Signals and Communication Technology

Mohammad Abdul Matin *Editor*

A Glimpse Beyond 5G in Wireless Networks

Springer

Signals and Communication Technology

This series is devoted to fundamentals and applications of modern methods of signal processing and cutting-edge communication technologies. The main topics are information and signal theory, acoustical signal processing, image processing and multimedia systems, mobile and wireless communications, and computer and communication networks. Volumes in the series address researchers in academia and industrial R&D departments. The series is application-oriented. The level of presentation of each individual volume, however, depends on the subject and can range from practical to scientific.

Indexing: All books in "Signals and Communication Technology" are indexed by Scopus and zbMATH.

For general information about this book series, comments or suggestions, please contact Mary James at mary.james@springer.com or Ramesh Nath Premnath at ramesh.premnath@springer.com.

Mohammad Abdul Matin
Editor

A Glimpse Beyond 5G in Wireless Networks

Springer

Editor
Mohammad Abdul Matin
Department of Electrical and Computer
Engineering
North South University, Dhaka, Bangladesh

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-031-13785-3 ISBN 978-3-031-13786-0 (eBook)
https://doi.org/10.1007/978-3-031-13786-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This one is for Zabeer Ahmed and Zawad Ahmed, My love, my heart, my light

Preface

Introduction

The beyond 5G (B5G) technology is the successor of 5G with new user requirements that provide up to Tbit/s peak data rate, expected bandwidth efficiency of two to three times compared with 5G, user experienced data rate of 10–100 Gb/s, latency of less than 0.1 ms, and mobility support of higher than 1000 km/h [1]. It facilitates completely new applications/use cases such as haptics, telemedicine, brain-machine interfaces, mixed reality, virtual reality, and augmented reality. Such requirements are deemed beyond 5G capabilities as they do not belong to any of the three main use cases which are mMTC, URLLC, and eMBB [2]. B5G is expected to be on the verge of the all-pervasive wireless Internet of Everything, capable of supporting mobile traffic in diversified scenarios with mixed key performance indicators. There are many challenges that must be resolved to achieve the targeted speed, power efficiency, spectrum efficiency, reliability, and latency requirements of B5G. The aim of this book is to gather the latest research findings on emerging trends in wireless systems. It will present and assess different enabling technologies, capabilities, and anticipated communications and computing solutions for 5G and beyond 5G. Topics discussed include new frequency bands, new network deployment, "Deep Physical Layer" (DPL) located below the physical layer of the OSI model, multiuser massive MIMO, relay-based cooperative communication, NOMA for massive IoT (mIoT), vehicular communications (from V2V to V2X), and so on. These will help the readers to understand more advanced research materials for developing new ideas for B5G to make a contribution in this field for themselves. It is hoped that this book will serve as a virtual and effective bridge between academic research in theory and engineering development in practice.

Research Contributions

This book attempts to explore the cutting-edge research developments in 5G and beyond technologies and is organized into 10 chapters.

The chapter "***A Comprehensive Study on 5G: RAN Architecture, Enabling Technologies, Challenges and Deployment***" provides an insight into 5G E2E architectures for private and public networks based on the 3GPP, 5GPPP, and O-RAN alliance standards. Besides network virtualization and network slicing, strategies have also been presented along with R&D progress on 5G technologies. A comparative analysis of discussed RAN architectures that have been proposed by different standardization bodies helps to comprehend categorically the architectural differences, limitations, benefits, and deployment scenarios.

The chapter "***Information Flows at the Deep Physical Layer Level***" describes the wireless communication process from a physical point of view introducing the "Deep Physical Layer" (DPL), located below the physical layer of the OSI model. Based on this model, the numerous solutions proposed for 6G that include intelligent modifications of the propagation environment can be seen as "data processing" at the DPL level in order to optimize the performance of the telecommunications system.

The chapter "***FBMC: A Waveform Candidate for Beyond 5G***" investigates the design of FBMC-OQAM waveform based multiple-input multiple-output (MIMO) and multi-user massive MIMO systems. It starts with a brief description of the important features along with the differences of FBMC waveform over widely popular OFDM waveform, followed by the discussion over key challenges in designing FBMC-based MIMO and massive MIMO systems. To compare the performance of FBMC and OFDM waveforms in the uplink transmission, the achievable sum rates are derived for multi-user (MU) massive MIMO technology relying on FBMC waveform with maximum ratio combining (MRC) and zero-forcing (ZF) receivers. The corresponding power scaling laws for MU massive MIMO-FBMC are also found. It is shown that in practical impairments such as carrier frequency offset, massive MIMO-FBMC systems significantly outperform their OFDM counterparts.

The chapter "***Full-Duplex Multi-hop Communication for Beyond 5G***" presents full-duplex massive multi-input multi-output (MIMO) wherein multiple pairs of full-duplex users are communicating with each other via a full-duplex massive relay. Relay-based cooperative communication improves the diversity gain, expands the coverage, enhances capacity, and reduces the total transmission power of wireless communication systems. The full-duplex (FD) system can facilitate increased performance than its half-duplex (HD) counterpart. The input-output equation derived in this chapter can be applied to many other 5G and beyond technologies.

Chapter "***NOMA for 5G and Beyond Wireless Networks***" investigates an analytical framework of heterogeneous cellular network (HCN) with three tiers, namely macro base station (MBS) tier underlaid with femto base station (FBS) tier and device-to-device (D2D) tier. NOMA principle is applied in the FBS tier and the D2D tier, while the MBS tier does not use NOMA. Offloading from the

MBS tier to the FBS tier aids in tackling congestion at the MBS tier. It is observed that using the proposed D2D cooperation, the performance of offloading in HetNet can be further enhanced, therefore is suitable for the 5G and beyond networks. Furthermore, offloading with NOMA leads to massive connectivity, which is vital for the 5G and beyond networks.

In chapter "***Energy Harvested Device-to-Device MIMO Systems for Beyond 5G Communication***," the energy harvesting (EH) has been addressed in terms of linear and non-linear methods. EH through ambient resources such as RF signal-based simultaneous wireless information and power transfer (SWIPT) provides a viable solution by enabling the energy constraint nodes to scavenge the energy. The EH implementation protocols have been discussed in this chapter before introducing the device-to-device (D2D) multiple-input and multiple-output (MIMO) relays along with SWIPT. This chapter also examines the practical constraints such as feedback errors, imperfect channel state information over the system performance, and the useful insights in system design considering the impacts of channel conditions, multiple antennas, and energy harvesting parameters.

The chapter "***Vehicular Communications in the B5G Era***" presents the requirements and challenges in the design and analysis of advanced wireless communication systems for vehicular communications (from V2V to V2X). Different types of vehicular network configurations, in terms of antenna subsystems and transceivers, have been described in relation with the impact of dense urban, urban, and rural scenarios in system implementation. Examples of coverage/capacity relations for vehicular communication scenarios have also been presented based on volumetric geometric/stochastic modelling techniques.

The chapter "***Implementation of Context-Aware Environments with Massive IoT Systems***" describes the framework for IoT evolution, from current LPWAN/5G connectivity to future B5G systems, taking advantage of sub THz (mainly in the 100–300 GHz frequency range) and THz bands (up to 10 THz). The current needs and future evolution of IoT applications require device integration, node density, interference, and energy handling. Coverage/capacity estimations for different case uses within dense urban/urban/sub-urban settings are being presented based on deterministic volumetric wireless channel estimation. Different application scenarios, such as the evolution of current IoT applications towards sensing networks, are also discussed based on the description of three realistic use case scenarios.

The chapter "***Feasibility of LDM to Serve User-IoT Pairs in the Future Wireless Network***" presents a transmission framework where the Layer division multiplexing (LDM) layer serves IoT-user pairs. LDM is considered as a potential technology that can enhance network capacity by taking advantage of this inherent heterogeneity of future wireless networks. The IoT devices are served using an LDM upper layer (UL), and the users are served using a lower layer (LL). The authors have developed a physical layer model incorporating LDM and tested its performance for the intended usages scenario. Both UL and LL performance show the capability to serve IoT devices and users to justify the proposed transmission scenario.

The chapter "***Wide Band THz Antenna Design Using Salp Swarm Algorithm for 6G Communications Systems***" provides a complete framework for circular

polarized antenna design in the low THz band. THz communications have the potential to provide an order of magnitude capacity improvement and thus considered to be an attractive candidate for 6G wireless networks. The presented optimization framework is based on a recently introduced swarm intelligence algorithm, namely the Salp Swarm Algorithm (SSA). The numerical results show that the SSA has been successfully applied in designing antenna with wide band operation and circular polarization.

Conclusions

Moving from 5G to B5G, the expectations are increasing about what the current and future generations of wireless communication can do for a wide range of applications. It is believed that the students, professionals, practitioners, and academic and industrial researchers who desire to work in the latest development of 5G and beyond technologies will need this book. The volume

- Reviews 5G network design methodologies as well as offers an in-depth analysis to understand future wireless networks
- Focuses on different emerging technologies and their applications under 5G and B5G
- Provides state-of-the-art solutions to the theoretical and practical challenges in network design, operation, management, and optimization.

References

1. Shuangfeng Han, Tian Xie, Chih-Lin I, Li Chai, Zhiming Liu, Yifei Yuan, Chunfeng Cui, “Artificial-Intelligence-Enabled Air Interface for 6G: Solutions, Challenges, and Standardization Impacts”, vol. 58, no. 10, pp. 73–79, IEEE Communications Magazine, October 2020.
2. Mohammad A. Matin, Mohammad S. Sharawi, Sotirios K. Goudos, Jiang Zhu “Antenna Systems for 5G and Beyond” IEEE Communications Magazine, Jan. 2022

Dhaka, Bangladesh Mohammad Abdul Matin

Acknowledgments

In the first place, I am grateful to my Lord Almighty ALLAH who helped and guided me throughout my life. I would like to acknowledge the effort and time invested by all contributors for their excellent work in completing this book, also thanks to the authors whose manuscripts were not included due to rigorous review process. I am indebted to numerous reviewers for valuable suggestions and comments during the review process. The authors were so cooperative during different stages of book development. I am also beholden to the Springer team, in particular Mary James for her cordial assistance throughout this project. I wish to include here a word of appreciation to them including Mr. Arun Siva Shanmugam for smoothly and efficiently handling the project from the beginning until the final stage of production and keeping it in line with the publisher's policies. I would also like to thank my wife Momtaz Begum and my children Zabeer Ahmed and Zawad Ahmed for their constant encouragement, patience, and understanding throughout the project. My special thanks to Mrs. Sufia Khaton (dearest mother) for the motivation that has encouraged me to keep going.

North South University, Bangladesh Mohammad Abdul Matin

Contents

Editor and Contributors

About the Editor

Dr. Mohammad Abdul Matin (Senior Member, IEEE) is a professor in the Department of Electrical and Computer Engineering at North South University (NSU), where he has been since 2008. He was first appointed as an assistant professor and then promoted to associate professor in 2011 and later on professor at North South University. While in that post, he was also the coordinator of EEE program. During 2012–2017, he was an associate professor at Universiti Teknologi Brunei (UTB), Brunei Darussalam. He received his BSc degree in electrical and electronic engineering from BUET (Bangladesh), MSc degree in digital communication from Loughborough University, UK; and PhD in wireless communication from Newcastle University, UK. Dr. Matin has been a visiting academic staff at the National University of Malaysia (UKM), the University of Malaya (UM), and other universities and institutions. He has published more than 100 peer-reviewed journal and conference papers. He is the author/editor of 17 academic books and 19 book chapters. Dr. Matin serves as a member of the editorial board of several international journals including IEEE Communications Magazine and IET Wireless Sensor Systems. He has received several prizes and scholarships including the Best Student Prize (Loughborough University), Commonwealth Scholarship, and Overseas Research Scholarship (ORS) conferred by the Committee of Vice Chancellors and Principals (CVCP) in the UK.

Contributors

R. Abozariba School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

A. V. Alejos Campus Universitario, Escola de Ingeniería de Telecomunicación, Universidad de Vigo, Vigo, Spain

M. Alfaqawi Altran Labs, Velizy-Villacoublay, France

A. T. Asyhari School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

L. Azpilicueta School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL., Mexico

V. Bhatia Indian Institute of Technology Indore, Indore, Madhya Pradesh, India

M. Celaya-Echarri School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL, Mexico

S. Davai Altran ACS, Saint-Herblain, France

F. Falcone EEC Department, Universidad Publica de Navarra, Edificio Los Tejos, Pamplona, Navarra, Spain

Electric, Electronic and Communication Engineering Department, Public University of Navarra, Pamplona, Navarra, Spain

M. Gateau Altran ACS, Saint-Herblain, France

S. K. Goudos ELEDIA@AUTH, Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece

P. Huard Altran ACS, Saint-Herblain, France

M. S. Islam School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

M.-C. Le Mercier Altran ACS, Saint-Herblain, France

P. Lopez-Iturri Electric, Electronic and Communication Engineering Department, Public University of Navarra, Pamplona, Navarra, Spain

M. B. Mabrouk Altran Labs, Velizy-Villacoublay, France

M. A. Matin Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

M. D. Migliore DIEI, University of Cassino and Souther Lazio, Cassino, Italy

M. Patwary Department of Computer Science, University of Wolverhampton, Wolverhampton, UK

I. Picallo Electric, Electronic and Communication Engineering Department, Public University of Navarra, Pamplona, Navarra, Spain

P. Reungoat Altran ACS, Saint-Herblain, France

P. Shaik Department of Electrical Engineering, IIT, Indore, India

E. Sharma Indian Institute of Technology, Roorkee, India

P. Singh International institute of Information Technology, Bangalore, India

P. Swami Indian Institute of Technology Indore, Indore, Madhya Pradesh, India

C. Vargas-Rosales School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL., Mexico

Chapter 1
A Comprehensive Study on 5G: RAN Architecture, Enabling Technologies, Challenges, and Deployment

Mohammed Alfaqawi, Martine Gateau, Patrick Huard, Pascal Reungoat, Marie-Christine Le Mercier, Stéphane Davai, and Mouna Ben Mabrouk

1.1 Introduction

Different from its predecessors, the 5G eco-system has been developed not only to enable human-to-human (H2H) communication but also to provide services for human-to-machine (H2M) and machine-to-machine (M2M) communications. In order to fulfill the diverse requirements of the vast and wide-scale applications of mobile network operators (MNOs), vertical industries, and third-party tenants, 5G is targeting three main families of use cases (UCs): (1) enhanced mobile broadband (eMBB) for the classical applications of H2H communication with data rate up to 10–20 Gbps, (2) massive machine-type communication (mMTC) to provide services for 100 billion end-user devices with density of 10^6 devices/km^2 and stringent requirements on energy consumption, and (3) ultra-reliable low latency communication (URLLC) for critical and delay-sensitive applications. These three UCs can be considered as an umbrella that covers a myriad of applications with varying requirements such as tactile Internet, video gaming and streaming, virtual/augmented reality (VR/AR), autonomous cars, smart transportation, smart agriculture, smart home and factory, medical applications, and so on.

The multiplicity of applications and connected devices with diverse requirements in terms of latency, bandwidth, traffic volume, resources, and energy mandated drastic changes on the cellular network. These changes involve new radio (NR), flexible, programmable, virtualized, and automated radio access network (RAN). New electromagnetic bands above 6 GHz, numerologies, modulation and coding

M. Alfaqawi (✉) · M. B. Mabrouk
Altran Labs, Velizy-Villacoublay, France
e-mail: mohammed.alfaqawi@capgemini.com

M. Gateau · P. Huard · P. Reungoat · M.-C. Le Mercier · S. Davai
Altran ACS, Saint-Herblain, France
e-mail: martine.gateau@capgemini.com

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_1

schemes, and waveforms with mini-slots have been proposed for the 5G NR. In addition, small cells (SCs), massive MIMO, edge computing, and distributed and centralized units with several deployment options have been introduced to the next generation-RAN (NG-RAN) to boost data rate and minimize latency. Meanwhile, software-defined networking (SDN), network functions virtualization (NFV), and network slicing technologies have been adopted within the 5G network in order to virtualize, orchestrate, and automate the applications and services operating on the same infrastructure.

Global telecommunication bodies, including the international telecommunication union (ITU), 5G public-private partnership (5GPPP), next-generation mobile network (NGMN) alliance, European telecommunications standards institute (ETSI), third-generation partnership project (3GPP), and open-RAN (O-RAN) alliance, are involved to standardize and accelerate the deployment of 5G networks. The ITU-R identified the requirements for 5G NR or international mobile telecommunications 2020 (IMT-2020). Furthermore, 5GPPP has published the first white paper for 5G architecture and launched several 5G projects, including 5G-Crosshaul, 5G Exchange (5GEx), flex5Gware, and SPEED-5G [1]. The NGMN alliance identified several UCs, scenarios, and business models for 5G and their requirements. Moreover, ETSI participates on standardizing the mobile or multi-access edge computing (MEC) technology and NFV. Meanwhile, 3GPP publishes several technical standards to identify the essential specifications and regulations for the 5G NR. In addition, it provides solutions for the vertical market such as automotive, transport, industry, and e-health. Considering several enhancements are still ongoing for the 5G eco-system, 3GPP develops release 17 to be completed in 2021. On the other hand, O-RAN alliance [2] aims at providing open, virtualized, and artificial intelligence (AI)-powered RAN.

1.1.1 Review of Related 5G Surveys

The state of the art on the 5G technology has been discussed in several surveys such as [3–20] and research works cited therein. This chapter addresses three main 5G topics involving 5G public and private architectures, key 5G enabling technologies, and 5G development and projects. Considering these topics, Table 1.1 presents and compares the scope, contributions, and limitations of the most relevant research works for the scope of this chapter.

The 5G architectural frameworks and challenging issues have been explored in several surveys such as in [3–8]. Habibi et al. [3] investigated comprehensively the 5G cloud-RAN (C-RAN) architecture. However, the 5G architecture in [3] addressed C-RAN for mobile operators or public networks considering remote radio head (RRH) and baseband unit (BBU). Similar to [3], Hossain et al. [4] reviewed several challenging issues and research works on enhancing C-RAN, including throughput, interference, energy efficiency, latency, and so on. In addition, the scope of [5] focused on reviewing extensively the resource allocation challenges

Table 1.1 Review for the most relevant surveys

Ref.	Theme	Scope and contributions	Issues
[3]	5G, RAN	A comprehensive comparison between the current and previous RAN architectures including in-depth study for C-RAN. It studied the 5G enabling technologies such as SDN, NFV, and network slicing	The various 5G RAN architectures presented in [3] are based on C-RAN that separates the 5G networks into RRH and BBU. However, in the NG-RAN provided by 3GPP and O-RAN alliance, the BBU is separated into DU and CU with several split options. These modifications on the NG-RAN have not been considered in [3]. Moreover, it presented the E2E 5G architecture for the mobile operators or public networks only. However, the E2E architecture and solutions for the enterprise's private networks have not been discussed in [3]. Meanwhile, other edge computing paradigms, e.g., MEC, have not been explored. Furthermore, an in-depth study for the research works to overcome the several challenges on SDN, NFV, and network slicing has not been addressed
[10]	5G, IoT	It studied the IoT E2E architecture considering 5G and LPWANs technologies. It classified the E2E 5G IoT architecture into sensors, network, communication, architectural, and application layers. Besides, it provided a comparative study for the LPWANs technologies and highlighted some IoT UCs and challenges such as security	The E2E architecture in [10] employed the LPWANs technologies to connect the sensor layer with the network layer, while the 5G technology was utilized to connect the network layer with the communication layer. However, the E2E architecture studied in [10] does not leverage the capabilities of the 5G eco-system. For instance, the LPWANs technologies cannot satisfy the requirements of URLLC UCs for a latency close to zero. In addition, the E2E architecture is cloud-based that does not consider edge computing paradigms for real-time processing and low latency services. Moreover, the enabling technologies for 5G technology have not been discussed in [10]
[20]	Edge computing, SDN	It provided a comprehensive and comparative study for several edge computing paradigms, including MEC and fog computing. Moreover, it discussed the SDN and NFV technologies	Edge computing paradigms, SDN, and NFV have not been investigated within the 5G architecture. Moreover, the 5G E2E architecture and network slicing have not been discussed in [20] nor the architecture and role of SDN and NFV in 5G networks

in C-RAN. Meanwhile, Tian et al. [6] provided an in-depth study on the security challenges, issues, and solutions in C-RAN. Furthermore, Agiwal et al. [7] discussed the architectural enhancements on the 5G RAN including the physical and medium access controller (MAC) layers. Moreover, Sharma et al. [8] highlighted the several challenges of mMTC such as quality of service (QoS), RAN congestion, handover, and so on. In [9–13], Internet of Things (IoT) networks have been investigated by considering 5G connectivity in conjunction with several IoT wireless technologies. In [14], the security protocols and advantages of the 5G IoT networks have been investigated. However, the research works in [9–14] did not consider the several enhancements on 5G networks in order to fulfill the diverse IoT applications and requirements. In contrast to these surveys, we investigate, herein, not only C-RAN architectural framework for the public 5G networks but also other public 5G architectures such as O-RAN and the implementation of MEC with NG-RAN as well as private 5G networks, which has not been discussed in [3–14]. Moreover, unlike [9–11], edge computing paradigms have been addressed for the NG-RAN. In [15, 20] and research works cited therein, 5G MEC approaches for vertical industries have been studied. In [20], cloud computing, edge computing paradigms, SDN, and NFV technologies have been explored comprehensively but not within the 5G networks. For instance, the 5G end-to-end (E2E) architecture, network slicing, and the role of SDN and NFV in 5G networks have not been discussed in [20]. Moreover, the 5G enabling technologies, e.g., SDN, NFV, and network slicing, have been addressed briefly in [3, 7, 11] and [16–19]. Unlike these research works, herein, the concept, architectural frameworks, challenges, and recent research works on the key 5G enabling technologies are addressed in depth.

1.1.2 Contributions and Chapter Organization

The main scope of this chapter is the 5G E2E architectures for public and private networks considering the standards of 3GPP, 5GPPP, and O-RAN alliance. In addition, SDN, NFV, and network slicing, artificial intelligence/machine learning (AI/ML), and MEC technologies have been investigated comprehensively. Furthermore, we review the R&D progress on 5G technology to deliver E2E services including projects and simulators. Despite our comprehensive study on the NG-RAN, some technologies for the NG-RAN, e.g., massive MIMO and beamforming, are out of the scope of this chapter. The contributions of this chapter are manifold as follows:

1. Investigating and comparing the 5G architectures for centralized and partially cloud-RAN (C-RAN), heterogeneous C-RAN (H-CRAN), virtualized C-RAN (V-CRAN), and fog C-RAN (F-CRAN) considering benefits, limitations, and deployment scenarios provided by 3GPP for public networks.

2. An in-depth study on the 5G E2E architecture proposed by 3GPP, O-RAN, and 5GPPP involving MEC and O-RAN deployment scenarios.
3. Exploring the private 5G networks for vertical industries including architectural frameworks and spectrum options.
4. A comprehensive review on the key 5G enabling technologies, i.e., SDN, NFV, network slicing, AI/ML methods, and MEC, including concept, standards, architectural frameworks, and recent research works.
5. Addressing the 5G R&D software and projects for RAN and core including features, benefits, and limitations.
6. Summary on the challenging issues for 5G deployment.

This chapter is organized as follows:

Section 1.2 presents several architectural frameworks for the public 5G networks involving the concept, standards, benefits, and limitations. The explored 5G architectures in Sect. 1.2 are classified into cloud-based and edge-based frameworks. In Sect. 1.2.1, several cloud-RAN (C-RAN) architectural frameworks are reviewed including centralized and partially C-RAN, H-CRAN, and V-CRAN. In Sect. 1.2.2, 5G edge cloud architectural frameworks are investigated involving F-CRAN, O-RAN, and NG-RAN with MEC. These architectural frameworks are compared in Sect. 1.2.3 in terms of deployment scenario, architecture differences, benefits, and limitations.

Section 1.3 addresses the architectural frameworks and spectrum options for private 5G networks. Section 1.3.1 presents the proposed architecture classifications by MNOs for private 5G networks including non-public network (NPN) and shared-RAN. In addition, Sect. 1.3.2 highlights the spectrum options for industry and private 5G networks including shared and unlicensed spectrum.

Section 1.4 investigates the key 5G enabling technologies including SDN, NFV, network slicing, AI/ML, and MEC. In Sects. 1.4.1 and 1.4.2, the concept, architectural framework, benefits, challenges, and recent research works on SDN and NFV, respectively, are explored. Then, in Sect. 1.4.2, network slicing is discussed involving concept, architecture, benefits, and challenges. After that, Sect. 1.4.3 presents AI/ML strategies and methods and 5G UCs to integrate AI/ML. Furthermore, recent AI/ML research works are reviewed to tackle 5G challenging issues. Section 1.4.4 highlights MEC within 5G technology including concept, benefits, architectural framework, and research works.

Section 1.5 presents software (SW) and projects in 5G research and development (R&D). Section 1.5.1 explores and compares the features of several development platforms and SW including OpenAirInterface (OAI), network simulator-3 (NS-3), and MATLAB. Then, Sect. 1.5.2 reviews two 5G projects to provide programmable, flexible, and agile 5G services.

