2017) Deep android malware detection. In: Proceedings of the seventh ACM on conference on data and application security and privacy. ACM, pp 301–308
Mennes R, Camelo M, Claeys M, Latre S (2018) A neural-network-based mf-tdma mac scheduler for collaborative wireless networks. In: 2018 IEEE wireless communications and networking conference (WCNC). IEEE, pp 1–6
Mohammadi M, Al-Fuqaha A, Guizani M, Oh J-S (2017) Semisupervised deep reinforcement learning in support of iot and smart city services. IEEE Internet Things J 5(2):624–635
Naparstek O, Cohen K (2017) Deep multi-user reinforcement learning for dynamic spectrum access in multichannel wireless networks. In: GLOBECOM 2017-2017 IEEE global communications conference. IEEE, pp 1–7
Nie L, Wang X, Wan L, Yu S, Song H, Jiang D (2018) Network traffic prediction based on deep belief network and spatiotemporal compressive sensing in wireless mesh backbone networks. Wireless Commun Mob Comput 1–10:2018
Ordóñez F, Roggen D (2016) Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors 16(1):115
Paolini M (2017) Mastering analytics: how to benefit from big data and network complexity. http://content.rcrwireless.com/20170620MasteringAnalyticsReport

Pham TAQ, Hadjadj-Aoul Y, Outtagarts A (2018) Deep reinforcement learning based QOS-aware routing in knowledge-defined networking. In: International conference on heterogeneous networking for quality, reliability, security and robustness. Springer, pp 14–26

Pierucci L, Micheli D (2016) A neural network for quality of experience estimation in mobile communications. IEEE Multimedia 23(4):42–49

Rutagemwa H, Ghasemi A, Liu S (2018) Dynamic spectrum assignment for land mobile radio with deep recurrent neural networks. In: 2018 IEEE international conference on communications workshops (ICC Workshops). IEEE, pp 1–6

Saied A, Overill RE, Radzik T (2016) Detection of known and unknown ddos attacks using artificial neural networks. Neurocomputing 172:385–393

Schmidhuber J (2015) Deep learning in neural networks: An overview. Neural Netw 61:85–117

Singh A, Verma S (2017) Graph laplacian regularization with procrustes analysis for sensor node localization. IEEE Sens J 17(16):5367–5376. ISSN 1530-437X. https://doi.org/10.1109/JSEN.2017.2717919

Thing VL (2017) IEEE 802.11 network anomaly detection and attack classification: a deep learning approach. In: 2017 IEEE wireless communications and networking conference (WCNC). IEEE, pp 1–6

Wang J, Zhang X, Gao Q, Yue H, Wang H (2016) Device-free wireless localization and activity recognition: a deep learning approach. IEEE Trans Veh Technol 66(7):6258–6267

Wang J, Tang J, Xu Z, Wang Y, Xue G, Zhang X, Yang D (2017a) Spatiotemporal modeling and prediction in cellular networks: a big data enabled deep learning approach. IEEE INFOCOM 2017—IEEE conference on computer communications, pp 1–9

Wang W, Zhu M, Wang J, Zeng X, Yang Z (2017b) End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In: 2017 IEEE international conference on intelligence and security informatics (ISI). IEEE, pp 43–48

Wang X, Gao L, Mao S, Pandey S (2015) DeepFi: deep learning for indoor fingerprinting using channel state information. In: IEEE wireless communications and networking conference (WCNC). IEEE, pp 1666–1671

Wang Z (2015) The applications of deep learning on traffic identification. BlackHat USA, 24

Wei Y, Zhang Z, Yu FR, Han Z (2018) Joint user scheduling and content caching strategy for mobile edge networks using deep reinforcement learning. In: IEEE international conference on communications workshops (ICC Workshops). IEEE, pp 1–6

Wu CL, Fu LC, Lian F-L (2004) WLAN location determination in e-home via support vector classification. In: IEEE international conference on networking, sensing and control, vol 2. IEEE, pp 1026–1031

Xia S, Liu Y, Yuan G, Zhu M, Wang Z (2017) Indoor fingerprint positioning based on Wi-Fi: an overview. ISPRS Int J Geo-Inf 6(5):135

Xiao J, Wu K, Yi Y, Ni LM (2012) FIFS: fine-grained indoor fingerprinting system. In: 21st international conference on computer communications and networks (ICCCN). IEEE, pp 1–7

Xiao J, Zhou Z, Yi Y, Ni LM (2016) A survey on wireless indoor localization from the device perspective. ACM Comput Surveys (CSUR) 49(2):25

Youssef M, Agrawala A (2005) The Horus WLAN location determination system. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services. ACM, pp 205–218

Yu Y, Wang T, Liew SC (2019) Deep-reinforcement learning multiple access for heterogeneous wireless networks. IEEE J Sel Areas Commun 37(6):1277–1290

Zhang C, Patras P (2018) Long-term mobile traffic forecasting using deep spatio-temporal neural networks. In: Proceedings of the eighteenth ACM international symposium on mobile ad hoc networking and computing. ACM, pp 231–240

Zhang W, Liu K, Zhang W, Zhang Y, Gu J (2016) Deep neural networks for wireless localization in indoor and outdoor environments. Neurocomputing 194:279–287

Zhou G-B, Wu J, Zhang C-L, Zhou Z-H (2016) Minimal gated unit for recurrent neural networks. Int J Autom Comput 13(3):226–234