Lastly, Sect. 1.6 summarizes the contribution of this chapter including challenging issues on 5G technology.

1.2 5G Architecture for Public Networks

The 5G technology aims to provide agile, scalable, programmable, and automated network services with ultralow latency and ultrahigh data rate, coverage, availability, reliability, and positioning accuracy. In order to achieve high QoS, the 5G ecosystem adopts drastic changes on the NG-RAN and core involving new radio, enabling technologies and deployment options. Due to the several enhancements required on the 5G RAN and core, the 5G technology is envisaged to be deployed in a non-standalone (NSA) phase, first, to pave the way for the standalone (SA) phase. The NSA phase is to facilitate a smooth transition from the evolved packet core (EPC) of the long-term evolution (LTE) technology into the 5G core (5GC). In the SA deployment scenario, the 5G network operates independently of the EPC where the NG-RAN will be connected directly into the 5GC [21–23]. Based on the SA and NSA deployment phases, eight architectural options have been discussed in [24] for 5G and LTE networks as follows:

1. Option 2 or SA: The gNB is connected directly into the 5GC and isolated from the LTE architecture.
2. Option 3, 3a and 3x or NSA: It is the first commercial 5G architecture to address the mobile broadband (MBB) services. In this option, the gNB is connected to the LTE EPC.
3. Option 4: The LTE BS or the eNB is connected to the 5GC via the gNB.
4. Option 5: The eNB is connected directly to the 5GC.
5. Option 7: The gNB is connected to the 5GC via the eNB.
6. Options 6 and 8:[1] The gNB is connected directly to the EPC.

This section investigates the 5G architectures considering the SA deployment scenario. Several 5G architectural frameworks have been proposed by the 5G telecommunication bodies for 5G SA deployment. Table 1.2 presents the 5G SA architectures proposed by the international business machines (IBM) corporation, O-RAN alliance, 5GPPP, and 3GPP for public 5G networks as well as MuLTEfire

Table 1.2 References for 5G SA architectures

Architecture	References	Section
C-RAN	[25–28]	1.2.1
H-CRAN	[27], and [29–33]	1.2.1.1
V-CRAN	[27], and [34–38]	1.2.1.2
F-CRAN	[39–41]	1.2.2.1
O-RAN	[2], and [42–44]	1.2.2.2
NG-RAN	[1], and [21, 45, 46]	1.2.2.3
Private 5G	[47–54]	1.3

[1] Options 6 and 8 have been withdrawn of further study by 3GPP due to the various limitations on compatibility between the 5G RAN and the EPC.

and MNOs for private 5G networks. These architectures will be explored in depth in this section and Sect. 1.3.

1.2.1 Cloud-Based 5G Architecture

The vast and massive number of devices and applications are foreseen to overwhelm the resources of traditional wireless networks. In 2025, 60 billion devices are expected to be connected to Internet, including 82 UCs, 20 industry-specific UCs, 110 million connected cars, and 1.2 million connected homes [55]. In order to provide a high QoS, e.g., high capacity and coverage, for the end-users, cell densification has been studied as a solution by installing more BSs in RAN [27]. However, cell densification was found to increase the complexity of interference management and intercell coordination. In order to overcome the high density of RAN, cloud-RAN (C-RAN) has been considered for 5G networks and beyond [27].

The C-RAN concept is based on allocating minimal processing of the evolved-universal terrestrial radio access network (E-UTRAN) stack to RRHs while the intensive processing to be performed by BBUs [3, 27]. Typically, the RRH is a small base station (BS) that is composed of radio units and antennas placed anywhere in the network. It performs minimal processing of E-UTRAN stack, such as L1 or physical (PHY) layer. The rest of the processing of E-UTRAN protocol is migrated to be executed by the BBU. The BBU represents a pool of central and shared computational resources placed in the 5GC or the cloud to perform the intensive signal processing tasks and coordination functionalities [26, 28]. Figure 1.1 presents the C-RAN architecture proposed by IBM corporation [25]. In Fig. 1.1, the RRH communicates with the BBU via a fronthaul link that represents an optical fiber with common public radio interface (CPRI) protocol or wireless link.

The C-RAN is categorized into fully and partially centralized C-RAN, as depicted in Fig. 1.2. In the fully centralized C-RAN, the operation and management (O&M) as well as all the E-UTRAN stack layers are processed on the BBU pool, including L1, L2, and L3.[2] Meanwhile, in the partially centralized C-RAN, O&M, L2, and L3 are processed by the BBU, while L1 and the radio functionalities are managed by the RRH [1, 3]. Thanks to the shared resources in C-RAN architecture, the 5G eco-system is foreseen to gain several benefits in terms of capital and operational expenditure (CAPEX/OPEX) reduction, boosting the network performance, enhancing the spectrum and energy efficiencies, supporting several protocols and standards, and so on [3, 37, 56, 57].

Despite these benefits, several challenges and issues are facing C-RAN including security, privacy, and low latency [3]. A shared BBU pool demands a universal security and privacy mechanism. In this context, the security challenges and

[2] L2 represents the MAC, radio link control (RLC), and packet data convergence protocol (PDCP). In addition, L3 represents the radio resource control (RRC).

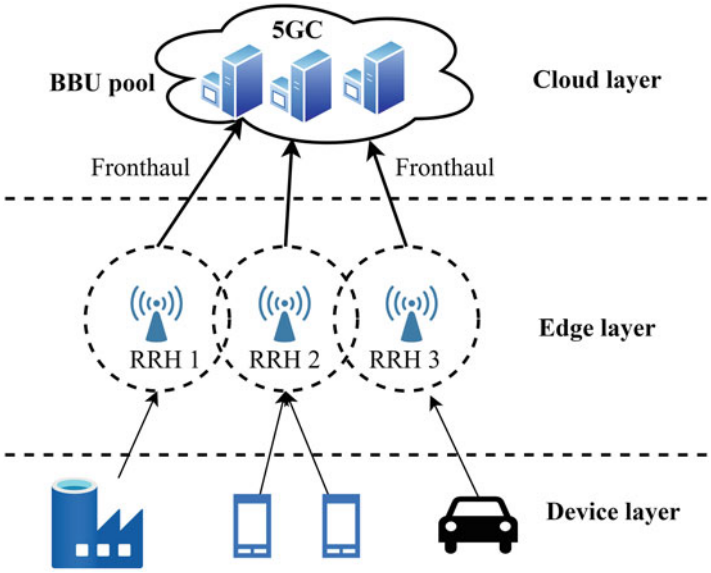


Fig. 1.1 C-RAN architecture

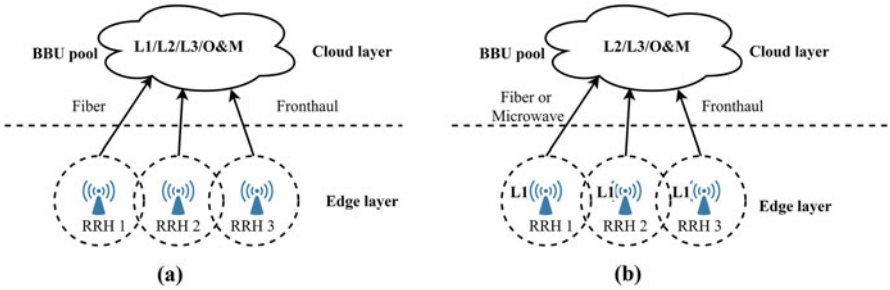


Fig. 1.2 (**a**) Fully centralized C-RAN and (**b**) partially centralized C-RAN

solutions for C-RAN have been investigated in [6]. Another challenging issue for C-RAN is the demand for efficient frameworks for BBU resource allocations in order to achieve high QoS for the end-users. This challenging issue has been explored in depth in [5]. On the other hand, the high data traffic generated from the massive number of end-user devices might overwhelm the fronthaul link between the RRH and the BBU in C-RAN. In order to cope with this challenge, 5G provides H-CRAN and adapts virtualization and softwarization of the network by providing V-CRAN.

1.2.1.1 Heterogeneous C-RAN (H-CRAN)

5G promises to achieve more than 1000 Gbit/s/km^2 area spectral capacity with stringent requirements on the E2E latency and energy consumptions [32]. In order

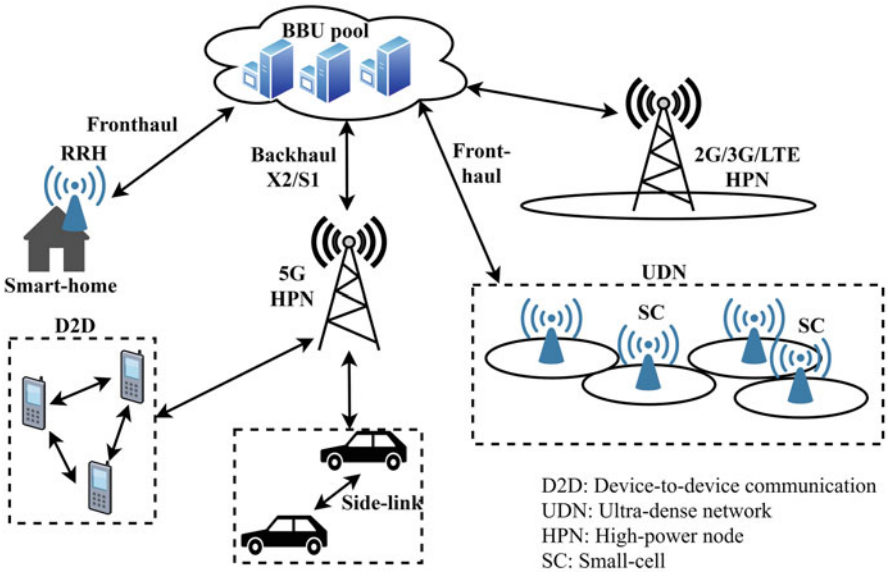


Fig. 1.3 5G network heterogeneity with H-CRAN design

to provide connectivity for the massive number of devices with high capacity and QoS, network densification[3] has been proposed in the past years [58]. However, the ever increasing of devices and applications leads the macro-cells to reach the limit of theoretical capacity [29]. In order to overcome this challenge, a dense heterogeneous network (HetNet) was proposed by deploying low power nodes (LPN) or small cells (SCs), such as pico- and femto-cells, with high power nodes (HPN) such as micro- and macro-cells [30]. Figure 1.3 shows the heterogeneity of the 5G network with several applications, such as device-to-device (D2D) and vehicular communication, and wireless technologies, such as 2G/3G/LTE. However, several challenges have been raised for the HetNets such as resource allocation, intercell coordination, inter-tire interference management, network capacity, coverage, energy efficiency, and CAPEX/OPEX [31, 59]. In order to cope with these challenges, H-CRAN architecture with centralized cloud and virtualization has been proposed for the 5G networks and beyond to manage the intercell coordination and facilitate the deployment of macro and SCs [27, 30, 31].

H-CRAN enhances the performance of the HetNets by connecting the SCs into the BBU via high-speed optical fibers and deploying massive MIMO with the macro-cells [30]. In [3, 29] and [30], architectural frameworks for H-CRAN have been presented and depicted in Fig. 1.3. In the architectural framework of H-CRAN in Fig. 1.3, the SCs have been deployed beside the macro-cell or HPN in order to

[3] Network densification concept increases the number of BSs in a given area.

increase the network capacity and provide high QoS. The SCs communicate with the BBU pool via wireless or wired fronthaul, while the macro-cell communicates with the BBU via X2/S1 backhaul interface [5]. On the other hand, centralized and shared resources of C-RAN facilitate the decoupling of the user plane (UP) from the control plane (CP), i.e., SDN concept [27, 56]. In H-CRAN, the SCs can handle the user data, while the macro-cell manages the control signaling [5]. The splitting of control and user planes improves the network management and reducing the signaling on the fronthaul [3, 5]. In addition, in H-CRAN, the cross-tier interferences between the LPN and HPN can be managed by cloud computing-based cooperative RRM (CC-CRRM) [30]. Despite the benefits of H-CRAN on the QoS, it faces several challenges and issues in terms of optimal planning to avoid interference, energy consumption, and demanding a low latency and high capacity fronthaul and backhaul links [3, 29, 32].

1.2.1.2 Virtualized C-RAN (V-CRAN)

In 2012, the telecom operators observed several problems to launch new network services due to the heterogeneity, multiplicity, different proprietary, and short life cycles of the network hardware. Consequently, virtualization of infrastructure and computing resources has been studied widely, in the literature, thanks to several benefits of providing scalable, multi-tenancy, and cost-effective solutions compared to dedicated hardware at the cell site [60]. Virtualization concept decouples the network functions (NFs) from the underlying hardware and running them as virtual machines (VMs) or containers [61, 62]. In [34], virtualization of wireless networks has been considered with layer L3 in E-UTRAN protocol stack that connects the BSs into the core network (CN). In addition, ETSI in [34] discussed the virtualization of layer L2 in E-UTRAN operated on the BS. Currently, virtualization has been investigated for L1 [27, 35]. Consequently, in 5G networks, an E2E virtualization is considered to virtualize the CN, RAN, backhaul, spectrum, and air interface [3, 5, 27].

By adopting SDN and NFV concepts within C-RAN, the 5GC manages and orchestrates the network services and replaces the dedicated hardware solutions with virtualized software. In [37], an architectural framework for V-CRAN was illustrated that is composed of radio unit (RU), digital unit, virtual-BS (V-BS), and time-wavelength-division multiplexed passive optical network (TWDM-PON). The RU is responsible on transmitting and receiving the data. Meanwhile, the V-BS virtualizes the resources in the cloud digital unit. The digital unit, in turns, is responsible on the baseband processing and the functions of L2/L3. Furthermore, the TWDM-PON is a fronthaul solution to connect the RU with the digital unit in V-CRAN at low latency, cost, and energy consumption [37]. In [38], an architecture for SDN called SoftAir has been proposed that employs SDN and NFV concepts to provide a scalable and flexible C-RAN architecture. Adopting SDN and NFV

concepts within C-RAN introduces several benefits for C-RAN in terms of:

1. Optimizing the network resources by employing a dedicated resource to an application or service.
2. Simplifying the management of end-user devices by decoupling the UP and CP where these devices serve as packet-forwarding units, while the CP will be processed by an external controller.
3. Softwarization of the network by running the functionalities of the SDN controller and the BBU pools on NFs separated from the underlying hardware.
4. Plug-in and play by self and quick initialization and configuration of NFs.
5. Reducing the cost by softwarization and optimization of the network resources.

Despite the several benefits of V-CRAN, it still faces challenging issues. Several challenges for V-CRAN have been discussed in [3, 27] and [36] in terms of latency and heterogeneity of fronthaul links, unified protocol, security and privacy, resources management and allocation, VMs migration, placement and management of SDN controllers, and network isolation. Another challenging issue of the various C-RAN architectures is regarding the high round-trip delay to process the end-user devices data by the cloud/BBU pool [39]. This limitation of high delay of C-RAN can be tackled by adopting edge computing paradigms within the C-RAN architecture where part of the data can be processed or stored at the edge servers [40, 63]. Consequently, next subsection investigates the deployment of edge computing with the NG-RAN.

1.2.2 Edge Computing with 5G Architecture

For several years, cloud computing has been considered as the most predominant computing technology. However, with the arrival of IoT, an evolution toward edge computing was mandated to remedy the several constraints on the computing, storage, and networking resources. Due to the massive number of IoT devices and applications, distributed computing paradigms are mandated to support the new requirements of mMTC and URLLC UCs and to tackle the several cloud computing drawbacks. These drawbacks involve latency, congestion, real-time processing, and scalability [64]. Instead, edge computing brings the computational resources of the cloud near to the end-users. By employing edge computing within 5G networks, several benefits can be gained in terms of latency, location awareness, scalability, reliability, mobility, bandwidth, energy efficiency, and cost [64–66]. Furthermore, edge computing is vital for the applications that demand fast and real-time processing, ultralow latency, intelligent services, and smart solutions [67]. Consequently, edge computing promises to fulfill the various requirements for healthcare, smart factory, AR/VR, streaming and video gaming, autonomous vehicles, smart farming, and so on [68, 69]. To this end, this section investigates 5G architecture within fog computing and MEC.

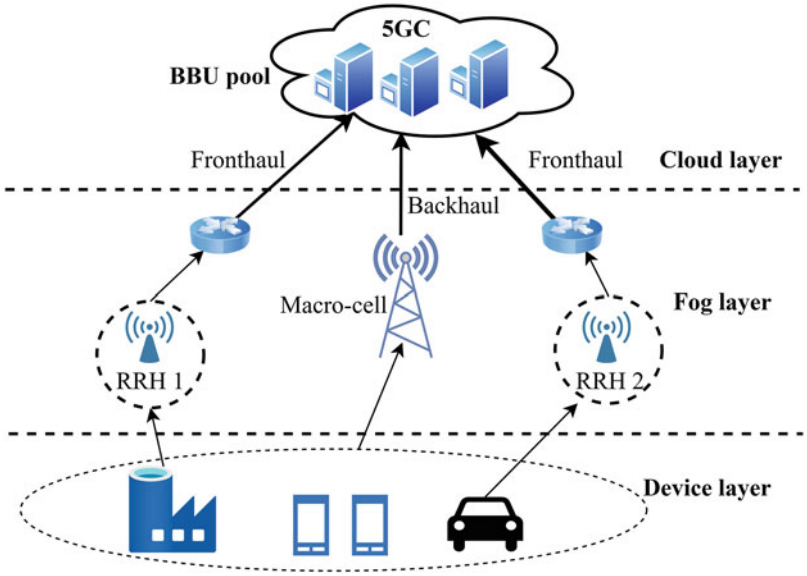


Fig. 1.4 Fog C-RAN 5G architecture

1.2.2.1 Fog-Based 5G Architecture Within C-RAN

Fog computing has been adopted within C-RAN architecture in order to enhance the end-users' mobility and facilitate the interconnection between all the network elements [65, 67]. Fog computing is highly geo-distributed non-standalone architecture proposed by Cisco [70] that represents an intermediate level between the end-user devices and the cloud. It is implemented on a dedicated gateway at the edge of the network in order to bring the storage, computing, control, and networking resources in the proximity of the end-users [64, 71].

As depicted in Fig. 1.4, fog C-RAN (F-CRAN) is composed of device, fog, and cloud layers. The end-users in the device layer offload their data to be processed on virtual machines on the fog devices. The fog tier represents an intermediate layer between the cloud and the device layers. It is composed of macro-cell, RRH, SCs, and networking devices such as access points (APs), gateways, and switches. This intermediate layer executes the delay-sensitive applications that cannot be handled by the end-user devices. The macro-cell, in fog layer, communicates with the cloud layer via a backhaul link, while the networking elements are communicating with the cloud through a fronthaul link[4] [39, 40]. The core network hosts the cloud servers that forms the BBU pools in order to execute the intensive and delay-tolerant tasks forwarded from the fog layer.

[4] The fronthaul and backhaul links are considered as Ethernet connection.

In order to alleviate the limitation of overwhelmed fronthaul links in C-RAN, F-CRAN permits to execute some functions, e.g., cooperative radio resource management (CRRM) and collaboration radio signal processing (CRSP), on the fog layer instead of the BBU pool. Furthermore, in order to support URLLC applications, F-CRAN adopts caching concept where some content can be downloaded directly from the RRH instead of demanding it from the BBU pool [39]. Fog computing solutions can be exploited not only by the mobile operators but also by the private networks. For this reason, OpenFog consortium [72] has been formed to address the various challenges within fog computing and to provide solutions for deploying private and public fog computing systems.

1.2.2.2 Cloud-Edge 5G RAN Based on O-RAN and 3GPP

Similar to the architectural framework proposed in [25] and adopted for C-RAN in [3] and [26], O-RAN alliance [42] decouples the NG-RAN into O-CU, O-DU, and O-RU. O-RAN provides an AI/ML-based architecture that aims to provide flexible, agile, automated, and optimized 5G network. As depicted in Fig. 1.5, O-RAN architecture is composed of O-Cloud, service management and orchestration (SMO), non-real-time RAN intelligent controller (Non-RT RIC), near-real-time RIC (Near-RT RIC), O-CU, O-DU, and O-RU. 3GPP in [24] defines the CU or O-CU as a logical node that operates the gNB functions and controls the operations of the DU or O-DU. Meanwhile, the Near-RT RIC hosts xApps and is responsible for the low latency processing, i.e., 10 ms to 1 s, as well as it monitors and controls the O-RAN nodes. In addition, O-RAN architecture defines the SMO entity that contains the Non-RT RIC. The role of the Non-RT RIC is to perform data analytics and AI/ML training that demands processing higher than 1 s [2].

O-RAN provides O1, A1, and E2 interfaces in order to enable intelligent processing and control of the traffic routing. The O1 interface collects data from O-CUs and O-DUs for the ML models in Non-RT RIC. In addition, A1 and E2 interfaces optimize the quality of experience (QoE) based on AI models [2, 42]. The open-fronthaul M-plane interface between the O-RU and the O-DU is defined by O-RAN alliance. Meanwhile, the interface F1 between the O-DU and O-CU and the interface NG between the O-CU and the 5GC are defined by 3GPP. However, the interface between the SMO and the 5GC has not been defined and out of the scope of O-RAN.

The entities of O-RAN architecture, such as O-DU, O-CU, near-RT RIC, and 5G user plane function (UPF), are considered as virtualized network functions (VNFs) operating on VMs or containers. In order to specify the distribution of these entities on the cloud hierarchy, O-RAN alliance proposed six deployment scenarios in [42] that are based on regional cloud and edge cloud. The edge cloud is placed in the proximity of the cell site and communicates with a regional cloud hosted in data centers [2]. Among the six deployment scenarios of O-RAN, O-RAN planned to deploy scenario B, as depicted in Fig. 1.5. In this deployment scenario, the O-DU

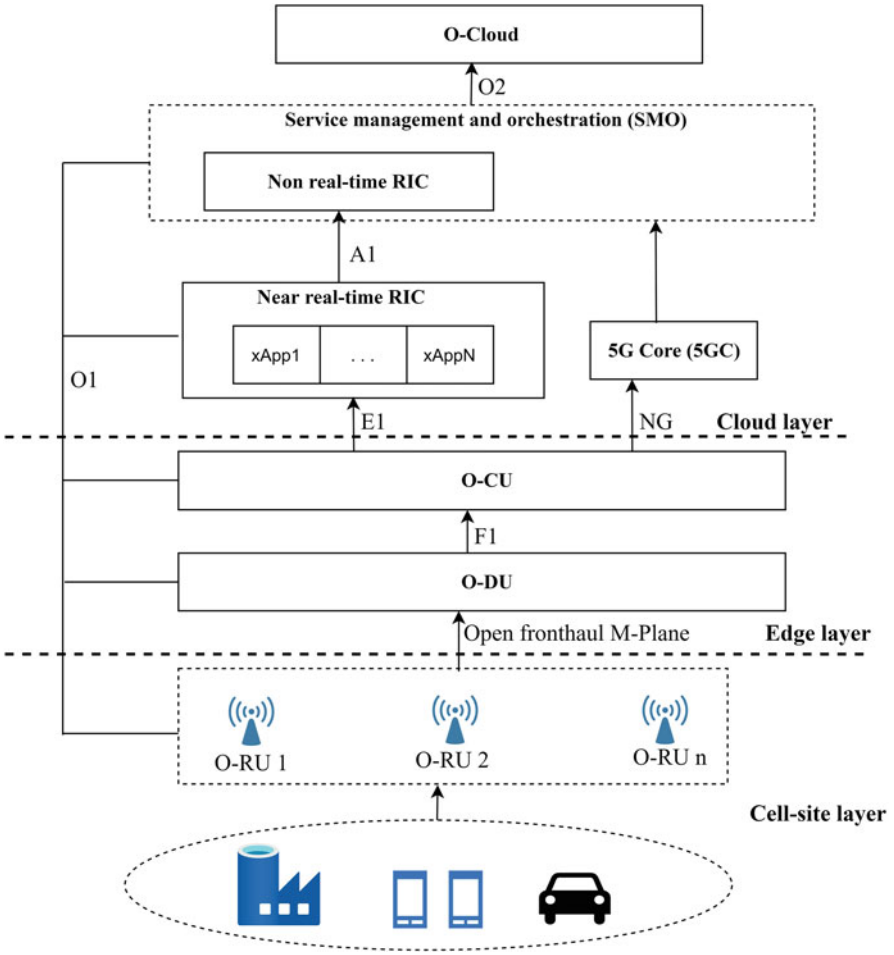


Fig. 1.5 O-RAN 5G architecture

and O-CU functionalities are processed at the edge, while the UPF and near-RT RIC are executed at the regional cloud.

In order to process the E-UTRAN stack on the DU and CU, 3GPP proposed eight split options as presented in Fig. 1.6 [24]. Among these split options, O-RAN alliance adopted the split option 7.2 [2, 43]. For Option 7, 3GPP separates the physical (PHY) layer into high-PHY and low-PHY for the downlink (DL) and uplink (UL). However, low layer split options have been postponed by 3GPP due to the requirements for bandwidth (BW) larger than 100 MHz and larger antennas to transmit the in-phase and quadrature (IQ) samples from the DU to the CU [1, 24].

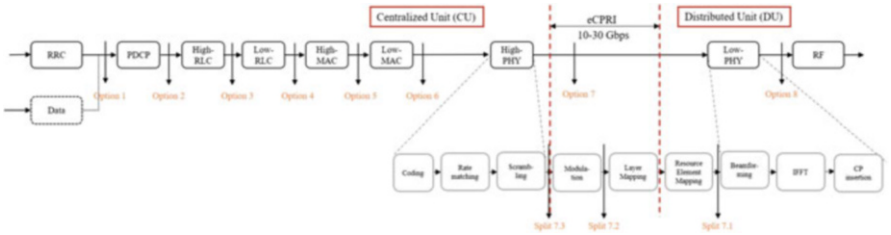


Fig. 1.6 The separation options of E-UTRAN stack protocol for DL

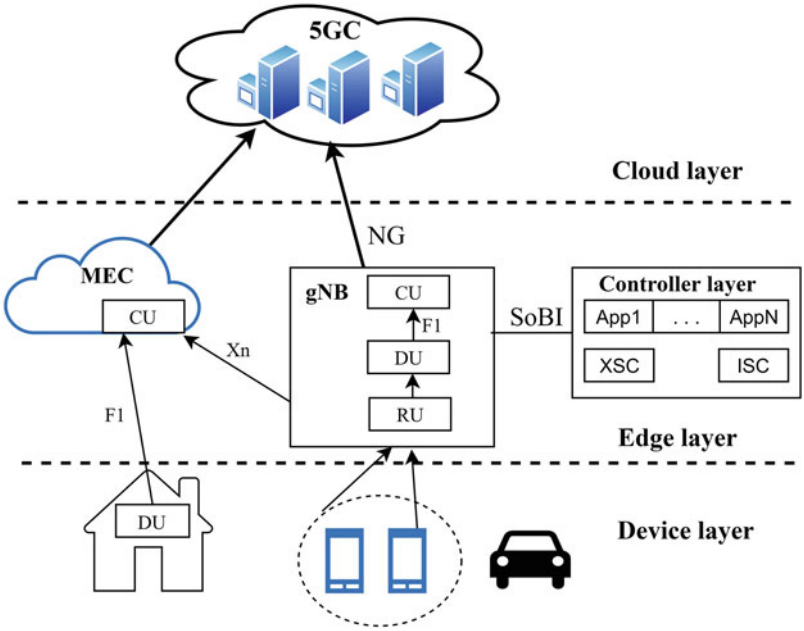


Fig. 1.7 5GPPP architecture for 5G with MEC

1.2.2.3 Multi-Access Edge Computing within 5G RAN Based on 5GPPP and 3GPP

Unlike fog computing servers that can be placed anywhere in the network, MEC is owned by the mobile operators and installed near to the BSs [73]. MEC is standardized by ETSI with the industry specification group (ISG) [68]. It aims to provide cloud capabilities and services at the edge servers. Consequently, adopting MEC enhances the QoS, energy efficiency, E2E delay, location awareness and handling of massive number of devices.

Figure 1.7 presents an example of the 5G architecture with MEC servers proposed by 5GPPP [45]. Similar to F-CRAN, the architecture of 5GPPP is composed of three tiers, i.e. end-user devices, edge, and cloud. The first-tier of the

5G RAN architecture represents the end-user devices such as automated cars, smart homes, UEs, and so on. The edge tier represents the NG-RAN that contains the DU, CU, controller, and MEC to provide services for the delay-critical applications. The cloud-tier involves the 5GC and executes the power-hungry and bulky tasks that demand high level of processing.

As depicted in Fig. 1.7, the DU and CU might be combined in one unit or separated to have the CU within the MEC and the DU close to the end-user devices.[5] In addition, the CU might be connected to a single or multiple DUs by F1 interface, while the CUs communicate via Xn interface. Moreover, the CU can be decoupled into CU-control plane (CU-CP) and CU-user plane (CU-UP) that facilitates the processing of the control and user planes on different locations.[6] Due to the separation of the CU into CU-CP and CU-UP, the NG interface to communicate with the 5GC is also separated into NG-C or N2 interface and NG-U or N3 interface. The 5G RAN interfaces are presented and explained in [24].

The end-users in Fig. 1.7 can be connected to their corresponding services by network slicing. In this regard, the 5GC eco-system adopts new services such as network slice selection function (NSSF) [46, 74]. The NSSF provides dedicated QoS and application functions (AFs) to interact with the network functions (NFs) [23, 45]. In order to enable network slicing and implement the RAN control functions as specific applications (App), 5GPPP provides a controller layer composed of cross-slice controller (XSC) and intra-slice controller (ISC). The control layer communicates with the RAN NFs via RAN controller agent (RCA) [45]. The RCA, in turns, is located at the CU and used to monitor and store RAN NF information. Accordingly, the controller layer interacts with the RCA via southbound interface (SoBI) in order to collect information regarding UE's power level, link quality, buffer state information, and so on [45].

Despite the several benefits and popularity of employing MEC within the 5G eco-system, it is still facing several challenges. These challenging issues involve the management of third-party applications, bringing the security and privacy drawbacks of the cloud into the MEC servers, and high maintenance cost in case of MEC failure [20, 75].

1.2.3 Comparison Among Public 5G Architectures

This section compares the various 5G architectures proposed by Telecom bodies and 5G standards for public networks. Table 1.3 compares the 5G architectural frameworks proposed by IBM, O-RAN, and 3GPP with 5GPPP in terms of deployment scenarios, benefits, and limitations. IBM provided the wireless cloud network (WCN) that has been re-explored by China mobile research institute

[5] The CU might be deployed as a dedicated hardware or as part of the MEC.

[6] The separation of CU-CP and CU-UP occurs via E1 interface.

Table 1.3 Comparison among the 5G architectures proposed by telecommunications bodies

Comparison	IBM	O-RAN alliance and 3GPP	5GPPP with 3GPP
Architecture	WCN or C-RAN	O-RAN	NG-RAN with MEC
Deployment scenario	Fully and partially centralized C-RAN. In addition, several architectures have been considered based on C-RAN such as H-CRAN, V-CRAN, and F-CRAN	6 deployment scenarios in [42]	3GPP in [21] provided 6 SA and NSA deployment scenarios. Meanwhile, 5GPPP provided an overall 5G architecture with different implementation based on various UCs
Architecture differences	C-RAN is composed of RU and BBU	O-RAN is divided into O-RU, O-DU and O-CU, Near-RT and Non-RT RIC, SMO, and O-cloud	MEC and a controller layer have been considered at the network edge
Benefits	Several architectural frameworks based on C-RAN such as H-CRAN, V-CRAN, and F-CRAN to improve energy and spectral efficiencies and minimize latency and CAPEX/OPEX	Several deployment options with AI/ML models to provide efficient radio resource management (RRM), low latency, and real-time processing	It addresses several UCs and vertical industry requirements within the NG-RAN architecture. It provides low latency and high processing services. Furthermore, it provides a controller layer to obtain information regarding UEs such as energy, buffer, and link quality
Limitations	Overwhelming the backhaul interface between the BBU and the CN, demanding fast fronthaul links and high round-trip delay to process the data of the end-user devices by the cloud	The integration of some functionalities with O-RAN is still in the early stage such as self-organized network (SON). In addition, E2E control plane services still cannot be created by O-RAN [76]	It provides different deployment scenarios for each UC. In addition, some functions and interfaces have not been defined by 5GPPP such as the northbound interface (NBI) and southbound interface (SoBI)

(CMRI) and so-called C-RAN. After that, several modifications and contributions have been proposed for C-RAN by Telecom bodies, e.g., Ericsson, ZTE, and Huawei. These contributions on C-RAN involve the deployment, advantages and challenges, architectures requirements, virtualization, and so on. As presented in Table 1.3, several deployment scenarios have been implemented based on C-RAN involving V-CRAN, H-CRAN, and F-CRAN. However, six deployment scenarios have been considered for O-RAN based on the placement of the architecture components whether to be processed at the edge or cloud. On the other hand, C-

RAN is composed mainly of RRHs and BBUs, while O-RAN and 3GPP classify the NG-RAN into RU, DU, and CU.

The various architectures and state of the art for C-RAN, including V-CRAN, H-CRAN, and F-CRAN, have been compared in [3–6]. These 5G architectures have been compared, therein, in terms of energy consumption, latency, security, resource allocation, reliability, spectrum and interference management, and CAPEX/OPEX. In [3], the latency and CAPEX/OPEX of C-RAN and H-CRAN were concluded to be high to very high, while they found to be medium to low for F-CRAN and V-CRAN. In addition, H-CRAN was found to have higher level of energy consumption than F-CRAN and V-CRAN. In order to enhance the CAPEX/OPEX, security, energy efficiency, and so on, several methods and strategies have been proposed for the 5G networks. Table 1.4 presents recent research works to tackle 5G network's challenges on energy efficiency, resource allocation, security, CAPEX/OPEX, and spectrum management.

1.3 Private 5G Networks

Several wireless technologies have been deployed within the enterprise's networks to provide solutions for their industrial applications. These wireless technologies can be categorized into 3GPP technologies, e.g., LTE for machines (LTE-M) and narrow band-IoT (NB-IoT), and non-3GPP technologies. Non-3GPP technologies can be classified into long-range, e.g., LoRa and SigFox, and short-range, e.g., Wi-Fi and Bluetooth low energy (BLE). However, these wireless technologies cannot fulfill the diverse demands of the industrial IoT applications. For instance, the long-range wireless technologies, e.g., LoRa, SigFox, and NB-IoT, prolong the battery lifetime but provide low data rate and high latency that might reach to 7 s [12]. On the other hand, the short-range wireless technologies have limitations in terms of battery lifetime, capacity, and coverage. In addition, LTE-M faces limitation in terms of cost and geo-localization accuracy. Above all, none of these wireless technologies provide solutions for the delay-sensitive IoT applications. For the aforementioned reasons, 5G technology is foreseen as a promising technology to enable the various demands of the industrial IoT applications. In 3GPP specification [94], the requirements of several industrial applications have been identified such as the reliability, latency, positioning accuracy, and battery lifetime.

Deploying a private 5G network in the enterprise's defined premises is fundamental to meet the demands of the various industrial applications. In addition, it offers a full control on the 5G RAN and core, and thereby several benefits can be gained in terms of coverage, deployment flexibility, availability, privacy, and security. Furthermore, the cost of accessing the licensed spectrum might be eliminated, thanks to the NR-unlicensed (NR-U) and the new unlicensed bands. The main benefits of the private 5G networks can be summarized as follows:

Table 1.4 Research works to tackle challenges for 5G architectures

Reference	Challenge	Methodology and outcomes
[77–80]	Resource allocation	In [77], a game theory algorithm has been proposed to maximize the utilization of the computational resources in C-RAN by distributing the resources of BBU pool efficiently. The simulation results showed that this algorithm improves the bottleneck management and the fulfillment level by 13%. In [78], a resource allocation algorithm for H-CRAN was proposed by using cooperative game approach and the constructive greedy approach. The proposed algorithm was found to outperform earlier research works in the literature in terms of average data rates, access rates, and throughput. In [79], joint resource allocation and admission control scheme was proposed to maximize the number of UEs in a network slice subject to a desired level of QoS. By comparing this algorithm to the existing baseline, the maximum number of UEs in a slice was increased effectively. In [80], a reinforcement learning algorithm has been developed to optimize the radio resource management (RRM) in O-RAN
[81–84]	Energy efficiency	In [81], an energy allocation and cooperation algorithm was proposed to improve the long-term energy efficiency in C-RAN. The problem has been formulated as a stochastic optimization problem and solved by using Lyapunov framework with the drift-plus-penalty method. In [82], a two-step energy efficiency algorithm for H-CRAN was proposed based on the underlaid and derived by using Dinkelbach, Lagrange multiplier, and alternating direction method of multiplier methods. This algorithm was found to outperform earlier research works in terms of energy efficiency. In [83], an energy efficient maximization algorithm was proposed for F-RAN. The maximization problem has been solved by using augmented Lagrangian with a heuristic solution to reduce the computational complexity. This algorithm was found to achieve a comparable energy efficiency result to AI algorithms. In [84], a joint optimization energy efficiency algorithm was proposed for O-RAN and solved as mixed-integer linear problem (MILP). The proposed algorithm was found to have low energy consumption at large network size
[85–87]	Spectrum management	In [85], an algorithm to optimize the spectrum pooling in C-RAN was addressed. In [86], a cooperated HetNet and C-RAN architecture was deployed, and a user-weighted probability-based algorithm was proposed in order to utilize the spectrum efficiently and manage the interference. In [87], spectrum pricing and allocation scheme was proposed for Fog-RAN by considering a Stackelberg game. The proposed scheme was found to improve the utilization of the network resources including spectrum

(continued)

Table 1.4 (continued)

Reference	Challenge	Methodology and outcomes
[6, 88] and [89]	Security	In [6], the security threats and attacks of the physical, control, and service planes in C-RAN have been comprehensively reviewed including challenges, methods, solutions, and research works. In [88], a network coding scheme was developed and implemented within H-CRAN in order to increase the transmission security by reducing the probability of data theft over multipath. In [89], a MACsec standard security protocol was integrated with O-RAN to secure the fronthaul data between RU and DU. The deployed security protocol was found to outperform the existing security protocols, such as IPsec, in terms of confidentiality, integrity, authenticity, and availability
[90–93]	CAPEX/ OPEX	In [90], a hybrid fronthaul approach by using optical fibers and free-space optics has been proposed in order to minimize C-RAN deployment costs. The proposed approach was analyzed in dense urban area scenario where it was found to reduce the CAPEX/OPEX effectively. In [91], a cost-effective solution has been proposed for the resource allocation between C-RAN and mobile cloud computing (MCC). In [92], a low complexity and successive elimination algorithm has been proposed to reduce the number of RRHs in H-CRAN under a QoS constraint. In [93], a cost-effective scheme was proposed for C-RAN in order to minimize the scheduling cost subject to resource availability, application requirements, deadlines, and load balancing

1. Design flexibility to boost the network coverage and capacity.
2. Optimize the QoS, reliability, latency, and so on.
3. High level of security and privacy because the data will be processed without interaction with the MNOs.
4. Enable customized and dedicated services.

1.3.1 5G Architecture for Private Networks

The 5G communities and alliance permit the deployment of local 5G networks with dedicated equipment to support the various applications of private enterprises and vertical industries. In this regard, 3GPP studies broad initiatives to enable the deployment of 5G private networks, including non-public network (NPN) management, NR-U, NR industrial IoT, and so on [47]. Moreover, the 5G alliance for connected industries and automation (ACIA) [48] and 5G-Smart [49] have considered two main architectural options for the industrial and private 5G networks.

The first option is the standalone NPN, while the second is NPN in conjunction with public network [48]. In the standalone NPN, all the enterprise functions related to user and control planes are processed on-premises without interaction with the MNOs or public networks. Nevertheless, in case some services demand an access to the public network, the NPN can access to the public network through a firewall [50]. An example of the standalone private architecture is the neutral-host network (NHN) proposed by muLTEfire alliance [51, 95]. In addition, Nokia [96] has proposed a standalone NPN defined as autonomous private wireless network.

The second deployment option for the private 5G networks is dependent on the agreement with the MNOs and can be classified into shared-RAN, shared-RAN and core control plane, and also NPN hosted by the mobile operators [49]. In shared-RAN and control plane, some functions, e.g., data flow, might be executed on-premises, while other functions, e.g., control plane, can be transferred to be processed by the public network [50]. An example of the shared-RAN architecture has been investigated in [48, 49, 52] and [54]. Furthermore, Nokia [96] provided a shared-RAN solution that is defined as a private wireless as a service (PWaaS). On the other hand, the mobile operators provide solutions to host the NPN as third-party services via network slicing, access point name (APN), and NFV. Among the various architectures of the private 5G networks, the NHN has the highest security and cost than the shared-RAN and hosted-NPN. In [48], a comparison between the various private 5G architectures is drawn in terms of security, availability, connectivity, and latency.

1.3.2 Spectrum Options for the Private 5G Networks

The UEs can communicate with the gNBs in the private 5G networks via licensed, shared, or unlicensed spectrum. The licensed spectrum can be dedicated for the private enterprises and third parties through agreement with the national regulator or the MNOs [47]. In this regard, the UK has licensed the spectrum 3.8–4.2 GHz for the private 5G networks, while Germany has dedicated a bandwidth of 100 MHz on bands 3.7–3.8 GHz for industry. Meanwhile, spectrum sharing enables multi-users to share a regulated spectrum at different times and places based on the national regulations and the commercial agreement. Three main frameworks have been considered for spectrum sharing. These frameworks are the citizens broadband radio service (CBRS), licensed shared access (LSA), and concurrent shared access [97]. The USA has dedicated the 3.5 GHz for CBRS, while Europe has considered the 2.3 GHz for LSA. On the other hand, the unlicensed spectrum has been investigated for LTE by considering three technologies as follows:

1. LTE-unlicensed (LTE-U) [98]: It has been standardized by LTE-U forum and considered in 3GPP releases 10, 11, and 12 [99]. It is based on carrier aggregation in order to boost the mobile network capacity. Accordingly, it enables the subscribers to access to the unlicensed 5 GHz band with an anchor carrier on

the licensed bands. This technology has been proposed for countries that do not mandate to perform the listen-before-talk (LBT) protocol such as the case in the USA, but, instead, it is based on the duty cycle.
2. Licensed-assisted access (LAA) [100]: Unlike LTE-U, LAA mandates to perform LBT, and thereby it is adopted in Europe and Japan. It has been standardized by the 3GPP in releases 13, 14, and 15 [101]. LAA and enhanced-LAA (eLAA) are based on carrier aggregation. Accordingly, LAA enables a supplementary UL/DL carrier on the unlicensed 5 GHz band with an anchor carrier on the licensed bands.
3. MuLTEfire [53]: Unlike LTE-U and LAA, MuLTEfire operates on standalone mode on the unlicensed sub-1, 2.4, and 5 GHz bands. It has been standardized by MuLTEfire alliance [53, 95]. Similar to LAA, it mandates to perform LBT protocol before accessing to the unlicensed channel.

Analogously with LTE-LAA and muLTEfire, anchored NR-U and standalone NR-U have been considered by 3GPP in releases 16 and 17 for 5G operating on unlicensed spectrum. NR-U in 3GPP release 16 operates on the unlicensed 2.4, 3.5, 5, and 6 GHz bands. These bands are extended in 3GPP release 17 for above-6 GHz bands such as 60 and 71 GHz [102, 103]. In comparison with LTE-LAA and muLTEfire, NR-U supports several deployment scenarios including carrier aggregation, dual connectivity, and standalone [101]. Moreover, NR-U reduces latency and enables URLLC applications by deploying several numerologies with mini-slots. Another advantage of NR-U is regarding the exposed node problem by permitting directional carrier sensing which increases the spectrum efficiency. Despite the various benefits of standalone NR-U, it still faces several challenging issues. These challenges were highlighted in [101] and [104].

1.4 5G Enabling Technologies

The SDN, NFV, network slicing, AI/ML, and MEC have been considered as key enabling technologies to automate, virtualize, softwarize, and optimize the 5G networks. This section investigates these enabling technologies and addresses the concept, standards, protocols, architectural frameworks, challenges, and recent research works.

1.4.1 Software-Defined Networking (SDN)

In the traditional cellular networks, data and control planes are tightly coupled on the same device that increases the complexity of managing and controlling the end-user devices. In contrast, the SDN technology enables the 5GC to configure and manage the control plane separately from the data plane [45]. According to open networking

foundation (ONF),[7] the main advantage of SDN is enabling a centralized and a fully controlled network by decoupling the data plane from the control plane of several nodes. Consequently, the end-user devices act as forwarding-packet devices where the control logic is moved from the devices onto a centralized layer [3].

The centralized control layer is defined as SDN controller that acts as the intelligence of the network. Consequently, the data plane delivers the packets by checking the flow tables that are controlled by the SDN controller [105]. The control plane, in turn, sends commands to manage the forwarding rules of the data plane [3, 106]. In [105], an example for the data plane rules and forwarding tables in SDN-based sensor networks has been presented. The data plane rules are configured by the control plane through southbound application platform interface (API) such as OpenFlow protocol. The control plane in the SDN-based sensor networks manages the sensors through forwarding data that is based on specified flow table and rules [105].

The SDN offers to the network administrators the ability to manage and control the data flow and to modify the characteristics of switching or routing devices from a centralized location by using software modules [105]. Several benefits can be gained by adopting SDN within the 5G RAN as follows [107]:

1. Decoupling the control and data planes.
2. Forwarding decisions flow-based instead of destination-based.
3. Control logic is moved to an external entity defined as SDN controller.
4. Several nodes can be controlled and managed by a SDN controller.
5. Enabling programmability where the network can be programmed via software modules running on the top of the SDN controller.

In the meanwhile, several benefits can be gained by adopting SDN within MEC. SDN within MEC has been found to improve the devices management, reduce the delay, and eliminate the bottlenecks not only in the data plane but also in the control plane. The MEC can host device management functions that facilitate the control of a subset of end-user devices. In this context, the adoption of SDN shows an improvement in terms of fault detection and device provisioning time, energy consumption, and CPU usage [108]. Furthermore, the flexibility and programmability features of SDN facilitate the deployment of MEC-related functions at the edge. In addition, the SDN multi-tenancy supports the creation of an open environment across multi-vendor platforms at the edge [109, 110].

1.4.1.1 SDN Architecture

The ONF has defined an architectural framework for SDN as presented in Fig. 1.8. The SDN architecture in Fig. 1.8 is composed of infrastructure layer, control layer, and application layer. The infrastructure or data plane layer is composed of the

[7] ONF is the consortium for SDN standardization.

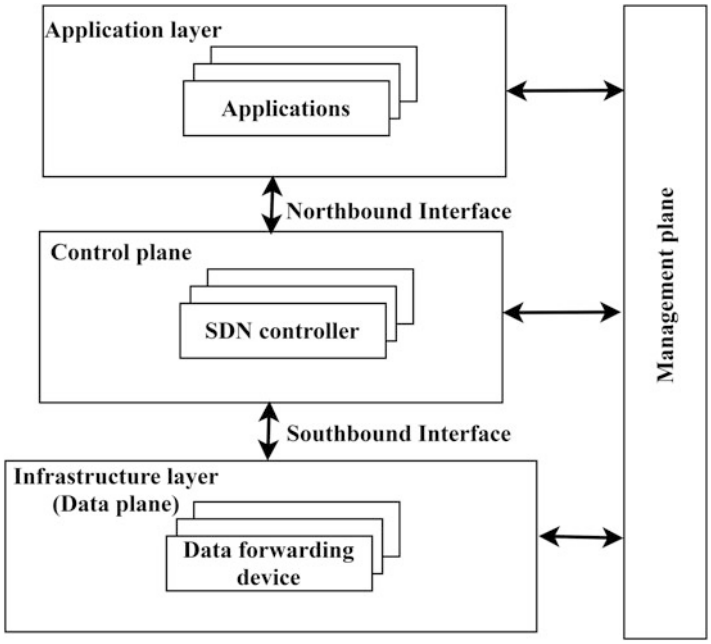


Fig. 1.8 Software-defined networking architecture

network devices or forwarding elements and virtual switches such as Open vSwitch. The switching devices in the infrastructure layer are responsible for collecting network status and sending it to the SDN controllers. The second function of these devices is to process the data packets based on rules provided by the SDN controller [3]. The infrastructure layer interacts with the control layer via southbound interfaces such as OpenFlow[8] protocol and forwarding and control element separation (ForCES) protocol. Meanwhile, the control layer contains several SDN controllers that are responsible for managing and monitoring all the virtual and physical network resources [3]. The control layer communicates with the application layer via northbound interface.[9] The application layer permits the service providers and network operators to manage the network to fulfill the various application's requirements such as traffic and access control, bandwidth, QoS, energy usage, and so on.

[8] OpenFlow is the most popular SDN protocol that defines a standard API to separate logically the controller and network devices. It offers an API between the SDN controller and the network devices to communicate over TCP/IP stack for any network topology.

[9] The northbound interface provides a network abstraction to the applications and the management systems. Various open-source projects have been proposed for the northbound APIs including RESTFul and Frenetic.

The SDN controller is a fundamental component of the SDN architecture. However, a centralized controller that manages all the devices of a network has a key drawback of a single point failure and scaling limitations. In order to alleviate these drawbacks, a distributed architecture with several SDN controller nodes has been proposed as an alternative to the centralized SDN controller architecture [107]. However, a distributed SDN poses design and implementation complexities. Therefore, the OpenFlow protocol has addressed these drawbacks by supporting multiple controller architecture [111]. In large networks where multiple controllers are deployed, an east-west communication interface is mandated to share network information between the several SDN controllers and to coordinate their decision-making processes [107].

The SDN concept can also be represented by three abstraction layers, i.e., forwarding, distribution, and specifications layers [107]. The forwarding layer hides the details of the underlying hardware and permits the business application to design a dedicated forwarding model. On the other hand, the distribution abstraction has two main functions: (1) configure the control commands on the forwarding devices and (2) collect a status information, e.g., network devices and links, from the forwarding layer [107]. Meanwhile, the specifications layer permits the business application to design a dedicated network framework that represents a simplified model of the real implementation [107].

1.4.1.2 Research Works on SDN

In order to guarantee an efficient and reliable communication in the dynamic and heterogeneous 5G RAN, various challenges have been raised regarding the implementation of routing and discovery protocols. Table 1.5 presents several algorithms and protocols that were proposed, in the literature, to tackle the various SDN challenges on topology discovery, failover duration, load balancing, controller placement, and security.

OpenFlow is the most popular SDN protocol that defines a standard API to separate logically the controller and the network devices. It offers an API between the SDN controller and the network devices to communicate over TCP/IP stack for any network topology [111]. Similar protocols that can be considered as an extension or dependent on OpenFlow were investigated in [125]. These southbound interface protocols include open virtual switch database management protocol (OVSDB), ForCES, protocol oblivious forwarding (POF), OpFlex, OpenState, and NetConf. All the SDN controllers are adopting the OpenFlow discovery protocol (OFDP) to facilitate the topology discovery by using link layer discovery protocol (LLDP) frames [126]. However, the OFDP has some limitations in terms of reliability and efficiency in heavy-loaded and large networks. The OFDP has been enhanced by OFDPv2, OFDPv2A, and OFDPv2B as discussed in [126] and research works cited therein.

Several algorithms have been proposed in the literature for SDN routing in order to reduce latency and save energy as in [127–131]. In [127], a ML algorithm was

Table 1.5 Research works to tackle SDN challenges

Reference	Challenge	Proposed algorithm and performance
[112] and [113]	Topology discovery	In [112], OFDPv2 algorithm has been proposed to reduce the number of control messages to be handled by the SDN controller and switches. In comparison with OFDP, it reduces significantly the induced CPU load on the SDN controller for the topology discovery by 40–45%. In [113], a hierarchical distributed topology discovery protocol was investigated for orchestrating multi-domain SDN. It showed a reduced provisioning delay while preserving confidential intra-domain information
[114–116]	Failover duration	In [114], a greedy incorporation and pre-partitioning strategies among controllers to minimize the failover duration have been proposed. The interruption time for the pre-partitioning mechanism was found to be significantly shorter than the greedy algorithm especially with a high number of orphan switches. In [115], fast failover mechanisms of both the control and data channels in the SDN-enabled data centers were presented. In [116], a fast and load-aware controller failover was proposed to resolve the SDN single point failure problem
[117–119]	Load balancing	In [117], a controller adaption and migration decision (CAMD) framework and a switch migration-based load balancing strategy have been proposed. This efficient switch migration technique improved the load balancing with reduced response time, and also it improved the end-users' QoS. In [118], an improved switch migration decision algorithm has been proposed for elephant flow or extremely large and continuous flow. By comparing it with CAMD, it is found to improve the SDN controller throughput, average response time, number of migration space, and the packet loss. In [119], a hierarchical two-phase algorithm was proposed. This algorithm integrated matching theory and coalitional games in order to achieve a near-optimal Nash equilibrium within tens of iterations. It reduced the total cost by 46% and achieved better load balancing among controllers in comparison to static assignment
[120–122]	SDN controller placement	In [120], a particle swarm optimization (PSO) and firefly population-based meta-heuristic algorithms have been investigated in order to achieve optimal placement of the SDN controllers. The results showed that the firefly algorithm performs better than PSO and random approach under multiple conditions. In [121], a framework for optimal controller placement with respect to different performance metrics was addressed by a heuristic approach. It is found to be less accurate but achieves rapid computation times. In [122], an algorithm based on hierarchical K-means was proposed to obtain more balanced controller placement
[123] and [124]	Security	In [123], a feedback-based scheduling scheme to determine the appropriate encryption algorithm based on free-to-add mechanism has been proposed. Employing the feedback-based IPsec was found to improve the throughput and latency for the networks traffic that fluctuates significantly. In [124], extended support and defensive algorithm were proposed to make SDN OpenFlow communication more secure

proposed in order to improve latency in smart city environments by predicting the routing paths. Furthermore, it presented and compared different routing protocols such as SDN-enabled connectivity-aware geographic-aware routing protocol (SCGRP). In [128], SDN-based multi-hop device-to-device (D2D) routing protocol has been proposed that offers high scalability and reduces the energy consumption for mobile nodes. Minimizing the energy consumption has been discussed in [129] by proposing an energy-efficient and secured routing protocol for software-defined Internet of vehicles. Meanwhile, Anjum et al. [130] proposed a real-time multipath transmission protocol to improve the best path selection in data center networks. On the other hand, secure and efficient topology discovery protocol (sOFTDP) has been proposed in [132] that offers a high level of security and reduces the topology discovery time in several orders of magnitude compared to the OFDP. Several discovery algorithms have been addressed in [133] such as fuzzy topology discovery protocol (FTDP) that improves the efficiency of software-defined wireless sensor network (SDWN).

1.4.2 Network Functions Virtualization (NFV)

The NFV standardization has been launched in 2012 by ETSI and ISG. It has been initiated by seven leading telecom network operators and became the most relevant standard for NFV. The NFV and the SDN technology are the key enablers for network softwarization. The NFV facilitates decoupling the NFs from dedicated hardware by implementing the NFs as software components running on VMs [134]. NFV promises to optimize the network resources, offer portability and design flexibility, reduce the CAPEX/OPEX, and fulfill the fast-varying user service requirements [135, 136]. In cellular networks, the virtualization is not only limited to the infrastructure and computing resources, but it includes also the spectrum, air interface, and multiple radio access virtualization [3, 60]. In this context, 3GPP deployed a virtual-EPC (vEPC) [137]. Moreover, ETSI NFV ISG discussed a zero-touch network concept in order to automate the operational tasks and management of 5G networks [45]. Providing a zero-touch orchestration, operations, and management (ZOOM) enables an E2E management solution across telecom operators [138]. This section highlights the NFV architecture provided by ETSI ISG and recent research works to overcome NFV challenges such as NFV orchestration and management.

1.4.2.1 NFV Architecture

The ETSI NFV ISG developed the requirements for NFV within telecom networks such as the requirements for hypervisor-based virtualization and security management [138, 139]. Furthermore, the ETSI NFV specification in [137] defines the NFV architectural framework, as presented in Fig. 1.9. The ETSI NFV architectural

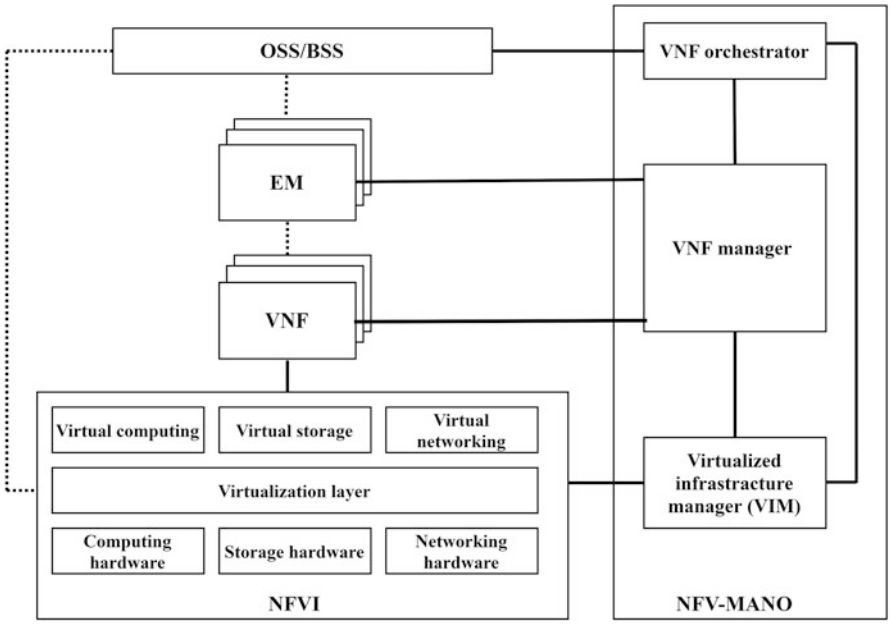


Fig. 1.9 Network functions virtualization architectural framework

framework has been integrated in 3GPP TS 28.500 [140] and considered as a reference for interoperability among all the telecom stakeholders. The NFV architecture in Fig. 1.9 is composed of four layers and can be summarized as follows [137, 141]:

1. VNF layer: It corresponds to the application plane and consists of two main parts as follows:
   (a) Virtual network functions (VNF): It is the virtualization of the NFs as the case for the network elements of 5GC, including mobility management entity (MME), serving gateway (SGW), and packet data network gateway (PGW).
   (b) Element management (EM) system: It is responsible for the management functionality, including fault, configuration, performance, and security, for one or several VNFs.
2. NFV infrastructure (NFVI) layer: It corresponds to the data plane and contains the hardware and software resources as follows:
   (a) Hardware resources: It provides processing, storage, and connectivity resources to the several VNFs through the virtualization layer or a hypervisor.
   (b) Virtualization layer or the hypervisor: It abstracts the hardware resources and decouples the VNF software from the underlying hardware. Therefore, the software can be deployed on different physical resources.

(c) Virtualization resources: It is virtual resources of computing, storage, and networking.

3. NFV management and orchestration (MANO) layer: It corresponds to the control plane and manages the physical and/or software resources for NFVI and VNFs. It includes:

   (a) Virtualized infrastructure manager (VIM): It performs the control operations and manages the interaction of VNF with the underlying resources such as allocation and management of infrastructure resources. Moreover, VIM guarantees the network connectivity and manages the NFVI resources, including allocation of VMs, computing, and storage resources.
   (b) NFV orchestrator (NFVO): It is responsible for the management of NFV infrastructure and software resources. It performs the orchestration and management of both VNFs and network services. NFVO is responsible for creating E2E services across multiple VNFs.
   (c) VNF managers: It manages the VNF instance life cycle including instantiation, update, scaling, and termination.

4. Operations support systems (OSS) and business support systems (BSS):

   (a) OSS is responsible for several management services involving network, fault, and configuration management.
   (b) BSS provides a customer product, order management, and so on.

1.4.2.2 Research Works on NFV

Various challenges have been addressed in [60] for an efficient integration of NFV within the 5G eco-system in terms of management, orchestration, resource optimization, and so on. Table 1.6 presents recent research works addressing these challenges. An efficient implementation of NFV within the 5G eco-system is required to ensure portability, interoperability, and consistent MANO of different applications and services to be executed through a unified interface. Therefore, standardizing of a unified northbound interface for NFV is considered as a challenging issue [139, 159]. In this regard, the 5G exchange (5GEx) project investigates proposing a unified European 5G infrastructure services [160, 161]. It involves cross-domain orchestration platforms and unification of NFV/SDN for multi-domain orchestration. In addition, the 5GPPP in [162] has analyzed several projects and platforms, e.g., OpenWhisky and Kubernetes, for 5G cloud services that are adopting the virtualization technology.

The design complexity of NFV within 5G networks is a challenging issue due to the diverse configurations of vertical industry's applications. This issue is envisaged to result in tremendous number of independent virtual networks operating on the same infrastructure which leads, in turn, to security threats. In this context, the ETSI NFV solutions (SOL) working group provides a set of specifications and open API

Table 1.6 Research works to tackle NFV challenges

Reference	Challenge	Proposed algorithm and performance
[142–144]	NFV orchestration	In [142], VNF orchestration algorithm based on variable neighboring search has been proposed to provide VNF as a service (VNFaaS) for any type of topologies. Furthermore, a demonstration composed of multiple OpenStack clusters showed an improvement in terms of QoS and a high level of availability to access the VNFaaS anytime and anywhere. In [143], an E2E integration system has been designed to orchestrate the heterogeneous resources of multi-vendors considering latency and availability constraints. It presented a full-automated E2E VNFaaS. In [144] a dynamic orchestration mechanism of SFC in hybrid NFV networks was proposed based on Markov renewal process (MRP) in order to reduce the computing time
[145–147]	VNF management	In [145], an algorithm to achieve an optimal reliability with minimum cost was proposed. The simulation results presented better performance by less computational resources and bandwidth consumptions in case of high number of services. In [146], a management method for the NFV MANO was proposed by employing the VNF group table. This method showed an efficient VNF management in terms of traffic statistics. In [147], a dynamic management algorithm for the VNF instances was proposed by deploying the model predictive control framework for multiple service chains. It presented better performance than threshold method in terms of data loss
[148–150]	VNF provisioning	In [148], a genetic algorithm to minimize the service blocking rate and the CPU usage was proposed. It is found to minimize the service blocking rate and increase the efficiency of the computation resources usage. In [149], an integer linear programming algorithm was deployed to solve the NP hard problem of the reliability-aware VNF instances provisioning in MEC. Experimental results showed the outperformance of the algorithm. In [150], deep Q-learning algorithms were employed to address the VNF provisioning problem in MEC, service chaining, and throughput maximization.
[151–153]	Service function chaining (SFC)	In [151], a genetic algorithm to find the efficient routing path under multi-conditions was proposed. It is found to reduce the computational complexity in comparison with the backpropagation algorithm that has been often used to find the optimal path in neural networks. In [152], the SFC composition and mapping problem have been formulated as a weighted graph matching and solved by using a Hungarian-based algorithm. In [153], a framework was proposed to increase the reliability of SFC while eliminating the consumption of resources

(continued)

Table 1.6 (continued)

Reference	Challenge	Proposed algorithm and performance
[154–156]	NFV placement	In [154], an online heuristic topology-aware placement algorithm has been proposed for VNF. It is found to improve the efficiency of resource allocation in cloud-based environment and minimize the computation time. In [155], a decision-tree algorithm was implemented to find the optimal placement of the OpenFlow containerized SDN-VNFs. It is found to detect the optimal placement in short and constant delay that outperforms the cbench approach. In [156], a model for dynamic reliability-aware service placement was proposed and formulated as an infinite horizon Markov decision process (MDP) in order to minimize the cost of the cloud server placement
[157] and [158]	VNF life cycle management	In [157], an online VNF scale-out/scale-in and dynamic live-migration scheduling algorithm was proposed to assign the VNF to an actual cloud or MEC hypervisor. The proposed algorithm is found to utilize the MEC and cloud resources efficiently and enable a high number of URLLC applications. In [158], a framework was proposed to combine the softwarized networks by SDN and NFV with ML-based management

to manage and orchestrate all the different management aspects of NFV system including configuration, life cycle, performance, and fault management [139].

1.4.3 Network Slicing

The decoupling of data and control planes via SDN and the separation of the NFs from the underlying hardware via NFV permit the 5G networks to offer flexible, programmable, and self-contained E2E virtual network services [163]. Network slicing is defined by 3GPP as differentiated treatment for each tenant requirement based on the service level agreement (SLA) with the MNOs [164]. From UE perspective, network slicing represents the devices with similar design requirements gathered into a slice. On the other hand, network slicing from infrastructure perspective can be defined as partitioning the physical resources of the network into a set of logically segregated virtual networks [165]. Accordingly, the slice can be considered as a set of virtualized or non-virtualized NFs that contains the session management, application, user plane functions, and so on [166, 167]. Moreover, network slicing should be capable to provide resources for the various UCs of URLLC, eMBB, and mMTC. Several benefits can be gained by adopting network slicing concept within the 5G networks as follows [168]:

1. Multi-tenancy: The same physical infrastructure is shared between several operators.
2. Service differentiation: Each slice fulfills the desired QoS for a network service such as delay, data rate, and packet loss.
3. On-demand by facilitating the creation, modification, or removal of a slice in real time.
4. Resource optimization, reducing the CAPEX/OPEX, and energy savings.

This section highlights recent research works on network slicing architecture and presents the conceptual architecture of network slicing. Meanwhile, several challenging issues and recent research works on network slicing are addressed in Sect. 1.4.4 and Table 1.8.

1.4.3.1 Network Slicing Architecture

Network slicing within 5G technology has been discussed in several standards and research works. Table 1.7 presents standards and research works that discussed network slicing architectural framework and the modifications on RAN and the core network (CN). Unlike network slicing for LTE that supports only MBB services, network slicing in 5G networks provides services for vertical industries including the UCs of eMBB, URLLC, and mMTC. 3GPP in [164] defined the fundamental regulations to enable network slicing within the NG-RAN, including RAN awareness of slices, resource management and isolation between slices, selection of CN entity, and so on. Meanwhile, ITU-T in [179] specified the framework for the non-radio aspects of 4G LTE and 5G technologies including network slicing.

The architectural framework proposed by ITU-T [179] for network slicing is presented in Fig. 1.10. It considers three layers, i.e., service instance, network slice instance (NSI), and resource layers. The service instance layer represents the application or the end-user services provided by tenants or network services providers [165]. Meanwhile, the NSI provides the dedicated features for the service instance, and it could be fully static, partially dynamic, or constructed on-demand [180]. One or several instance services can be served by one or several NSI. An instance service can be identified and chained to a corresponding NSI by a network slice blueprint. The network slice blueprint provides a full description of the configuration and specifications of the instance service as well as it is used by the MNO to instantiate and manage the NSI.

1.4.4 Artificial Intelligence/Machine Learning (AI/ML)

AI/ML is vital for 5G networks and beyond because it provides efficient tools for network planning, diagnostics, and optimization [181]. The heterogeneity of 5G networks with the diverse and high number of services poses challenges on the

Table 1.7 Research works on network slicing architecture

Document type	Title	Description	References
Standard	3GPP TS38.300: NR and NG-RAN overall description-stage 2 3GPP TS23.501: System architecture for the 5G system-Stage 2	Describe the general principles, requirements, functions, and configurations for network slicing architecture in NG-RAN and CN	[164] and [169]
	ITU-T Y.3112 series Y: Global information infrastructure, Internet protocol aspects, next-generation networks, IoT, and smart cities	Present a high level of requirements and CN functions for network slicing	[170]
Articles	Adaptive Network Slicing in Multi-Tenant 5G IoT Networks	Describe the control and management network slicing architectural frameworks for 5G IoT	[171]
	Network slicing in 5G mobile communication: Architecture, prototype modelling, and challenges	Describe the network slicing architecture considering RAN, CN, and pairing functions slices	[172]
	An E2E network slicing framework for 5G wireless communication systems	Present a global E2E framework for network slicing in 5G network	[173]
	On radio access network slicing from a radio resource management perspective	Describe the network slicing deployment in NG-RAN	[174]
	Impact of network slicing on 5G radio access networks		[175]
	A new network slicing framework for multi-tenant heterogeneous cloud radio access networks		[176]
	Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges	Present the concepts of SDN and NFV within network slicing	[177]
Books	5G NR architecture technology implementation and operation	In chapter 1, network slicing conceptual architecture and the impact on RAN for URLLC, mMTC, and eMBB have been highlighted	[165]
	5G system design architectural and functional considerations and long-term research	In chapter 8, an implementation of network slicing in CN and the slicing impact on RAN were discussed	[178]

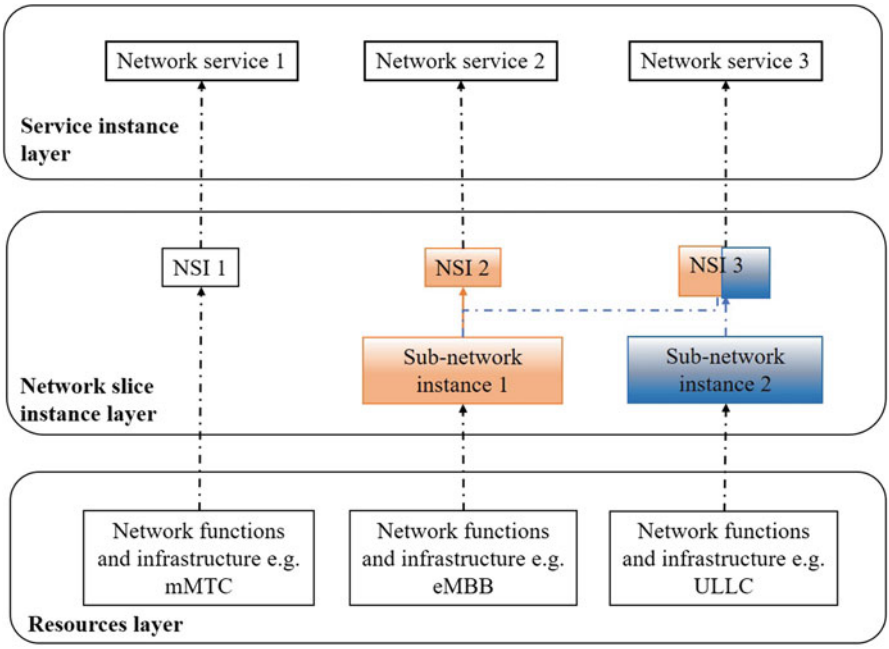


Fig. 1.10 Network slicing conceptual architecture

network planning including element placement problem and dimensioning of C-RAN clusters. Therefore, AI/ML is mandated to automate the network planning to find the optimal placement of network elements, reduce CAPEX/OPEX, and maximize BSs coverage. Meanwhile, AI/ML provides efficient tools for network diagnostics in order to forecast, detect, and alert network problems [181]. Consequently, AI/ML methods with network diagnostic overcome challenges regarding UEs localization and forecasting of network characteristics and events to improve the QoS and security. On the other hand, AI/ML with network optimization and control provides promising solutions for challenging issues in RAN slicing, radio resource provisioning, virtualization, applications, and so on [181]. Furthermore, AI/ML is crucial for 5G networks to provide the ability to analyze the data "Tsunami" in short time, automate the network processes without being explicitly programmed, and predict the network behavior with high level of accuracy [182]. At the sequel, AI/ML mechanisms bring several benefits for 5G networks in terms of network planning and diagnostics, RAN congestion, on-demand and adaptive network configuration, service creation and orchestration, fault detection, security, handover and mobility management, user experience enhancement, and dynamic policy adjustment [183].

Several AI/ML mechanisms have been proposed, in the literature, in order to optimize the performance of 5G networks. These AI/ML techniques can be classified into [181, 184, 185]:

1. Supervised learning: It uses a training data and comprises input and desired outputs. It includes support vector machine, Bayesian learning, regression model, K-nearest neighbor, and gradient boosting decision tree.
2. Unsupervised learning: It is based on unstructured or unlabeled data and used for data clustering, dimensionality reduction, and density estimation. It includes K-means clustering, spectral clustering, replicator neural network, and principal and independent component analysis.
3. Reinforcement learning (RL): It learns by interacting with the environment without human intervention in order to obtain the optimal policy. RL approaches include Q-learning, Markov decision process (MDP), and multi-armed bandits.
4. Deep learning (DL) or semi-supervised: It is a highly advanced unsupervised learning model that is based on neural networks.
5. Deep reinforcement learning (DRL): It is a combination of deep and reinforcement learning.

In [181] and [185–187], the role of AI/ML mechanisms in 5G networks and research works to overcome diverse 5G challenges have been addressed. 5GPPP in [181] has investigated AI/ML methods to tackle optimization problems in 5G and beyond involving network planning, diagnostics, optimization, and control. Meanwhile, the ITU focus group on ML for 5G (FG-ML5G) [188] has classified the UCs for ML into five categories as follows:

1. Network slicing and other network services-related use cases including cognitive heterogeneous networks and ML-based self-organized networks (SON).
2. User plane-related UCs such as traffic classification and emergency services based on ML.
3. Application-related UCs for ML data-driven architecture at the edge.
4. Signaling or management-related UCs involving ML-based mobility pattern prediction, ML-based QoE optimization, ML-based E2E network management, and so on.
5. Security-related UCs.

1.4.4.1 Research Works on Network Slicing with AI/ML Strategies

In this section, AI/ML research works are addressed to remedy challenges facing network slicing in 5G RAN. Various challenging issues were highlighted in [188] for network slicing such as optimization of RAN slicing, the prediction of allocation trajectory per network slice and cell, optimal radio resource allocation per slice, and so on. Table 1.8 presents recent AI/ML research works that were proposed in the literature to alleviate several network slicing challenges in terms of resource allocation and management. In [189], a Q-Learning algorithm to maximize the

Table 1.8 Research works adopting AI/ML for network slicing

Ref.	ML algorithm	Challenge	Outcome
[189]	Q-Learning	Slice admission policy optimization	Optimize the network infrastructure revenue while guarantee the QoS to third-party tenants. An adaptive admission algorithm based on Q-Learning framework was proposed to achieve high substantial gains for the network infrastructure provider compared to a smart heuristic method
[190]	Convolutional neuron network	Slice admission control and cross-slice resource allocation prediction and optimization	Provide an optimal balance between resource over-provisioning and insufficient resource allocation to minimize the economic cost for the infrastructure provider. A deep neural network structure was proposed to reach an efficient capacity forecast rather than the legacy mobile traffic prediction
[191]	Deep reinforcement learning	Maximization of the slice number	Deploy RAN resources allocation independent of the fluctuant number of RAN slices A DRL method was proposed to optimize the RAN resource allocation
[192]	Deep Q-learning	E2E QoS optimization	Dynamic resource allocation to maximize the number of access users. An E2E network slicing resource allocation algorithm based on deep Q-learning has been proposed to optimize the resource allocation on RAN and core sides
[193]	Deep reinforcement learning	Optimization of multi-provider physical infrastructures	The proposed RL algorithm optimizes the slice request partitioning for a slice deployed across multiple substrate networks with outperformance in terms of delay
[194]	Reinforcement learning	Slice admission optimization on RAN infrastructure	RL-based solution to maximize the profit of the provider infrastructure with different QoS constraints. A RL-based slice admission policy was proposed to maximize the provider infrastructure revenues by minimizing the rejected or degraded services
[195]	LSTM & A3C algorithm	RAN slicing optimization	Provide performance isolation among RAN slices and optimize the RAN resources. A collaborative learning framework using DL for large timescale networks in conjunction with RL for small timescale networks has been proposed and solved by asynchronous advantage actor critic (A3C) algorithm
[196]	Deep neural network	RAN slicing optimization	A deep Q-network was proposed for the dynamic radio resource slicing allocation. This algorithm improves significantly the radio resource utilization
[197]	Deep neural network	Dynamic SFC resource allocation	Optimization of SFC dynamic resource allocation for multilayer core network. RL algorithm was proposed to enhance the performance of dynamic SFC resources allocation

profits of the infrastructure provider and to satisfy the tenant-user services was proposed. The simulation results in [189] achieving a near-optimal performance for the admission policy. On the other hand, Bega et al. [190] presented an analytical tool based on the convolutional neural network in order to manage the decisions of the slice orchestrator. In [191], a DRL algorithm was proposed for RAN slicing in order to achieve the optimal resource allocation regardless of the slice number. The simulation results in [191] were found to maximize the number of slices that fulfills the requirements of each application. Meanwhile, in [196], a deep Q-learning algorithm was proposed for the RAN slicing in order to optimize the resource reservation. In [195], a DL algorithm with a RL method was proposed for RAN slicing to predict the traffic volume for resource allocation and scheduling.

1.4.5 Multi-Access Edge Computing (MEC)

With the arrival of IoT and the massive surge of applications, an evolution toward edge computing was mandated to alleviate the various challenges of cloud computing involving ultralow latency, network congestion, high QoE, energy savings, and scalability [64, 67]. Edge computing brings the computational resources in the proximity of end-users and can be located anywhere in the RAN to connect the end-users with the core network [165]. For instance, in [198], edge computing servers were considered to be inside a femtocell or the device itself. Consequently, several benefits can be gained by edge computing such as low latency, real-time processing, location awareness, high scalability, availability and reliability, high security and data privacy, and saving energy and cost [64–66]. Under the umbrella of edge computing, several computing paradigms were proposed such as mobile cloud computing (MCC), cloudlet, fog computing, and mobile edge or multi-access edge computing (MEC) [64]. In this section, we address MEC because it is the standardized technology for mobile networks.

MEC transfers the cloud capabilities, including computing, storage, and networking resources, into the network edge, e.g., BS or eNB/gNB [64, 199]. It has been standardized by ETSI ISG and European 5G infrastructure public-private partnership. In this context, ETSI ISG in [200] proposed several white papers for MEC principles and frameworks including several UCs, APIs requirements, implementation and tests, and architectural frameworks. In the early versions of MEC standardization by ETSI, MEC has been denoted as mobile edge computing which has been defined, later, as multi-access edge computing.

1.4.5.1 MEC Architecture

MEC architectural framework has been explored, in the literature, by several standards, research works, and books. Table 1.9 presents standards and research works that investigated the implementation of MEC within the 5G eco-system. The

Table 1.9 Research works on MEC architectural framework

Document type	Title	Description	References
Standard	ETSI GS MEC 003: Multi-access edge computing (MEC); framework and reference architecture ETSI white paper No. 28: MEC in 5G networks ETSI GR MEC 031: Multi-access edge computing (MEC) MEC 5G integration	ETSI is MEC standardization home. It investigates several issues, UCs, and architectural frameworks to realize the implementation of MEC with 5G network	[201–203]
	5GPPP: View on 5G architecture	An architectural framework for the implementation of MEC within the 5G network has been presented	[45]
Articles	How Can Edge Computing Benefit From Software-Defined Networking: A Survey, Use Cases, and Future Directions	Compare and investigate several paradigms for edge computing	[20]
	Mobile Edge Computing: A Survey	Provides a comprehensive study on MEC advantages, architectural framework, UCs and applications, issues, and research works	[204]
	A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Arts	Present a comprehensive survey on the integration of MEC with several technologies including 5G network and beyond	[205]
Books	5G NR architecture technology implementation and operation	In chapter 1, the deployment scenarios and architectural frameworks of MEC have been addressed	[165]
	Multi-access edge computing in action.	MEC motivations, dimensions and applications, architectural frameworks, deployment scenarios, and market have been investigated	[206]

main standard for MEC is ETSI ISG which is the home for MEC standardization. As presented in Table 1.9, ETSI has published several white papers that investigate the issues to implement and realize MEC with 5G. MEC reference architecture has been defined by ETSI in [201] as depicted in Fig. 1.11. MEC reference architecture in Fig. 1.11 can be categorized into three main entities as follows [165, 201, 206]:

1. Network-level entity: ETSI defined three main types of networks involving 3GPP, local, and external networks. 3GPP network is standardized by 3GPP such as 5G network. Meanwhile, the local network could be any local access network

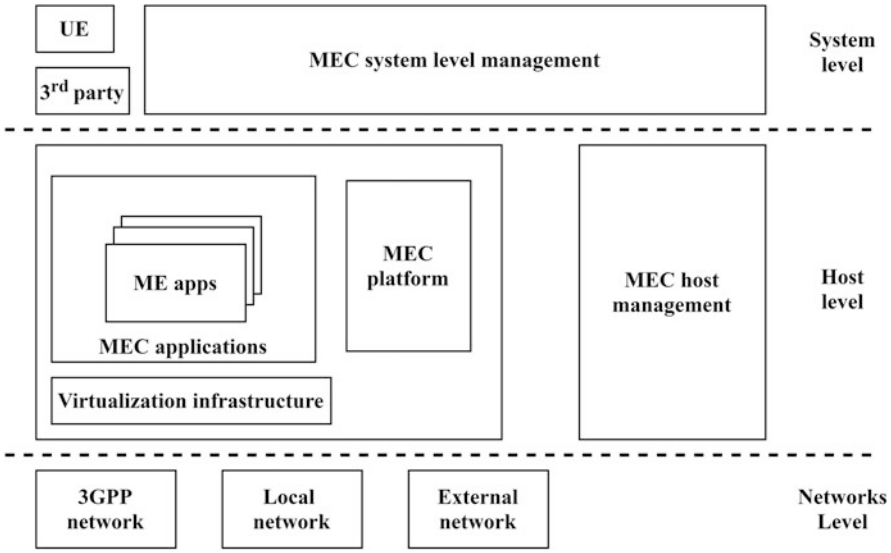


Fig. 1.11 MEC conceptual framework [201]

operated by wireless technologies such as Wi-Fi and Internet. On the other hand, the external network enables the third-party and industry networks such as private networks.

2. Host-level entity: It hosts the mobile edge (ME) components such as applications, platforms, virtualization infrastructure, and ME platform manager (MEPM). The virtualization infrastructure includes the computing, storage, and networking entities as well as hypervisor and OpenFlow switch protocol. ME applications operate as VMs on the virtualization infrastructure and interact with the MEC platform. The ME platform, in turns, provides the required functionalities and policies to execute the ME applications and access into MEC services. It provides ME service, service registry, and traffic rule control. On the other hand, MEPM is the main management entity at the MEC host level. It manages platform element and ME application life cycle, rules, and requirements. MEPM has been investigated by ETSI in [207] and [208].
3. System-level entity: It involves the user application life cycle management proxy, operations support system (OSS), and mobile edge orchestrator. The user application life cycle proxy is responsible on some application services such as instantiation and termination. In addition, it provides a security gateway between the third-party applications and OSS. Meanwhile, the OSS receives the requests of UE and third-party applications. It verifies and grants resources for these requests and then forward them into the MEC orchestrator. The MEC orchestrator has a global view on the available MEC resources, services, and topology. Furthermore, it manages the MEC system, functionalities, services, and resources as well as third-party and UE applications. In addition, it verifies and achieves the requested resources and QoE for the applications.

MEC entities communicate via reference points categorized into ME platform functionality (Mp), management reference points (Mm), and interfacing with external entities (Mx). These three reference points are detailed and explained in [165] and [201].

1.5 5G Development and Experimentations

Mobile operators, telecom enterprises, and Academia are collaborating on developing simulation tools to validate and analyze the cellular networks with or without an actual implementation of hardware. These simulation tools are targeting the RAN, core, edge/cloud, and application layer. This section presents several 5G RAN and core simulation tools and projects for the edge/cloud and application layer.

1.5.1 5G Development Platforms for R&D

Various 5G and LTE platforms, e.g., srsLTE and Radisys, for RAN and core have been investigated in depth in [209]. Besides these platforms in [209], additional simulation tools for 5G networks are compared in Table 1.10. As illustrated in Table 1.10, the LTE and 5G simulators for RAN involve three layers as follows [215]:

1. Link-level simulator (LLS): It analyzes the performance of PHY layer for UL and DL point-to-point wireless link between a BS and UE. It involves the channel coding, waveform, and modulation.
2. System-level simulator (SLS): It evaluates the performance of MAC layer such as scheduling and radio resource management.
3. Network simulator (NS): It analyzes the performance of high layers with the core network.

Table 1.10 compares LTE and 5G simulators including MATLAB, NS-3, Vienna simulator, OpenAirInterface (OAI), and 5G K-simulator. MATLAB provides LLS solutions including 5G NR DL waveforms, beam sweeping, and so on [210]. On the other hand, NS-3 LTE EPC network simulator (LENA) module provides a NS with an abstraction of the physical layer [211]. Meanwhile, Vienna simulator offers a multi-use MIMO, beamforming, and mm-Wave solutions for the 5G technology [212]. Moreover, OAI simulator provides solutions for all layers of E-UTRAN stack protocol as well as the core functionalities [213]. Similar to OAI, the 5G K-simulator develops LLS, SLS, and NS based on the 3GPP Release 15 [214]. Among the compared simulators in Table 1.10, NS-3, MATLAB, and OAI gain high credibility in the R&D community. The credibility of these simulators can be measured by the number of published articles each year by using these simulators. For this reason of high credibility, in this section, the solutions provided by OAI, NS-3, and MATLAB are investigated and compared.

Table 1.10 Comparison between 5G simulators

Simulator	NR (5G)	NR-U	LLS	NS	SLS	Credi-bility	Open source	Lang-uage	Compl-exity	User manual
MATLAB [210]	√		√			High		Fortran	Simple	√
NS3 [211]	√	√	Abstr-action	√		High	√	C++	High	√
Vienna [212]	√		√		√	Mid	Acad-emia	Fortran	Mid	√
OAI [213]	√		√	√	√	High	√	C	High	√
5G K-Sim [214]	√		√	√	√	Low	√	C++	Mid	√

1.5.1.1 OpenAirInterface (OAI)

It is an open-source software developed by EURECOM and OAI alliance, involving Orange, Nokia Bell Labs, Qualcomm, TCL, and others. It provides RAN and core functionalities for LTE and 5G systems including IoT technologies such as NB-IoT. It supports the features of LTE, LTE-Advanced, and LTE-Advanced-Pro as well as 3GPP releases 15, 16-onward for the 5G RAN and core [216]. Furthermore, it is compatible with several software-defined radios (SDRs) and commercial-of-the-shelf (COTS) cellular phones. OAI is licensed under OAI public license 1.1 for the RAN software and Apache V2.0 for the core software.

Since 2018, OAI has been developing the Openair5G software for 5G RAN to operate on the EPC considering NSA deployment. Currently, OAI started the development of the 5G SA option to be accomplished in 2022 onward. The deployment of the 5G RAN software on OAI considers the split option 7.2 [216]. In Openair5G software, the DU processes L1 and L2 that communicates via the functional application platform interface (FAPI)[10] protocol. Moreover, the CU processes L3 and communicates with the DU via F1 interface standardized by 3GPP. On the other hand, OAI is developing the 5G core functionalities for the 5G SA architecture within the software openairCN to be completed in 2022 onward. The 5GC is decoupled into UP and CP. In the first phase of the 5GC deployment, OAI considers the mandatory 5G core functions such as user-plane function (UPF), access and mobility management function (AMF), and session management function (SMF) [218].

Operating a real-time E2E network by using OAI demands a powerful PC for the gNB that can be interfaced to several types of SDR such as universal software radio peripheral (USRP) X310 and N310. Meanwhile, OAI supports the deployment of LTE and 5G software without actual SDRs by providing four simulators as follows [216]:

1. L1 simulator: It analyzes E-UTRAN stack protocol for multiple UE considering IF4.5 interface between the UE and the eNB/gNB.
2. Basic simulator or RFsimulator: It evaluates the E-UTRAN stack protocol for one or multiple UEs with emulation of the RU.
3. Uplink and downlink simulators (ulsim/dlsim): They analyze the physical layer of the UE and the eNB/gNB.
4. L2 simulator that simulates the upper layers for the UE and the eNB/gNB where the MAC and PHY layers communicates via nFAPI.[11]

Despite the possibility of implementing a 5G E2E network via OAI, it faces several limitations. For instance, OAI does not provide systematic documentations for the installation and configuration of OAI software with the various SDR platforms. Moreover, it provides few documentations to clarify the RAN threads

[10] FAPI protocol is standardized by the small cell forum in [217].

[11] nFAPI protocol is standardized by small cell forum in [219].

and software. In addition, unlike NS-3 that supports several RATs, OAI supports only 3GPP technologies for LTE and 5G.

1.5.1.2 Network Simulator-3 (NS-3)

NS-3 [220] is a discrete-event network simulator approved by Academia for various RATs such as Wi-Fi, WiMAX, and LTE. It provides several modules that are developed on C++ under the license GNU GPLv2 [220]. LTE/EPC network simulator (LENA) [211] is one of the modules provided on NS-3 to analyze LTE networks. It supports several features including DL&UL schedulers, radio resource management algorithms, intercell interference coordination, mobility management, and so on [221]. The E2E architecture for LENA project is composed of UEs, eNBs, and EPC that contains the SGW and PGW functions as well as the UP and CP interfaces [221].

NS-3 provides a 5G NR module that has been built on the top of the mm-wave and LENA modules. The NR module provides several features for the PHY and MAC layers including flexible frame structure of several numerologies and mini-slots, dynamic time division duplex (TDD), mm-wave with beam management-related operations, channel coding models, flexible MAC schedulers, and multiple bandwidth parts (BWP) [211, 222]. Furthermore, NS3 provides a NR-U module for the mm-wave bands.

Even though NS-3 provides several RAT modules and does not consider constraints on the PC requirements, the NR module on NS-3 has several limitations. For instance, it provides PHY layer abstraction based on look-up tables, employs LTE upper layers, and operates on EPC and mm-wave bands only. Nevertheless, national instrument (NI) has implemented a LTE project that employs the upper layer of LENA module with a physical layer operates on LabVIEW [223, 224]. However, this implementation demands LabVIEW's communication toolbox, two PCs operating NI Linux real-time (RT) controllers, and two USRPs 2974.

1.5.1.3 MATLAB 5G Toolbox

MATLAB provides a 5G toolbox to configure, measure, and simulate the LLS. It supports physical shared and control channels and other functionalities such as MAC scheduling strategies [210]. According to [225], several SDRs can be supported by MATLAB 5G toolbox including Xilinx Zynq and FPGA radios and USRP N, X, and E series. However, MATLAB does not provide full functionalities for L2 and L3 nor functionalities for the 5GC. Consequently, a 5G E2E service is not provided by MATLAB 5G toolbox. Moreover, the 5G toolbox is under MathWorks licensing.

1.5.2 5G R&D Projects

Several 5G consortiums of mobile operators, telecom services providers, and Academia have launched multiple projects to provide agile, programmable, virtualized, and scalable 5G E2E services. These consortiums are developing platforms for RAN, core, edge/cloud, and application layer based on the 5G standards such as 3GPP, O-RAN, and ETSI. In this part, two 5G projects are highlighted.

1.5.2.1 MOSAIC 5G

MOSAIC5G [226] is a 5G project that leverages the various 5G enabling technologies such as SDN, NFV, network slicing, edge computing, and AI. It provides several open-source platforms involving FlexRAN, low latency-MEC (LL-MEC), Juju-based orchestrator core (JOX), and Apps store. FlexRAN is based on SDN and NFV concepts to configure the 5G core and RAN for handover, resources, bandwidth, and so on. LL-MEC provides elastic and low latency services for the end-user devices. On the other hand, JOX offers network slicing solutions to cover several use cases including e-health and IoT. The vertical industries can parametrize their use cases by using open-source APIs and software development kits (SDKs) provided in the App store. MOSAIC5G is an open-source platform under Apache V2.0 license and can be summarized as follows:

1. OAI-RAN and OAI-CN: MOSAIC5G operates on the top of OAI software to deliver 5G E2E services.
2. FlexRAN: It is an O-RAN-compliant platform. It aims to provide flexible RAN control by enabling programmability, virtualization, and management of the 5G RAN based on SDN and NFV concepts. Decoupling the user and control planes is envisioned to provide a master control to manage and orchestrate the gNBs and UEs. It also facilitates the deployment of third-party and industry applications by providing plug&play control apps such as App-to-App and App-to-RAN.
3. LL-MEC: Besides the low latency services, LL-MEC platform aims to provide an abstraction layer between RAN and 5GC data plane. It monitors and controls the user data per service to provide real-time RAN information.
4. JOX: It is a Juju-based orchestrator that provides network virtualization, automation, slicing, and subslicing services. Furthermore, it provides northbound APIs based on 3GPP [227] and REST API to instantiate, prepare, configure, program, monitor, and control each slice.
5. Store: It aims to enable rapid deployment for 5G solutions by providing dedicated services via Apps and without knowledge of the underlying network. Moreover, it provides several services including 5G platform as a service (PaaS), datasets for ML and analytics, and SDKs for FlexRAN, LL-MEC, and JoX.

1.5.2.2 Open Networking Foundation (ONF)

Besides the standardization of OpenFlow protocol, ONF provides several open projects and solutions for mobile, broadband, and enterprise networks by leveraging the SDN, NFV, and edge cloud technologies. For the mobile networks, it provides software-defined-RAN (SD-RAN), open mobile evolved core (OMEC), and converged multi-access and core (COMAC) projects for RAN and core. In addition, it provides Aether solution for the edge in mobile and enterprise networks. Meanwhile, ONF offers SDN-enabled broadband access (SEBA) and virtual OLT hardware abstraction (VOLTHA) solutions for the broadband networks [228]. Moreover, it provides a central office re-architected as a datacenter (CORD) project to leverage the key 5G enabling technologies, e.g., SDN and NFV, at the network edge. The CORD project provides cloud-native, open, programmable, flexible, and agile services for the MNOs [229].

ONF mobile network projects can be summarized as follows:

1. COMAC: It is an open-source platform built under the CORD project for the RAN and core to provide convergence between the mobile and broadband networks. By leveraging SDN, RAN, core, and edge, COMAC project aims to optimize the QoE for the end-users.
2. SD-RAN: It provides an open-source RAN platform to develop the near-RT RIC entity based on O-RAN architecture (see Sect. 1.2.2.2), in order to control, configure, and manage the RAN by xApps.
3. OMEC: It is an open-source project to offer scalable and virtualized EPC compatible with 3GPP release 13.
4. Aether: It is a 5G open-source edge platform to provide edge computing services and wireless licensed and unlicensed connectivity for enterprises.

1.6 Challenging Issues and Directions

Tremendous research works have been conducted, in the literature, on 5G technology to boost the network performance and to fulfill the diverse demands of mobile operators and vertical industry. Herein, several challenging issues and future directions are highlighted on public and private 5G architectures as well as 5G enabling technologies, such as SDN, NFV, network slicing, and MEC. These challenging issues can be summarized as follows:

1. Cloud RAN: A main C-RAN issue is to achieve high QoS. The BBU pool might be placed faraway from some end-user devices which lead to degradation of the QoS. Therefore, efficient frameworks for BBU placement and resource allocations are mandated in order to achieve high QoS for the end-users. Meanwhile, a shared BBU pool demands a universal security and privacy mechanism. Accordingly, efficient and universal security and privacy preservation frameworks for C-RAN are challenging issue. In addition, the high data traffic

generated from the massive number of end-user devices might overwhelm the fronthaul link between the RRH and the BBU in C-RAN. Therefore, H-CRAN was proposed to manage the traffic between SCs or LPN and macro-cell or HPN. However, H-CRAN faces several challenging issues in terms of optimal planning to avoid interference, energy consumption, and demanding low latency and high capacity fronthaul and backhaul links.

V-CRAN was proposed to virtualize the E2E resources of 5G networks and thereby provides cost-effective solutions. However, V-CRAN faces several challenging issues in terms of latency and heterogeneity of fronthaul links, unified protocol, security and privacy, resources management and allocation, VMs migration, placement, and management of SDN controllers and network isolation. Trust mechanisms are mandated for V-CRAN to provide security for multiple operators sharing the same resources on the BBU pool. Another challenging issue of the various C-RAN architectures is regarding the high round-trip delay to process the end-user device data by the cloud/BBU pool. This limitation of high delay of C-RAN can be tackled by adopting edge computing paradigms within the C-RAN architecture where part of the data can be processed or stored at the edge servers. Consequently, O-RAN and 5GPPP have adopted edge computing within NG-RAN.

2. Edge RAN: O-RAN and 5GPPP proposed different RAN architectures than the C-RAN provided by IBM. O-RAN proposed six deployment scenarios for the installation of RAN and core functionalities at the edge or regional cloud. Meanwhile, 5GPPP proposed NG-RAN with MEC and several UCs. In addition, in O-RAN and 5GPPP architectures, the BBU is separated into DU and CU. Furthermore, edge computing has been considered in order to provide ultralow latency and real-time processing. We found that the 5G architectural frameworks of O-RAN and 5GPPP define different control layers with different interfaces. Moreover, the 5G RAN interfaces have been defined by 3GPP and O-RAN, but some interfaces are still undefined. Besides, we observed that the low physical layer split option has been postponed by 3GPP due to the demand of high transport bandwidth with large antennas. Among the eight deployment scenarios to operate the 5G networks, 3GPP decided to drop scenarios 6 and 8 due to complexities to operate the gNB directly on the EPC.
3. Private 5G networks: We observed that the 3GPP did not define a standard for the NR-U for frequency bands higher than 7 GHz. Moreover, several challenges have been identified due to LBT protocol failure involving the synchronization signals, paging, scheduling, and so on. Few research works on NR-U have been proposed in the literature to tackle the various challenging issues on efficient and new methods for the channel access procedures.
4. SDN and NFV: The separation of the control and user planes and a centralized SDN controller are foreseen to increase the potential security threats. In addition, a centralized SDN controller has a key challenge of a single point failure and scaling limitations. Furthermore, a distributed SDN controller increases the design complexity. In the meanwhile, several challenging issues have been addressed for efficient NFV integration, management, orchestration, and resource optimization

within the 5G networks. In addition, a standardizing of a unified northbound interface for NFV is considered as a challenging issue. Moreover, considering that the network functions will be executed on VMs, the security threat is found to be as a challenging problem for NFV. Accordingly, efficient frameworks for the placement of SDN controllers, unified northbound APIs, fair resource distribution, resource optimization, and security are recommended for future research works on SDN and NFV.

5. Network slicing: The mobility and handover between different RAN in the cellular networks are posing several challenges for network slicing in terms of mobility-aware slicing, anticipation of user mobility, pre-booking resources, and so on. Furthermore, analytical studies and economic models on the service-level agreement are required in order to maximize the revenue of MNOs and third parties.
6. MEC: It faces several challenging issues in terms of the management of third-party applications, bringing the security and privacy drawbacks of the cloud into the MEC servers, and high maintenance cost in case of MEC failure. Therefore, efficient optimization and security solutions are mandated to enhance the performance of MEC. In addition, in case of mobility, a universal framework for offloading the tasks or applications and coexistence between different MEC is required.

References

1. A. Toskala, H. Holma, T. Nakamura, *5G Technology: 3GPP New Radio* (John Wiley and Sons, Hoboken, 2020)
2. O-RAN Alliance, O-RAN use cases and deployment scenarios (2021). https://static1.squarespace.com/static/5ad774cce74940d7115044b0/t/5e95a0a306c6ab2d1cbca4d3/1586864301196/O-RAN+Use+Cases+and+Deployment+Scenarios+Whitepaper+February+2020.pdf. Accessed Jan 2021
3. M.A. Habibi, M. Nasimi, B. Han, H.D. Schotten, A comprehensive survey of RAN architectures toward 5G mobile communication system. IEEE Access **7**, 70371–70421 (2019)
4. M.F. Hossain, A.U. Mahin, T. Debnath, F.B. Mosharrof, K.Z. Islam, Recent research in cloud radio access network (C-RAN) for 5G cellular systems – a survey. J. Netw. Comput. Appl. **139**, 31–48 (2019)
5. W. Ejaz, S.K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, N.A. Chughtai, A comprehensive survey on resource allocation for CRAN in 5G and beyond networks. J. Netw. Comput. Appl. (2020). https://doi.org/10.1016/j.jnca.2020.102638
6. F. Tian, P. Zhang, Z. Yan, A survey on C-RAN security. IEEE Access **5**, 13372–13386 (2017)
7. M. Agiwal, A. Roy, N. Saxena, Next generation 5G wireless networks: a comprehensive survey. IEEE Commun. Surv. Tutorials **18**(3), 1617–1655 (2016)
8. S.K. Sharma, X. Wang, Toward massive machine type communications in ultra-dense cellular IoT networks: current issues and machine learning-assisted solutions. IEEE Commun. Surv. Tutorials **22**(1), 426–471 (2020)
9. I.B.F. de Almeida, L.L. Mendes, J.J.P.C. Rodrigues, M.A.A. da Cruz, 5G waveforms for IoT applications. IEEE Commun. Surv. Tutorials **21**(3), 2554–2567 (2019). Thirdquarter
10. L. Chettri, R. Bera, A comprehensive survey on internet of things (IoT) toward 5G wireless systems. IEEE Internet Things J. **7**(1), 16–32 (2020)

11. G.A. Akpakwu, B.J. Silva, G.P. Hancke, A.M. Abu-Mahfouz, A survey on 5G networks for the internet of things: communication technologies and challenges. IEEE Access **6**, 3619–3647 (2018)
12. J. Ding, M. Nemati, C. Ranaweera, J. Choi, IoT connectivity technologies and applications: a survey. IEEE Access **8**, 67646–67673 (2020)
13. F. Al-Turjman, E. Ever, H. Zahmatkesh, Small cells in the forthcoming 5G/IoT: traffic modelling and deployment overview. IEEE Commun. Surv. Tutorials **21**(1), 28–65 (2019)
14. M. Wazid, A.K. Das, S. Shetty, P. Gope, J.J.P.C. Rodrigues, Security in 5G-enabled internet of things communication: issues, challenges, and future research roadmap. IEEE Access **9**, 4466–4489 (2021)
15. F. Spinelli, V. Mancuso, Toward enabled industrial verticals in 5G: a survey on MEC-based approaches to provisioning and flexibility. IEEE Commun. Surv. Tutorials **23**(1), 596–630 (2021)
16. M. Shafi et al., 5G: a tutorial overview of standards, trials, challenges, deployment, and practice. IEEE J. Sel. Areas Commun. **35**(6), 1201–1221 (2017)
17. M.A. Adedoyin, O.E. Falowo, Combination of ultra-dense networks and other 5G enabling technologies: a survey. IEEE Access **8**, 22893–22932 (2020)
18. S. Chen, R. Ma, H. Chen, H. Zhang, W. Meng, J. Liu, Machine-to-machine communications in ultra-dense networks—a survey. IEEE Commun. Surv. Tutorials **19**(3), 1478–1503 (2017)
19. M. Kamel, W. Hamouda, A. Youssef, Ultra-dense networks: a survey. IEEE Commun. Surv. Tutorials **18**(4), 2522–2545 (2016)
20. A.C. Baktir, A. Ozgovde, C. Ersoy, How can edge computing benefit from software-defined networking: a survey, use cases, and future directions. IEEE Commun. Surv. Tutorials **19**(4), 2359–2391 (2017)
21. 3GPP TR38.913, Study on scenarios and requirements for next generation access technologies (2020). https://www.3gpp.org/ftp//Specs/archive/38_series/38.913/. Accessed Jan 2021
22. M. Agiwal, H. Kwon, S. Park, H. Jin, A survey on 4G-5G dual connectivity: road to 5G implementation. IEEE Access **9**, 16193–16210 (2021)
23. D. Soldani, M. Shore, J. Mitchell, M. Gregory, The 4G to 5G network architecture evolution in Australia. J. Telecommun. Digit. Econ. **6**, 1–30 (2018)
24. 3GPP TR 38.801, Technical Specification Group Radio Access Network; Study on new radio access technology: Radio access architecture and interfaces (2017)
25. Y. Lin et al., Wireless network cloud: architecture and system requirements. IBM J. Res. Develop. **54**(1) (2010)
26. China Mobile Res. Inst., C-RAN the Road Towards Green RAN-White Paper. China Mobile Beijing (2011)
27. H. Venkataraman, R. Trestian, *5G Radio Access Networks: Centralized RAN, Cloud-RAN and Virtualization of Small Cells* (CRC Press, Boca Raton, 2017)
28. J. Wu, Z. Zhang, Y. Hong, Y. Wen, Cloud radio access network (C-RAN): a primer. IEEE Netw. **29**(1), 35–41 (2015)
29. O. Alamu, A. Gbenga-Ilori, M. Adelabu, A. Imoize, O. Ladipo, Energy efficiency techniques in ultra-dense wireless heterogeneous networks: an overview and outlook. Eng. Sci. Technol. Int. J. **23**(6), 1308–1326 (2020)
30. M. Peng, Y. Li, J. Jiang, J. Li, C. Wang, Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. IEEE Wirel. Commun. **21**(6), 126–135 (2014)
31. Y. Li, T. Jiang, K. Luo, S. Mao, Green heterogeneous cloud radio access networks: potential techniques, performance trade-offs, and challenges. IEEE Commun. Mag. **55**(11), 33–39 (2017)
32. M. Peng, Y. Li, Z. Zhao, C. Wang, System architecture and key technologies for 5G heterogeneous cloud radio access networks. IEEE Netw. **29**(2), 6–14 (2015)
33. 3GPP TR 36.932, Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN (2020). https://www.3gpp.org/ftp/Specs/archive/36_series/36.932/. Accessed Nov 2021

34. ETSI, Network function virtualization: use cases (2013). www.etsi.org. Accessed Nov 2021
35. H. Hawilo, A. Shami, M. Mirahmadi, R. Asal, NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC). IEEE Netw. **28**(6), 18–26 (2014)
36. M. Arslan, K. Sundaresan, S. Rangarajan, Software-defined networking in cellular radio access networks: potential and challenges. IEEE Commun. Mag. **53**(1), 150–156 (2015)
37. X. Wang et al., Virtualized cloud radio access network for 5G transport. IEEE Commun. Mag. **55**(9), 202–209 (2017)
38. I.F. Akyildiz, P. Wang, S.-C. Lin, SoftAir: a software defined networking architecture for 5G wireless systems. Comput. Netw. **85**, 1–18 (2015)
39. E. Markakis, G. Mastorakis, C.X. Mavromoustakis, E. Pallis, Cloud and Fog Computing in 5G Mobile Networks: Emerging Advances and Applications (The Institution of Engineering and Technology, London, 2017)
40. M. Peng, S. Yan, K. Zhang, C. Wang, Fog-computing-based radio access networks: issues and challenges. IEEE Netw. **30**(4), 46–53 (2016)
41. M. Peng, Z. Zhao, Y. Sun, *Fog Radio Access Networks (F-RAN) Architectures, Technologies, and Applications* (Springer International Publishing, Berlin, 2020)
42. O-RAN Alliance Technical Specifications, O-RAN Architecture Description v02.00 (2020). https://www.o-ran.org. Accessed Jan 2021
43. O-RAN Alliance, https://www.o-ran.org. Accessed Jan 2021
44. A. Garcia-Saavedra, X. Costa-Perez, O-RAN: disrupting the virtualized RAN ecosystem. IEEE Commun. Stand. Mag. (2021). https://doi.org/10.1109/MCOMSTD.101.2000014
45. 5G PPP Architecture Working Group, View on 5G Architecture, Version 3.0 – February (2020)
46. 3GPP TS 23.501, TSG RAN, System architecture for the 5G System (5GS) Stage 2 v16.6.0 (2020)
47. Qualcomm, White Paper: Private 5G Mobile Networks for Industrial IoT (2019). www.qualcomm.com/media/documents/files/private-5g-networks-for-industrial-iot.pdf. Accessed Jan 2021
48. 5G-ACIA, White Paper: 5G Non-Public Networks for Industrial Scenarios (2019). https://5g-acia.org/wp-content/uploads/2021/04/WP_5G_NPN_2019_01.pdf. Accessed Jan 2021
49. 5G-Smart, D5.2: First Report on 5G Network Architecture Options and Assessments (2020). https://5gsmart.eu/wp-content/uploads/5G-SMART-D5.2-v1.0.pdf. Accessed Jan 2021
50. A. Aijaz, Private 5G: the future of industrial wireless. IEEE Indust. Electron. Mag. **14**(4), 136–145 (2020)
51. MulteFire, Technical Paper: A New Way to Wireless (2021). www.multefire.org/wp-content/uploads/MulteFire-Release-1.0-whitepaper_FINAL.pdf. Accessed Nov 2021
52. 3GPP TS 23.251, Network sharing; Architecture and functional description (2020). https://www.3gpp.org/ftp/Specs/archive/23_series/23.251/. Accessed Nov 2021
53. Multefire, Cellular-based technology — LTE or 5G NR — operating in unlicensed or shared spectrum (2020). https://www.multefire.org/. Accessed Dec 2020
54. 3GPP TR 23.734, Study on Enhancement of 5G System (5GS) for Vertical and Local Area Network (LAN) Services (Release 16) (2020). https://www.3gpp.org/ftp/Specs/archive/23series/23.734/. Accessed Dec 2020
55. IDC, Worldwide Internet of Things Forecast Update 2020–2024 (2020). www.reportlinker.com/p05352129/Worldwide-Internet-of-Things-Forecast-Update.html. Accessed Dec 2021
56. A.W. Dawson, M.K. Marina, F.J. Garcia, On the benefits of RAN virtualisation in C-RAN based mobile networks, in *2014 Third European Workshop on Software Defined Networks*, London (2014), pp. 103–108
57. E.J. Kitindi, S. Fu, Y. Jia, A. Kabir, Y. Wang, Wireless network virtualization with SDN and C-RAN for 5G networks: requirements, opportunities, and challenges. IEEE Access **5**, 19099–19115 (2017)
58. T.Q. Duong, X. Chu, H.A. Suraweera, *Ultra-Dense Networks for 5G and Beyond: Modelling, Analysis, and Applications* (John Wiley & Sons Ltd., Hoboken, 2019)

59. H. Zhang, Y. Dong, J. Cheng, M.J. Hossain, V.C.M. Leung, Fronthauling for 5G LTE-U ultra dense cloud small cell networks. IEEE Wirel. Commun. **23**(6), 48–53 (2016)
60. ETSI, NFV White paper: Network Functions Virtualisation, An Introduction, Benefits, Enablers, Challenges & Call for Action (2012). https://portal.etsi.org/NFV/NFV_White_Paper.pdf. Accessed Oct 2021
61. M. Masoudi, S.S. Lisi, C. Cavdar, Cost-effective migration toward virtualized C-RAN with scalable fronthaul design. IEEE Syst. J. **14**(4), 5100–5110 (2020)
62. S. Su, X. Xu, Z. Tian, M. Zhao, W. Wang, 5G fronthaul design based on software-defined and virtualized radio access network, in *2019 28th Wireless and Optical Communications Conference (WOCC)* (2019), pp. 1–5
63. S. Hung, H. Hsu, S. Lien, K. Chen, Architecture harmonization between cloud radio access networks and fog networks. IEEE Access **3**, 3019–3034 (2015)
64. M. De Donno, K. Tange, N. Dragoni, Foundations and evolution of modern computing paradigms: cloud, IoT, edge, and fog. IEEE Access **7**, 150936–150948 (2019)
65. S. Naveen, M.R. Kounte, Key technologies and challenges in IoT edge computing, in *2019 Third International Conference on I-SMAC* (2019), pp. 61–65
66. D. Loghin, L. Ramapantulu, Y.M. Teo, On understanding time, energy and cost performance of wimpy heterogeneous systems for edge computing, in *IEEE International Conference on Edge Computing (EDGE)* (2017), pp. 1–8
67. H. El-Sayed et al., Edge of things: the big picture on the integration of edge, IoT and the cloud in a distributed computing environment. IEEE Access **6**, 1706–1717 (2018)
68. Y. Liu, M. Peng, G. Shou, Y. Chen, S. Chen, Toward edge intelligence: multiaccess edge computing for 5G and internet of things. IEEE Internet of Things J. **7**(8), 6722–6747 (2020)
69. M. Caprolu, R. Di Pietro, F. Lombardi, S. Raponi, Edge computing perspectives: architectures, technologies, and open security issues, in *2019 IEEE International Conference on Edge Computing (EDGE)* (2019), pp. 116–123
70. Cisco, Cisco delivers vision of fog computing to accelerate value from billions of connected devices (2020). http://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1334100. Accessed May 2021
71. J.S. Preden, K. Tammemäe, A. Jantsch, M. Leier, A. Riid, E. Calis, The benefits of self-awareness and attention in fog and mist computing. Computer **48**(7), 37–45 (2015)
72. OpenFog Consortium, https://www.openfogconsortium.org. Accessed July 2021
73. M. Sapienza, E. Guardo, M. Cavallo, G. La Torre, G. Leombruno, O. Tomarchio, Solving critical events through mobile edge computing: an approach for smart cities, in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)* (2016), pp. 1–5
74. I. Badmus, M. Matinmikko-Blue, J.S. Walia, T. Taleb, Network slice instantiation for 5G micro-operator deployment scenarios, in *2019 European Conference on Networks and Communications (EuCNC)* (2019), pp. 133–138
75. H. Li, G. Shou, Y. Hu, Z. Guo, Mobile edge computing: progress and challenges, in *2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)* (2016), pp. 83–84
76. S. Kukliński, L. Tomaszewski, R. Kołakowski, On O-RAN, MEC, SON and Network Slicing integration, in *2020 IEEE Globecom Workshops (GC Wkshps)* (2020), pp. 1–6
77. M. Barahman, L.M. Correia, L.S. Ferreira, QoS-demand-aware computing resource management scheme in cloud-RAN. IEEE Open J. Commun. Soc. **1**, 1850–1863 (2020)
78. S.S. Arnob, I. Islam Shovon, T. Ahmed, M.S. Ullah, R. Shelim, Dual-order resource allocation in 5G H-CRAN using matching theory and ant colony optimization algorithm, in *IECON 2020 The 46th Annual Conference of the IEEE Industrial* (2020), pp. 2101–2107
79. Y. Ai, G. Qiu, C. Liu, Y. Sun, Joint resource allocation and admission control in sliced fog radio access networks. China Commun. **17**(8), 14–30 (2020)
80. F. Mungari, An RL approach for radio resource management in the O-RAN architecture, in *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (2021), pp. 1–2

81. B. Xu, P. Zhu, J. Li, D. Wang, X. You, Joint long-term energy efficiency optimization in C-RAN with hybrid energy supply. IEEE Trans. Veh. Technol. **69**(10), 11128–11138 (2020)
82. L. Ferdouse, I. Woungang, A. Anpalagan, S. Erkucuk, Energy efficient downlink resource allocation in cellular IoT supported H-CRANs. IEEE Trans. Veh. Technol. **70**(6), 5803–5816 (2021)
83. T.H.L. Dinh, M. Kaneko, E.H. Fukuda, L. Boukhatem, Energy efficient resource allocation optimization in fog radio access networks with outdated channel knowledge. IEEE Trans. Green Commun. Netw. **5**(1), 146–159 (2021)
84. T. Pamuklu, S. Mollahasani, M. Erol-Kantarci, Energy-efficient and delay-guaranteed joint resource allocation and DU selection in O-RAN, in *2021 IEEE 4th 5G World Forum (5GWF)* (2021), pp. 99–104
85. S. Park, O. Simeone, S. Shamai, Multi-tenant C-RAN with spectrum pooling: downlink optimization under privacy constraints. IEEE Trans. Veh. Technol. **67**(11), 10492–10503 (2018)
86. I. Al-Samman, R. Almesaeed, A. Doufexi, M. Beach, A. Nix, User weighted probability algorithm for heterogeneous C-RAN interference mitigation, in *2017 IEEE International Conference on Communications (ICC)* (2017), pp. 1–7
87. Y. Yu, S. Liu, Z. Tian, S. Wang, A dynamic distributed spectrum allocation mechanism based on game model in fog radio access networks. China Commun. **16**(3), 12–21 (2019)
88. D.S. Dong, K. Khatri, A. Gachhadar, Network coding based secure and efficient traffic flow in Heterogeneous Cloud Radio Access Network, in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (2017), pp. 584–589
89. D. Dik, M.S. Berger, Transport security considerations for the open-RAN fronthaul, in *2021 IEEE 4th 5G World Forum (5GWF)* (2021), pp. 253–258
90. F. Tonini, C. Raffaelli, L. Wosinska, P. Monti, Cost-optimal deployment of a C-RAN with hybrid fiber/FSO fronthaul. J. Opt. Commun. Netw. **11**(7), 397–408 (2019)
91. K. Wang, K. Yang, X. Wang, C.S. Magurawalage, Cost-effective resource allocation in C-RAN with mobile cloud, in *2016 IEEE International Conference on Communications (ICC)* (2016), pp. 1–6
92. D. Pliatsios, P. Sarigiannidis, I.D. Moscholios, A. Tsiakalos, Cost-efficient remote radio head deployment in 5G networks under minimum capacity requirements, in *2019 Panhellenic Conference on Electronics & Telecommunications (PACET)* (2019), pp. 1–4
93. O. Chabbouh, S. Ben Rejeb, N. Nasser, N. Agoulmine, Z. Choukair, Novel cloud-RRH architecture with radio resource management and QoS strategies for 5G HetNets. IEEE Access **8**, 164815–164832 (2020)
94. 3GPP TS 22.104, Service requirements for cyber-physical control applications in vertical domains (2021). https://www.3gpp.org/ftp/Specs/archive/22_series/22.104/. Accessed June 2021
95. C. Rosa, M. Kuusela, F. Frederiksen, K.I. Pedersen, Standalone LTE in unlicensed spectrum: radio challenges, solutions, and performance of multefire. IEEE Commun. Mag. **56**(10), 170–177 (2018)
96. Nokia, Industrial-grade Private Wireless (2021). https://www.nokia.com/networks/solutions/private-wireless/. Accessed Aug 2021
97. GSMA, Spectrum Sharing (2021). www.gsma.com/spectrum/wp-content/uploads/2021/06/Spectrum-Sharing-Positions.pdf. Accessed Aug 2021
98. LTE-U Forum, LTE-U Technical Report: Coexistence study for LTE-U SDL V1.0 (2015)
99. R. Zhang, M. Wang, L.X. Cai, Z. Zheng, X. Shen, L.-L. Xie, LTEunlicensed: the future of spectrum aggregation for cellular networks. IEEE Wirel. Commun. **22**(3), 150–159 (2015)
100. 3GPP TR 36.889, Study on Licensed Assisted Access to Unlicensed Spectrum (Release 13) (2015). https://www.3gpp.org/ftp/Specs/archive/36_series/36.889/. Accessed Aug 2021
101. 3GPP TR 38.889, Study on NR-based access to unlicensed spectrum (Release 15) (2018). https://www.3gpp.org/ftp/Specs/archive/38_series/38.889/. Accessed Aug 2021

102. 3GPP TSG-RAN 86, Document RP-193259: Study on Supporting NR From 52.6 GHz to 71 GHz (2019). https://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_101-e/Inbox/drafts. Accessed Aug 2021
103. 3GPP TSG-RAN 86, Document RP-193229: Extending Current NR Operation to 71 GHz (2019). https://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_90e/Inbox/Drafts/. Accessed Aug 2021
104. S. Lagen et al., New radio beam-based access to unlicensed spectrum: design challenges and solutions. IEEE Communl Survl Tutorials **22**(1), 8–37 (2020)
105. F. Hu, Q. Hao, K. Bao, A survey on software-defined network and openflow: from concept to implementation. IEEE Commun. Surv. Tutorials **16**(4), 2181–2206 (2014)
106. S. Sezer et al., Are we ready for SDN? Implementation challenges for software-defined networks. IEEE Commun. Mag. **103**(1), 14–76 (2015)
107. D. Kreutz, F.M.V. Ramos, P.E. Veríssimo, C.E. Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined networking: a comprehensive survey. Proc. IEEE **103**(1), 14–76 (2015)
108. A. Mavromatis et al., A software-defined IoT device management framework for edge and cloud computing. IEEE Internet Things J. **7**(3), 1718–1735 (2020)
109. T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, D. Sabella, On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration. IEEE Commun. Surv. Tutorials **19**(3), 1657–1681 (2017)
110. S.D.A. Shah, M.A. Gregory, S. Li, R.D.R. Fontes, SDN enhanced multi-access edge computing (MEC) for E2E mobility and QoS management. IEEE Access **8**, 77459–77469 (2020)
111. J. Tourrilhes, P. Sharma, S. Banerjee, J. Pettit, SDN and OpenFlow evolution: a standards perspective. Computer **47**(11), 22–29 (2014)
112. P. Farzaneh, P. Marius, T.W. Lum, I. Jadwiga, Efficient topology discovery in software defined networks, in *8th International Conference on Signal Processing and Communication Systems (ICSPCS)* (2014)
113. J.S. Choi, X. Li, Hierarchical distributed topology discovery protocol for multi-domain SDN networks. IEEE Commun. Lett. **21**(4), 773–776 (2017)
114. M. Obadia, M. Bouet, J. Leguay, K. Phemius, L. Iannone, Failover mechanisms for distributed SDN controllers, in *International Conference and Workshop on the Network of the Future (NOF)*, Paris (2014), pp. 1–6
115. R. Hwang, Y. Tang, Fast failover mechanism for SDN-enabled data centers, in *International Computer Symposium (ICS)* (2016), pp. 171–176
116. K. Fang, K. Wang, J. Wang, A fast and load-aware controller failover mechanism for software-defined networks, in *10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)* (2016), pp. 1–6
117. K.S. Sahoo, B. Sahoo, CAMD: a switch migration based load balancing framework for software defined networks. IET Netw. **8**(4), 264–271 (2019)
118. O. Adekoya, A. Aneiba, M. Patwary, An improved switch migration decision algorithm for SDN load balancing. IEEE Open J. Commun. Soc. **1**, 1602–1613 (2020)
119. T. Wang, F. Liu, H. Xu, An efficient online algorithm for dynamic SDN controller assignment in data center networks. IEEE/ACM Trans. Netw. **25**(5), 2788–2801 (2017)
120. K. Sagar Sahoo et al., Metaheuristic solutions for solving controller placement problem in SDN-based WAN architecture, in *14th International Joint Conference on e-Business and Telecommunications*, vol. 1 (2017), pp. 15–23
121. S. Lange et al., Heuristic approaches to the controller placement problem in large scale SDN networks. IEEE Trans. Netw. Ser. Manag. **12**(1), 4–17 (2015)
122. H. Kuang, Y. Qiu, R. Li, X. Liu, A hierarchical K-means algorithm for controller placement in SDN-based WAN architecture, in *10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (2018), pp. 263–267
123. X. Yang, D. Wang, W. Tang, W. Feng, C. Zhu, IPsec cryptographic algorithm invocation considering performance and security for SDN southbound interface communication. IEEE Access **8**, 181782–181795 (2020)

124. S. Midha, K. Triptahi, Extended TLS security and defensive algorithm in OpenFlow SD, in *9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (2019), pp. 141–146
125. Z. Latif, K. Sharif, F. Li, Md.M. Karim, Y. Wang, A comprehensive survey of interface protocols for software defined networks. J. Netw. Comput. Appl. (2020). https://arxiv.org/abs/1902.07913. Accessed Dec 2021
126. D. Hasan, M. Othman, Efficient topology discovery in software defined networks: revisited. J. Netw. Comput. Appl. **156**, 539–547 (2017)
127. L. EL-Garoui, S. Pierre, S. Chamberland, A New SDN-based routing protocol for improving delay in smart city environments. Smart Cities **3**(3), 1004–1021 (2020). https://doi.org/10.3390/smartcities3030050
128. M. Abdollahi, M. Abolhasan, N. Shariati, J. Lipman, A. Jamalipour, W. Ni, A routing protocol for SDN-based multi-hop D2D communications, in *16th IEEE Annual Consumer Communications & Networking Conference (CCNC)* (2019), pp. 1–4
129. K. Indira, P. Ajitha, V. Reshma, A. Tamizhselvi, An efficient secured routing protocol for software defined internet of vehicles, in *International Conference on Computational Intelligence in Data Science (ICCIDS)* (2019), pp. 1–4
130. M.J. Anjum, I. Raza, S.A. Hussain, Real-time multipath transmission protocol (RMTP): a software defined networks (SDN) based routing protocol for data centric networks, in *International Conference on Electrical, Communication, and Computer Engineering (ICECCE)* (2019), pp. 1–6
131. O. Lemeshko, O. Nevzorova, V. Rossikhin, A.M. Hailan, Hierarchical method of load balancing routing on SDN controllers with multicore architecture, in *International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)* (2018), pp. 457–460
132. A. Azzouni et al., sOFTDP: Secure and Efficient Topology Discovery Protocol for SDN (2017). https://hal.sorbonne-universite.fr/hal-01538564/file/sOFTDP.pdf. Accessed Dec 2021
133. N. Abdolmaleki, M. Ahmadi, H.T. Malazi, S. Milardo, Fuzzy topology discovery protocol for SDN-based wireless sensor networks. Simul. Model. Pract. Theory **79**, 54–68 (2017)
134. F.Z. Yousaf, M. Bredel, S. Schaller, F. Schneider, NFV and SDN—key technology enablers for 5G networks. IEEE J. Sel. Areas Commun. **35**(11), 2468–2478 (2017)
135. Y. Hu, J. Wang, Architectural and cost implications of the 5G edge NFV systems, in *IEEE 37th International Conference on Computer Design (ICCD)* (2019), pp. 594–603
136. L. Zhang et al., Characterizing and orchestrating NFV-ready servers for efficient edge data processing, in *IEEE/ACM 27th International Symposium on Quality of Service (IWQoS)* (2019), pp. 1–10
137. ETSI, ETSI GS NFV 002: Network Functions Virtualisation (NFV); Architectural Framework (2014). https://www.etsi.org/deliver/etsi_gs/nfv/001_099/002/01.02.01_60/gs_nfv002v010201p.pdf. Accessed Dec 2021
138. 5GPPP SN WG, Vision on Software Networks and 5G (2017). https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP_SoftNets_WG_whitepaper_v20.pdf. Accessed Dec 2021
139. ETSI, ETSI ISG NFV working group (2021). https://www.etsi.org/technologies/nfv. Accessed Dec 2021
140. 3GPP TS 28.500, Management concept, architecture and requirements for mobile networks that include virtualized network function (2020). https://www.3gpp.org/ftp//Specs/archive/28_series/28.500/. Accessed Dec 2021
141. K. Sienkiewicz, W. Latoszek, P. Krawiec, Services orchestration within 5G networks — challenges and solutions, in *Baltic URSI Symposium (URSI)* (2018), pp. 265–268
142. M. Casazza, M. Bouet, S. Secci, Availability-driven NFV orchestration. Comput. Netw. **155**, 47–61 (2019). https://doi.org/10.1016/j.comnet.2019.02.017
143. B. Gerő et al., The orchestration in 5G exchange – a multi-provider NFV framework for 5G services, in *IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)* (2017), pp. 1–2

144. X. Liang, X. Huang, D. Li, T. Yang, Dynamic orchestration mechanism of service function chain in hybrid NFV networks, in *Asia Communications and Photonics Conference (ACP)* (2018), pp. 1–3
145. G. Abolfazl, A. Behzad, T.M. Mahdi, Joint Reliability-aware and Cost Efficient Path Allocation and VNF Placement using Sharing Scheme (2019). https://arxiv.org/abs/1912.06742. Accessed Dec 2021
146. C. Park, D. Shin, VNF management method using VNF Group Table in Network Function Virtualization, in *19th International Conference on Advanced Communication Technology (ICACT)* (2017), pp. 210–212
147. M. Kumazaki, M. Ogura, T. Tachibana, VNF management with model predictive control for service chains, in *IEEE International Conference on Consumer Electronics – Taiwan (ICCE-TW)* (2019), pp. 1–2
148. Ruiz et al., A genetic algorithm for VNF provisioning in NFV-enabled cloud/MEC RAN architectures. Appl. Sci. **8**(12) (2018). https://doi.org/10.3390/app8122614
149. M. Huang, W. Liang, X. Shen, Y. Ma, H. Kan, Reliability-aware virtualized network function services provisioning in mobile edge computing. IEEE Trans. Mob. Comput. **19**(11), 2699–2713 (2020)
150. Q. Xia, W. Ren, Z. Xu, P. Zhou, W. Xu, G. Wu, Learn to optimize: adaptive VNF provisioning in mobile edge clouds, in *17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (2020), pp. 1–9
151. J. Incheol, K. Gu-In, Genetic algorithm for service function chaining in NFV. Adv. Sci. Technol. Lett. **129**, 223–228 (2016)
152. M. Wang, B. Cheng, B. Li, J. Chen, Service function chain composition and mapping in NFV-enabled networks, in *IEEE World Congress on Services (SERVICES)* (2019), pp. 331–334
153. A. OI, M. Nakajima, Y. Soejima and M. Tahara, Reliable design method for service function chaining, in *20th Asia-Pacific Network Operations and Management Symposium (APNOMS)* (2019), pp. 1–4
154. L. Ochoa-Aday, C. Cervelló-Pastor, A. Fernández-Fernández, P. Grosso, An Online Algorithm for Dynamic NFV Placement in Cloud-Based Autonomous Response Networks. Symmetry. **10**(5), 163 (2018)
155. D. Gedia, L. Perigo, Latency-aware, static, and dynamic decision-tree placement algorithm for containerized SDN-VNF in OpenFlow architectures, in *IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)* (2019), pp. 1–7
156. M. Karimzadeh-Farshbafan, V. Shah-Mansouri, D. Niyato, A dynamic reliability-aware service placement for network function virtualization (NFV). IEEE J. Sel. Areas Commun. **38**(2), 318–333 (2020)
157. I. Sarrigiannis et al., VNF lifecycle management in an MEC-enabled 5G IoT architecture. IEEE Internet Things J. **7**(5), 4183–4194 (2020)
158. S. Lange et al., A network intelligence architecture for efficient VNF lifecycle management. IEEE Trans. Netw. Ser. Manag. **18**(2), 1476–1490 (2021)
159. ETSI, ETSI NFV API Specifications (2021). https://nfvwiki.etsi.org/index.php?title=API'_specifications. Accessed Dec 2021
160. 5GPPP, On Board Procedure to 5G PPP Infrastructure Projects (2020). https://5g-ppp.eu/wp-content/uploads/2020/04/. Accessed Dec 2021
161. 5G PPP, 5G Exchange (5GEx) project (2021). [Online]. Available: https://5g-ppp.eu/5gex/. Accessed Dec 2021
162. 5G PPP, Cloud-Native and Verticals services5G-PPP projects analysis (2019). https://5g-ppp.eu/wp-content/uploads/2019/09/. Accessed Dec 2021
163. L.U. Khan, I. Yaqoob, N.H. Tran, Z. Han, C.S. Hong, Network slicing recent advances, taxonomy, requirements, and open research challenges. IEEE Access **8**, 36009–36028 (2020)
164. 3GPP TS 38.300, NR and NG-RAN Overall Desciption (2020). https://www.3gpp.org/ftp//Specs/archive/38_series/38.300/. Accessed Jan 2021
165. S. Ahmadi, *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards* (Academic Press, Cambridge, 2019)

166. D. Sattar, A. Matrawy, Optimal slice allocation in 5G core networks. IEEE Netw. Lett. **1**(2), 48–51 (2019)
167. R. Wen et al., On robustness of network slicing for next-generation mobile networks. IEEE Trans. Commun. **67**(1), 430–444 (2019)
168. X. Shen et al., AI-assisted network-slicing based next-generation wireless networks. IEEE Open J. Veh. Technol. **1**, 45–66 (2020)
169. 3GPP TS 23.501, System architecture for the 5G system (2020). https://www.3gpp.org/ftp//Specs/archive/23_series/23.501/. Accessed Jan 2021
170. ITU-T Y.3112 series Y, Global information infrastructure, internet protocol aspects, next-generation networks, internet of things and smart cities (2018). https://www.itu.int/rec/T-REC-Y.3112-201812-I/en. Accessed Jan 2021
171. A.M. Escolar, J.M. Alcaraz-Calero, P. Salva-Garcia, J.B. Bernabe, Q. Wang, Adaptive network slicing in multi-tenant 5G IoT networks. IEEE Access **9**, 14048–14069 (2021)
172. M.A. Habibi, B. Han, H.D. Schotten, Network slicing in 5G mobile communication: architecture, prot modeling, and challenges, in *Proceedings of the 14th International Symposium Wireless Communications System* (2017), p. 16
173. Q. Li, An end-to-end network slicing framework for 5G wireless communication systems (2016). https://arxiv.org/abs/1608.00572. Accessed Jan 2021
174. O. Sallent, J. Pérez-Romero, R. Ferrús, R. Agustí, On radio access network slicing from a radio resource management perspective. IEEE Wirel. Commun. **24**(5), 166–174 (2017)
175. I.D. Silva, Impact of network slicing on 5G radio access networks. Proc. Eur. Conf. Netw. Commun., 153–157 (2016)
176. Y.L. Lee, J. Loo, T.C. Chuah, A new network slicing framework for multi-tenant heterogeneous cloud radio access networks. Proc. Int. Conf. Adv. Electr. Electron. Syst. Eng., 414–420 (2016)
177. J. Ordonez-Lucena et al., Network slicing for 5G with SDN/NFV: concepts, architectures, and challenges. IEEE Commun. Mag. **55**(5), 80–87 (2017)
178. M. Boldi, O. Queseth, P. Marsch, Ö. Bulakci, *5G System Design: Architectural and Functional Considerations and Long Term Research* (John Wiley and Sons, Hoboken, 2018)
179. ITU-T Y.3102, Series Y: Global Information Infrastructure, Internet Protocol Aspects, Next-Generation Networks, Internet of Things and Smart Cities (2018). https://www.itu.int/rec/T-REC-Y.3102/en. Accessed Nov 2021
180. NGMN Alliance, 5G End-to-End Architecture Framework (2019). https://www.ngmn.org/publications/5g-end-to-end-architecture-framework-v3-0-8.html. Accessed Nov 2021
181. 5GPPP, AI and ML – Enablers for Beyond 5G Networks (2021). https://5g-ppp.eu/wp-content/uploads/2021/05/AI-MLforNetworks-v1-0.pdf.Accessed Nov 2021
182. ITU, Architectural framework for machine learning in future networks including IMT-2020 (2019). https://www.itu.int/rec/T-REC-Y.3172-201906-I/en. Accessed Dec 2020
183. V.P. Kafle, Y. Fukushima, P. Martinez-Julia, T. Miyazawa, Consideration on automation of 5G network slicing with machine learning, in *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)* (2018), pp. 1–8
184. M.E. Morocho-Cayamcela, H. Lee, W. Lim, Machine learning for 5G/B5G mobile and wireless communications: potential, limitations, and future directions. IEEE Access **7**, 137184–137206 (2019)
185. M.S. Mollel et al., A survey of machine learning applications to handover management in 5G and beyond. IEEE Access **9**, 45770–45802 (2021)
186. J. Kaur, M.A. Khan, M. Iftikhar, M. Imran, Q. Emad Ul Haq, Machine learning techniques for 5G and beyond. IEEE Access **9**, 23472–23488 (2021)
187. M. Kaaviya, S. Deepa, Machine learning approaches for 5G network challenges. Int. J. Res. Eng. Sci. **9**(4) (2021)
188. ITU, Machine learning in future networks including IMT-2020: use cases (2019). https://www.itu.int/rec/T-REC-Y.Sup55-201910-I. Accessed Dec 2020

189. D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, X. Costa-Perez, Optimising 5G infrastructure markets: the business of network slicing, in *IEEE INFOCOM 2017 – IEEE Conference on Computer Communications* (2017), pp. 1–9
190. D. Bega, M. Gramaglia, M. Fiore, A. Banchs, X. Costa-Perez, DeepCog: cognitive network management in sliced 5G networks with deep learning, in *IEEE INFOCOM 2019 – IEEE Conference on Computer Communications* (2019), pp. 280–288
191. Y. Abiko, T. Saito, D. Ikeda, K. Ohta, T. Mizuno, H. Mineno, Flexible resource block allocation to multiple slices for radio access network slicing using deep reinforcement learning. IEEE Access **8**, 68183–68198 (2020)
192. T. Li, X. Zhu, X. Liu, An end-to-end network slicing algorithm based on deep Q-learning for 5G network. IEEE Access **8**, 122229–122240 (2020)
193. G. Kibalya, J. Serrat, J. Gorricho, R. Pasquini, H. Yao, P. Zhang, A reinforcement learning based approach for 5G network slicing across multiple domains, in *2019 15th International Conference on Network and Service Management (CNSM)* (2019), pp. 1–5
194. M.R. Raza, C. Natalino, P. Öhlen, L. Wosinska, P. Monti, Reinforcement learning for slicing in a 5G flexible RAN. J. Lightwave Technol. **37**(20), 5161–5169 (2019)
195. M. Yan, G. Feng, J. Zhou, Y. Sun, Y. Liang, Intelligent resource scheduling for 5G radio access network slicing. IEEE Trans. Veh. Technol. **68**(8), 7691–7703 (2019)
196. G. Sun, Z.T. Gebrekidan, G.O. Boateng, D. Ayepah-Mensah, W. Jiang, Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks. IEEE Access **7**, 45758–45772 (2019)
197. S. Troia, R. Alvizu, G. Maier, Reinforcement learning for service function chain reconfiguration in NFV-SDN metro-core optical networks. IEEE Access **7**, 167944–167957 (2019)
198. C. Vallati, A. Virdis, E. Mingozzi, G. Stea, Mobile-edge computing come home connecting things in future smart homes using LTE device-to-device communications. IEEE Consumer Electron. Mag. **5**(4), 77–83 (2016)
199. S. Abdelwahab, B. Hamdaoui, M. Guizani, T. Znati, Replisom: disciplined tiny memory replication for massive IoT devices in LTE edge cloud. IEEE Internet Things J. **3**(3), 327–338 (2016)
200. ETSI, Industry Specification Group (ISG) on Multi-Access Edge Computing (MEC) (2021). https://www.etsi.org/committee/1425-mec. Accessed May 2021
201. ETSI, ETSI GS MEC 003: Multi-access Edge Computing (MEC); Framework and Reference Architecture (2019). www.etsi.org/deliver/etsi_gs/mec/001_099/003/02.01.01_60/gs_mec003v020101p.pdf. Accessed March 2021
202. ETSI, ETSI White Paper No. 28; MEC in 5G networks (2018). https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf. Accessed Mar 2021
203. ETSI, ETSI GR MEC 031: Multi-access Edge Computing (MEC); MEC 5G Integration (2020). www.etsi.org/deliver/etsi_gr/MEC/001_099/031/02.01.01_60/gr_MEC031v020101p.pdf. Accessed Mar 2021
204. N. Abbas, Y. Zhang, A. Taherkordi, T. Skeie, Mobile edge computing: a survey. IEEE Internet Things J. **5**(1), 450–465 (2018)
205. Q. Pham et al., A survey of multi-access edge computing in 5G and beyond: fundamentals, technology integration, and state-of-the-art. IEEE Access **8**, 116974–117017 (2020)
206. S. Dario, R. Alex, F. Rui, *Multi-Access Edge Computing in Action* (CRC Press, Boca Raton, 2020)
207. ETSI, ETSI GS MEC 010-1: Mobile Edge Computing (MEC); Mobile Edge Management; Part 1: System, host and platform management (2017). https://www.etsi.org/deliver/etsi_gs/mec/001_099/01001/01.01.01_60/gs_mec01001v010101p.pdf. Accessed Mar 2021
208. ETSI, ETSI GS MEC 010-2: Mobile Edge Computing (MEC); Mobile Edge Management; Part 2: Application lifecycle, rules and requirements management (2017). https://www.etsi.org/deliver/etsi_gs/mec/001_099/01002/01.01.01_60/gs_mec01002v010101p.pdf. Accessed Mar 2021
209. L. Bonati et al., Open, programmable, and virtualized 5G networks: state-of-the-art and the road ahead. Comput. Netw. **182** (2020)

210. MATLAB 5G Toolbox (2020). www.mathworks.com/products/5g.html. Accessed Jan 2021
211. 5G LENA Project on NS-3 (2020). https://5g-lena.cttc.es/. Accessed Jan 2021
212. Vienna 5G Simulator (2020). www.nt.tuwien.ac.at/research/mobile-communications/vccs/vienna-5g-simulators/. Accessed Jan 2021
213. Open Air Interface (OAI) (2020). www.openairinterface.org/. Accessed Jan 2021
214. 5G K-Simulator (2020). http://5gopenplatform.org/main. Accessed Jan 2021
215. J. Baek et al., 5G K-simulator of flexible, open, modular (FOM) structure and web-based 5G K-simplatform, in *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)* (2019), pp. 1–4
216. F. Kaltenberger, OpenAirInterface 5G Overview, Installation, Usage. OpenAirInterface Workshop 2019 (2019). https://www.openairinterface.org/docs/workshop/8_Fall2019Workshop-Beijing/Training/2019-12-03-KALTENBERGER-1.pdf. Accessed Jan 2021
217. SCF, 5G FAPI: PHY API Specification (2020). https://scf.io/en/documents/222_5G_FAPI_PHY_API_Specification.php. Accessed Jan 2021
218. F. Kaltenberger et al., OpenAirInterface: democratizing innovation in the 5G era. Comput. Netw. **176** (2020)
219. SCF, 5G nFAPI specifications (2020). https://scf.io/en/documents/225_5G_nFAPI_specifications.php. Accessed Jan 2021
220. Network simulator-3 (2020). https://www.nsnam.org/. Accessed Feb 2021
221. P. Solis, D4.4 Cognitive MAC Simulation, Evaluation and Optimization. 5G-RANGE Research and Innovation Action (2019). http://5g-range.eu/wp-content/uploads/2018/04/D4.4_final.pdf. Accessed Dec 2020
222. N. Patriciello, S. Lagen, B. Bojovic, L. Giupponi, An E2E simulator for 5G NR networks, in *Elsevier Simulation Modelling Practice and Theory (SIMPAT)* (2019)
223. C. Felber, Prototyping wireless systems with NI SDR and open source stacks (2019). https://www.openairinterface.org/docs/workshop/8_Fall2019Workshop-Beijing/Talks/2019-12-05-FELBER.pdf. Accessed Dec 2020
224. R. Gupta et al., NS-3-based real-time emulation of LTE testbed using LabVIEW platform for software defined networking (SDN) in CROWD Project. Association for Computing Machinery (2015).
225. MathWorks, 5G Development with MATLAB. eBook (2020)
226. Mosaic5G, Enabling Agile 4G/5G Service platforms (2018). https://mosaic5g.io/resources/mosaic5g.pdf. Accessed Jan 2021
227. 3GPP TR 28.801, Study on management and orchestration of network slicing for next generation network (2018). https://www.3gpp.org/ftp//Specs/archive/28_series/28.801/. Accessed Feb 2021
228. ONF, Broadband Projects (2021). https://opennetworking.org/onf-broadband-projects/. Accessed Feb 2021
229. ONF, CORD Project (2021). https://opennetworking.org/cord/. Accessed Feb 2021

Chapter 2
Information Flows at the Deep Physical Layer Level

Marco Donald Migliore

2.1 Introduction

5G has introduced many technological advances over previous generations of cellular communication systems. Among these, one of the most relevant is represented by the antennas. In 5G, antennas are no longer relatively simple devices with fixed beams, but become highly reconfigurable systems. The introduction of such sophisticated antennas represents a major step forward in an area that was previously of little interest in cellular systems.

However, 5G antennas are just the starting point of a new trend in personal wireless communication [1–3]. 6G promises a much more aggressive use of the communication channel with an *active control* of it. In this new and revolutionary vision, the communication channel ends up being an uncontrollable bottleneck for the communication system and becomes a resource to be optimized, with a potential disruptive impact on the performance of the communication system.

At the same time, antennas will be highly flexible elements that interact intelligently with the environment, which in turn plays an active role in the communication system [4, 5]. Concepts such as far field and radiation pattern, as well as simple propagation models, lose their ability to model communication channels, and more sophisticated analyses must be undertaken, prompting research on physical aspects [6–8]. This opens up new challenges in studying the interaction between radiating systems and the propagation environment which requires an analysis of the communication process at a finer level of detail than that normally considered in the communication system.

M. D. Migliore (✉)
DIEI, University of Cassino and Souther Lazio, Cassino, Italy
e-mail: mdmiglio@unicas.it

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_2

To clarify this point, let us consider the Open Systems Interconnection (OSI) model (see Fig. 2.1). The OSI model divides the data stream into seven layers starting from the highest level, the "application" layer, down to the lowest level, the "physical layer" (PL), which describes the implementation of bits transmission through the communication channel.

Many exciting new ideas have been proposed at the PL level for 6G. Some of them are based on the use of very high frequencies, f.i. sub terahertz frequencies and visible light, to obtain ultra-broadband communication systems. As in past generations of cellular systems, the improvements are concentrated in the *time channel*.

At the same time, extensive research is underway on innovative radiating systems to obtain a high degree of control of the electromagnetic environment. As mentioned above, the idea is to modify the electromagnetic environment by controlling the space channel. This is a disruptive change of perspective in communication theory, which promises extremely high performance in terms of data rate and number of users served by the network without cost in terms of bandwidth. Radiating systems can be "packaged" in a limited area, or they can be spread over larger areas and work cooperatively, as happens in so-called "user centric cells." All these solutions are focused on the efficient use of the *space channel* by increasing the amount of information that can be reliably transmitted per "unit of space."

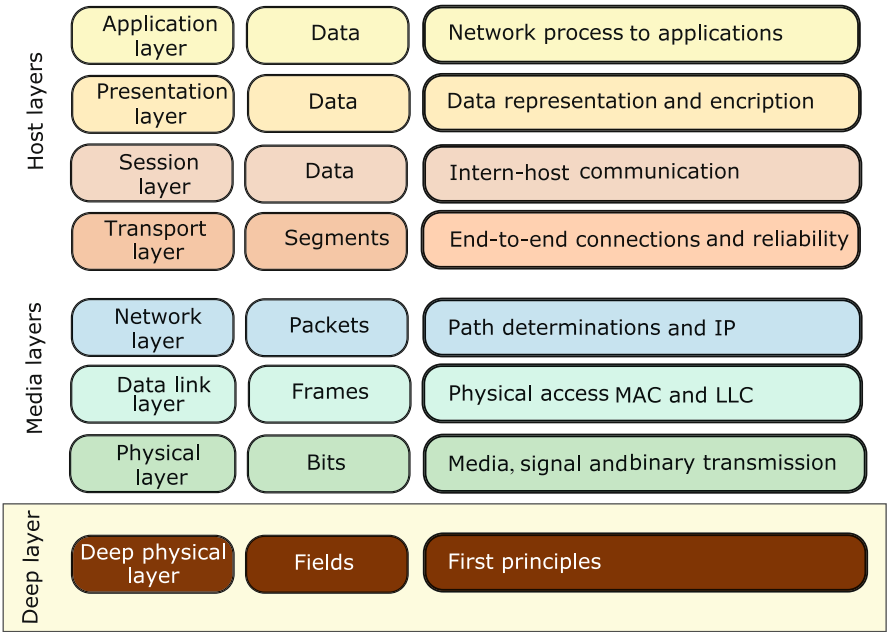


Fig. 2.1 The OSI model extended including the Deep Physical Layer [19]

With reference to the efficient use of the space channel, the different ideas regarding 6G antennas can be grouped in two main research directions: Active Large Apertures (ALAs) and Passive Large Apertures (PLAs). ALAs include a number of technologies such as Massive Very Wide Aperture MIMO, Distributed MIMO, Massive Cell Free MIMO, Radio Stripe, Holographic Massive MIMO, and Orbital Angular Moment Antenna, while PLAs include technologies such as reflective smart surface, reconfigurable reflectors, software-controlled metasurfaces, and smart skins. A short list with some of the main references for each type of technology is shown in Table 2.1. However, it should be emphasized that the literature on the subject is extremely vast and rapidly expanding.

Research at the PL is therefore characterized by a plethora of solutions. Despite the complexity at the PL, there is an element that unifies all these techniques. Basically, all these solutions are different ways of using resources available at a lower, and somehow more fundamental, level than the physical level. The description of the impact of this level in the communication process requires to consider a further layer placed *below* the PL. This additional layer will be called the *Deep Physical Layer* (DPL) [9].

The introduction of the DPL leads to an extended version of the OSI stack in which the communication process is analyzed in terms of distinguishable electromagnetic field configurations on the receiving antenna. In particular, the DPL contains all the resources potentially available to the telecommunication system in terms of space/time/polarization field configurations, i.e., all the possible spatial/temporal/polarization waveforms that can be distinguished at the receiver, while the PL implements specific solutions to effectively exploit the potential offered by the DPL.

Table 2.1 Some directions of research in 6G physical layer

Research directions	Examples
Very high frequency	Millimeter and sub-millimeter communication [49] [50]
Active Large Aperture (ALA)	Very large Massive MIMO [51]
	Cell-free Massive MIMO [52]
	Radio Strips [53]
	Distributed MIMO [54]
	Active Holographic MIMO [54]
	OAM antennas [55]
Passive Large Apertures (PLA)	Holographic beamforming [56]
	Intelligent reflecting surface [57]
	Large intelligent surface [58, 59]
	Reconfigurable reflectarrays [60]
	Reconfigurable Intelligent Surfaces [1]
	Intelligent walls [57]
	Software-controlled metasurfaces [61]
	Smart skins [62]

Clearly, the potential in terms of information transmission present in the DPL is limited only by the physical laws, which are inviolable. Consequently, the limits imposed at the level of the DPL are absolute. In this view, as already noted the DPL is a more fundamental level than the PL.

The results reported in this chapter represent a further contribution in the long path towards a physical approach to communication, started soon after the publication of Claude Shannon's fundamental papers on information theory [10, 11].

This effort has shown interesting connections between information theory and a large number of different fields of physics. In particular, in 1969 Toraldo di Francia introduced the concept of Degrees of Freedom of an optical image [12]. Starting with this article, the optical community has actively studied the connections between optical images and information [12–18].

The work of Toraldo di Francia on optical degrees of freedom was also the starting point for the introduction of the Number of Degrees of Freedom (NDF) in electromagnetism in 1989 [20]. The NDF of the electromagnetic field was introduced in the framework of approximation theory as the minimum number of orthogonal basis functions required to represent the electromagnetic field within a desired error using a linear representation. The paper [20] belongs to a large set of papers devoted to the electromagnetic field approximation theory that includes also efficient bandlimited approximation of the field [21, 22], with a large number of applications in the framework of antenna measurement [23, 24].

The introduction of space multiplexing triggered extensive research on the connections between information theory and electromagnetism in both the communication and the electromagnetic communities [25–30]. The approach followed in a first phase was based on spherical expansion and was consequently limited to spherical sources and spherical observation domains [29, 31, 32]. In 2006, the Degrees of Freedom of the Field were introduced to quantify the amount of information transmitted by sources radiated toward observation domains both having a general shape on the basis of numerical simulations [33] and analytical approach [34]. Successively, Kolmogorov information theory was introduced in 2008 in order to analyze spatial information in case of a small Number of Degrees of Freedom [35]. Limitations regarding different aspects of communication using spatial channels have been further clarified in a number of papers [9, 36–45].

The results reported in the above literature are the backbone of the Deep Physical Layer, introduced in [19] as a common frame for the analysis of information carried by the electromagnetic field. It is worth pointing out that this chapter is focused on the Deep Physical Layer and does not discuss specific technological aspects. Instead, the goal is to identify an upper limit in terms of spatial information that can be reliably transmitted by the electromagnetic field.

A list of the main variables and acronyms used in this chapter is shown in Table 2.2.

Table 2.2 List of variables and acronyms employed

$\mathbf{A}$	Radiation operator
ALA	Active Large Aperture
B	Time-varying signal bandwidth
C_ϵ	Kolmogorov ϵ capacity
CSI	Channel State Information
DPL	Deep Physical Layer
D	Source domain
$\mathbf{E}$	Electric field on the observation domain
$\mathbf{F}$	Reduced electric field on the observation domain
NCSI	No Channel State Information
L	Length of a linear electromagnetic source
NDF	Number of Degrees of Freedom
NDF_ϵ	Number of Degrees of Freedom at ϵ level
MIMO	Multiple Input Multiple Output
$\mathbf{J}$	Current density distribution on the electromagnetic source
PL	Physical Layer
PLA	Passive Large Aperture
PNDF	Polarization Number of Degrees of Freedom
RIS	Reconfigurable Intelligent Surfaces
SNR	Signal Noise Ratio
SISO	Single Input Single Output
SNDF	Space Number of Degrees of Freedom
STPNDF	Space-Time-Polarization Number of Degrees of Freedom
TNDF	Time Number of Degrees of Freedom
$\mathbf{X}$	Set of all the possible current density distribution on the source
$\mathbf{Y}$	Set of all the possible electric field distribution on the observation domain
T	Length of observation time
W	Space-varying signal effective bandwidth
W'	Space-varying signal effective angular bandwidth
c	Speed of the light in vacuum
f	Frequency
f_c	Central frequency
s	Curvilinear abscissa along the observation curve
$\hat{\mathbf{t}}$	Tangent unit vector on the observation curve
u_k	kth left singular function of the radiation operator
v_k	kth right singular function of the radiation operator
x	Functional point in the X set
$\mathbf{x}$	Vector collecting the x_k coefficients
x_k	kth coefficient of the expansion of the current density distribution $\mathbf{J}$
y	Functional point in the Y set
$\mathbf{y}$	Vector collecting the y_k coefficients
w'	Angular spatial bandwidth
$\hat{w}'$	Minimum angular spatial bandwidth assuring that the band limitation representation error of the field tends to zero

(continued)

Table 2.2 (continued)

y_k	kth coefficient of the expansion of the electromagnetic field **E**
$\Delta\xi$	Nyquist spatial step
Ξ	Length of the parameterized observation domain
Ω	Observation domain
β	Free space wavenumber
λ	Free space wavelength
ξ	Observation curve parameterization function in the reduced field representation
ξ_k	Observation curve parameterization function evaluated in $k\pi/w'$
σ_k	kth singular value of the radiation operator
ψ	Phase function along the observation line in the reduced field representation

2.2 Communication at the DPL Using the Space Resource

In this section the communication process will be analyzed using Kolmogorov information theory [35, 46–48]. In short, the waveform that encodes the transmitted information is represented by a point x in a proper functional space. On the receiver side, the received waveform is represented by a point y in a proper functional space. The sets of all possible transmitted and received waveforms are called X and Y, respectively. The communication process maps $x \in X$ to $y \in Y$. The presence of noise is modeled as an open sphere of uncertainty having radius ϵ around the "noiseless" point y. Received waveforms associated with points whose distance is less than ϵ are considered indistinguishable from y. Consequently, the maximum number of distinguishable waveforms on the receiver is given by the maximum number of functional points $y \in Y$ whose distance is larger than 2ϵ. This number is called the *Kolmogorov ϵ-capacity* of the set Y [47] and is the solution of a *sphere packing* problem, which consists in identifying the maximum number of non-overlapping open spheres having radius ϵ.

In order to apply the Kolmogorov theory to our problem, let us consider the model of the communication system at the DPL level drawn in Fig. 2.2. The transmitter is a harmonic electromagnetic source placed in the volume D having a surface area Σ, while the receiver is an "observer" who has access to the value of the electromagnetic field on the observation curve Ω placed at a finite distance from the source D. Information is encoded in spatial variations of the electromagnetic field (Fig. 2.3) obtaining a "pure space communication system" [34].

Let us consider a component of the electric field in the phasor domain. Loosely speaking, the propagation process links the physical quantity associated with information in the transmission process, i.e., the charges in the antenna volume, to the physical quantity associated with the reception, i.e., the electromagnetic field configuration on the receiving antenna generated by the electric charges that are accelerated in the transmitting volume. This physical process is modeled by a compact operator [34, 46, 63, 64], called the "radiation operator":

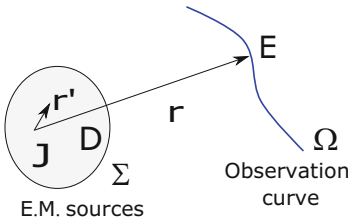


Fig. 2.2 Communication system model; D, electromagnetic source bounded by the surface Σ; Ω, observation domain; J, current density in D; **E**, electromagnetic field on Ω

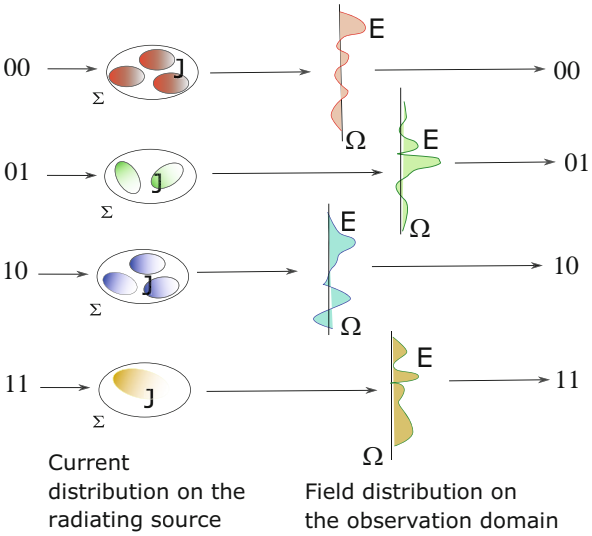


Fig. 2.3 In the "pure space communication system," each information block is encoded in a different current distribution of the source Σ and hence on a different radiated field spatial configuration on the observation domain Ω; the field configurations received on Ω can be considered the "spatial waveforms" used by the communication system

$$E(\mathbf{r}) = \int_D G(\mathbf{r}, \mathbf{r}')\, J(\mathbf{r}')d\mathbf{r}' \tag{2.1}$$

wherein $J(\mathbf{r}')$ is the current distribution in the source domain D, $E(\mathbf{r})$ is the field observed on the domain Ω, and G is Green's function, i.e., the spatial impulse response. We suppose that E and J are square integrable functions. Standard L_2 norm in Hilbert spaces is considered in the following.

Since the energy of the currents is finite, the set of currents is bounded. A compact operator with bounded domain can be diagonalized using the Hilbert-Schmidt expansion [46, 63]. Let $\mathbf{v}(\mathbf{r}') = \{v_1(\mathbf{r}'), v_2(\mathbf{r}'), \ldots, v_k(\mathbf{r}'), ...\}$ and $\mathbf{u}(\mathbf{r}) = \{u_1(\mathbf{r}), u_2(\mathbf{r}), \ldots, u_k(\mathbf{r}), ...\}$ be the left and the right singular basis of the Hilbert-Schmidt expansion. The current distribution on D and the field distribution on Ω can be expanded as:

$$J(\mathbf{r}') = \sum_{k=1}^{\infty} x_k v_k(\mathbf{r}') \tag{2.2}$$

$$E(\mathbf{r}) = \sum_{k=1}^{\infty} y_k u_k(\mathbf{r}) \tag{2.3}$$

wherein x_k and y_k are the expansion coefficients of the currents and of the field distribution. As noted earlier, the energy of the current is limited. For the sake of simplicity, we will suppose $\|\mathbf{x}\| < 1$.

After the diagonalization of the radiation operator, the kth coefficient of the field is just a scaled version of the k^{th} coefficient of the current:

$$y_k = \sigma_k x_k \tag{2.4}$$

wherein σ_k is the $k-th$ singular value of the radiation operator. The singular values are positive quantities sorted in descending way ($\sigma_1 \geq \sigma_2 \geq ...$) and $\sigma_k \to 0$ when $k \to \infty$.

The current distribution $J(\mathbf{r}')$ can be represented by a point x having coordinates $\{x_1, x_2, \ldots, x_k, ...\}$ in an infinite dimensional Hilbert space (Fig. 2.4). Similarly, the field distribution $E(\mathbf{r})$ can be represented by a point y, having coordinates $\{y_1, y_2, \ldots, y_k, ...\}$ in an infinite dimensional Hilbert space (Fig. 2.4). Accordingly, the radiation process maps points of the bounded spherical set X of all the possible current distributions on the source to points of the set Y that turns out to be a hyperellipsoid whose main kth semi-axis is σ_k long (Fig. 2.4).

Now, let us come back to the communication problem.

Information is encoded in different values of the x_k coefficients (Fig. 2.4) and is retrieved observing the field distribution on Ω, i.e., the values of the y_k coefficients.

However, following the Kolmogorov approach, due to the presence of noise and measurement uncertainties, the received field is

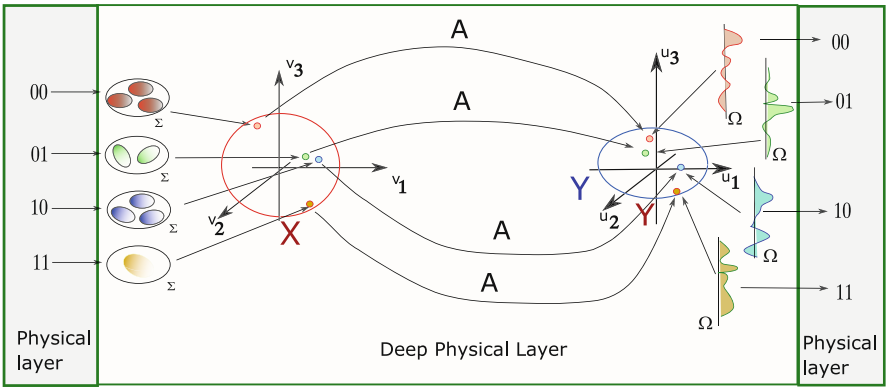


Fig. 2.4 Scheme of a functional communication system; **A** is the radiation operator, X is an infinite-dimensional ball having unit radius, Y is an almost finite-dimensional hyperellipsoid having effective NDF_ϵ dimensions and whose semi-axes have length equal to σ_k; the receiver divides Y in subsets given by ϵ-balls centered on an ϵ distinguishable set

$$E' = \sum_{k=1}^{\infty} \sigma_k x_k u_k + \epsilon \tag{2.5}$$

Intuitively, since $\|x_k\| < 1$, if $\sigma_k < \epsilon$, we have $\sigma_k \|x_k\| < \epsilon$ for any value of x_k, making it impossible to transmit information using the k^{th} coefficient of the expansion.

We define "Number of Degrees of Freedom of the field at the ϵ level" (NDF_ϵ) the number of singular values higher than ϵ [34].

It is worth noting that the definition of the Number of Degrees of Freedom of the field at ϵ level of uncertainty adopted in our discussion is different compared to the NDF defined by Bucci and Franceschetti in [20]. The problem discussed in [20] regards the approximation of the electromagnetic field, and its solution involves a covering problem [35] (i.e., the minimum number of balls having radius ϵ required to cover the Y set), instead of a packing problem. The dimension of the effective spaces in the two cases differs by a singular value. Furthermore, Bucci's definition of NDF does not explicitly consider the noise level since Bucci's theory was developed mainly in the case of free-space propagation in radiative region, where the singular values gently decrease until reaching a value, e.g., $\hat{\sigma}_k$ after which there is an extremely rapid decrease toward zero. In this case the number of singular values that must be retained is almost independent of the noise level and equal to $\hat{\sigma}_k + 1$ in Bucci definition. Obviously, under the same conditions, also the NDF_ϵ is almost independent of ϵ. However, in presence of scatterers in the environment, the number of singular values that must be taken can be quite sensitive toward the noise level. This sensitivity is particularly high if the observation domain is in the reactive region of the source [35]. It is also interesting to note that when the electrical dimension of the source tends to infinity, the knee of the singular values becomes a step and the value of the NDF_ϵ as defined in this chapter coincides with the value of the NDF by Bucci.

The maximum number of field configurations that can be distinguished on Ω is given by the maximum number of points on Y whose distance is larger than ϵ [46, 47]. Such a set of points will be called Y_ϵ, and, as previously discussed, is the solution of a "sphere-packing" problem, since it requires the identification of the maximum number of non-overlapping ϵ-balls (i.e., balls having ϵ radius) whose centers are elements of Y. Accordingly:

the maximum amount of bits that can be transmitted in presence of an ϵ level of uncertainty is the *the Kolmogorov ϵ-capacity of Y* [35].

The geometric structure of the functional spaces involved in the propagation model has an important impact on the characteristics of the communication process seen at the level of the PL.

For example, the parallel spatial communication subchannels that arise in MIMO communication systems [65] are a consequence of the separability of the Hilbert space. In particular, each coefficient y_k of y is able to encode independent information, giving a spatial communication subchannel at the PL level [35].

However, the number of spatial subchannels is limited by the number of coefficients that can be retrieved in presence of noise. Equation 2.5 shows that there

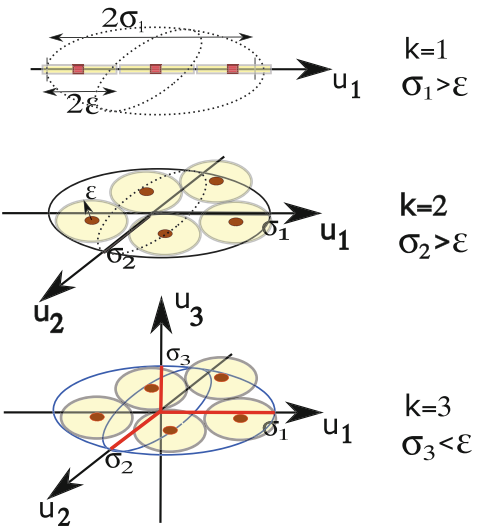


Fig. 2.5 An intuitive explanation of the meaning of NDF_ϵ; when k increases we consider more dimensions of the hyperellipsoid Y; we can add more balls adding dimensions until $\sigma_k > \epsilon$; the NDF_ϵ is the largest k value such that $\sigma_k > \epsilon$; in the example shown in the figure, $NDF_\epsilon = 2$

is a value k', such that $\sigma_{k'}$ becomes smaller than ϵ making it impossible to add further spheres (see Fig. 2.5). Since σ_k is a decreasing sequence of positive numbers, all the dimensions $k > k'$ do not provide any contribution to the packing process. This intuitive explanation has its rigorous mathematical counterpart in the theory of Kolmogorov information theory and in the n-widths theory [35, 66].

Accordingly, the number of parallel spatial communication subchannels at the PL is limited by the NDF_ϵ at the DPL.

Summarizing, the field can be approximated by the first NDF_ϵ basis functions [9]

$$E(\mathbf{r}) \simeq \sum_{k=1}^{NDF_\epsilon} \sigma_k x_k u_k(\mathbf{r}) \tag{2.6}$$

while the ϵ-channel capacity turns out to be [9]

$$C_\epsilon(Y) \approxeq \sum_{k=1}^{NDF_\epsilon} \log_2\left(1 + \frac{\sigma_k}{\epsilon}\right) \tag{2.7}$$

2.3 Bandlimited Approximation in the Space Domain

The analysis carried out in the above section is based on the Kolmogorov approach instead of Shannon's one usually adopted in Information Theory. The reason is related to a slight asymmetry between information in time domain and in space domain.

Loosely speaking, in the time domain, the information is associated with different time-varying waveforms chosen so that they are distinguishable by the receiver even in the presence of noise. The bandwidth of the communication channels limits the set of waveforms that can be received. In practice, the bandlimiting operator plays a role equivalent to that played by the propagation operator in limiting the set of space-varying waveforms.

Like the radiation operator, also the bandlimiting operator is a compact operator [67] whose singular values are almost constant up to a value after which an extremely rapid decrease occurs [68]. The number of singular values at the knee is called the Number of Degrees of Freedom of the (time domain) signal, and in the following it will be denoted as $TNDF$. In case of a signal having effective bandwidth B and observed for a time interval T, we have that the $TNDF$ is given by the time-bandwidth product, i.e., $TNDF \simeq 2BT$. A more accurate and rigorous analysis is carried out in [67]. Since the time is unlimited, it is possible to *arbitrarily increase the value of* $TNDF$ by sending waveforms of arbitrarily long duration. In Shannon's theory, this allows to exploit the properties of high-dimensional spaces in terms of energy concentration, transforming "soft spheres" of uncertainty into "hard spheres" of uncertainty [69]. Consequently, the ability to send arbitrarily long waveforms (i.e., arbitrarily long codes) is a key factor in Shannon's theory.

The analysis of the degrees of freedom of the electromagnetic field basically shows that the NDF of space-varying waveforms used to encode information is limited by the effective rank of the radiation operator. Consequently, from a mathematical point of view, it is not possible to apply the law of large numbers to transform "soft spheres" of probability to "hard spheres" in "pure" space coding communication. Kolmogorov theory overcomes this problem.

When the NDF of the field can be arbitrarily increased, as happens in Massive MIMO [70] wherein arbitrarily electrically[1] large antennas are considered, we can apply the statistical approach used in Shannon's theory considering arbitrarily long data block transmitted in the space domain.

Shannon theory in case of bandlimited channels is a well-developed and consolidated field of research. Consequently, if we are able to obtain a bandlimited approximation of the field in space, we can extend to the space domain the results available in literature regarding communication in the time domain.

This is certainly possible. In fact, the set of bandlimited functions with bandwidth w is dense for $w \to \infty$ in the set of the fields on the receiving domain [21]. Accordingly, it is possible to obtain a representation of the field within any degree of accuracy using bandlimited functions.

The sampling representation of the electromagnetic field is routinely used, for example, in planar near-field antenna measurement systems, wherein a Nyquist distance $\simeq \lambda/2$ is usually adopted under the hypothesis that the reactive energy of the plane-wave spectrum is negligible [71].

[1] The electrical dimension is the dimension measured in wavelengths.

Such a simple "a priori information" on the radiated field gives a number of sampling positions much larger than the NDF of the field. This is a clear indication that the sampling representation is redundant. While in antenna measurements the oversampling of the representation is acceptable (and in some cases advantageous since the goal is the interpolation of the data, and oversampling reduces the effect of noise in measured data), in spatial communication it gives correlation between the samples. Loosely speaking, it is not possible to associate independent information to each spatial sample.

Consequently, we are interested in a sampling representation of the field at an ϵ level of approximation with a number of samples equal to the NDF_ϵ of the field. In this case it is possible to associate independent information to each spatial sample, obtaining the maximum number of parallel spatial subchannels.

However, this is generally not possible, and bandlimited approximations of the field require a number of samples larger than NDF_ϵ at an ϵ level of approximation. In fact, the optimal basis, given in Eq. 2.4, generally requires quite complicated functions that depend on the geometry of both the radiating and the receiving domain, while bandlimited approximation uses basis functions that do not depend on the geometry of the problem.

Clearly, since we have no flexibility in the choice of basis functions, the only way to improve the bandlimited representation is to include source and observation domains in the sampling step. This solution has been proposed in a series of elegant papers by Bucci, Franceschetti, and co-workers [20–22]. The papers are very accurate and complete from a mathematical point of view, and the interested reader is invited to refer to these articles to go into the details of the theory.

However, the papers do not focus on the physical processes that underlie the results of the theory. In the following, some key points of the theory will be recalled with the aim of clarifying these aspects.

As preliminary step, let us consider the field on an observation curve Ω. The phase of the field in a point depends on two effects. The first one is the value of the current distribution on the source. Note that information is encoded in the current distribution. Consequently, this effect must be "preserved" in the field representation. The second one depends on the delay of propagation and hence on the geometry of the source and on the position of the observation point.

Let us focus our attention on the consequences of the delay, whose effect in the harmonic domain is a phase shift.

The phase variation of the field on Ω depends on the geometry of the source and of the observation curve. Since information is not encoded in the shape of the source and of the observation curve, the effect of the geometry can be subtracted from the phase of the field without losing information.

There is a further effect of the phase that is relevant in a bandlimited representation of the field. To clarify this effect, let us consider a linear source and a linear observation domain parallel to the linear source. The dimension of the source seen from a point along the observation line decreases when the point moves away from the position in front of the source. Intuitively, the source seen from the position of the observation point becomes "smaller," and its ability to transmit information

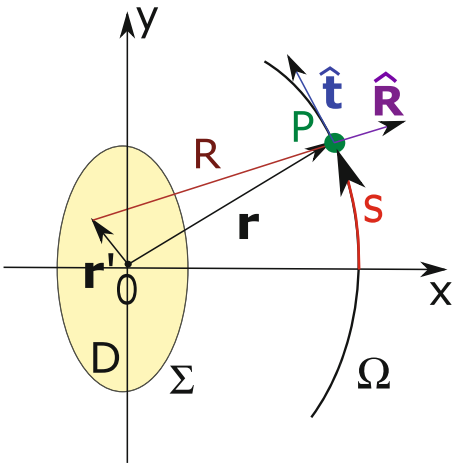


Fig. 2.6 Relevant for the bandlimited representation of the field; D is the domain containing the electromagnetic sources, bounded by the surface Σ; Ω is the observation curve; $P(s)$ is the position of the observation point, wherein s is the curvilinear abscissa along the curve Ω; $\hat{\mathbf{t}}$ is the tangent to the curve Ω in the observation point

tends to decrease. This suggests that the information is not evenly distributed along the observation domain, but is more concentrated in front of the source. The relevant consequence is that the Nyquist distance tends to change along the line of observation.

In accordance with the intuitive discussion outlined above, and following the reference [22], instead of directly considering the field on the observation curve, we consider the *reduced* field F [19, 22]:

$$F(\xi) = E(\xi)e^{j\psi(\xi)} \tag{2.8}$$

in which we introduce the parameterization function along the observation curve $\xi = \xi(\mathbf{r})$, and the phase function $\psi(\xi) = \psi(\boldsymbol{\xi}(\mathbf{r}))$ (see Fig. 2.6). As discussed above, these functions take into account the effect of the geometry. The parameterization function changes the distance between two points on the curve, while the phase function allows to subtract the phase specified by ψ from the field.

Our final goal is the identification of ξ and ψ functions that give the best bandlimited approximation of the field. Mathematically, this requires to identify the ξ and ψ functions that minimize the distance between the reduced field F and the bandlimited approximation of the reduced field F_w, wherein w is the bandwidth.

The first step is the evaluation of the spatial bandlimited version of the reduced field [19, 21]:

$$\begin{aligned} F'_w(\xi) &= \frac{1}{\pi}\int_D \left(\frac{\sin(w(\xi-\xi'))}{\xi-\xi'}\hat{G}(\xi',\mathbf{r}')d\xi'\right)\cdot J(\mathbf{r}')d\mathbf{r}' \\ &= \int_D \hat{G}'_w(\xi,\mathbf{r}')\cdot J(\mathbf{r}')d\mathbf{r}' \end{aligned} \tag{2.9}$$

wherein w' is the angular spatial bandwidth,[2] $\hat{G}$ is the reduced Green's function [22] and $\hat{G}_w$ is the bandlimited approximation of $\hat{G}$

This formula deserves some comments from a physical point of view. In particular, it shows that the bandlimited property of the field is directly related to a bandlimited property of Green's function. Indeed, Green's function acts as a "spatial filter" that limits the amount of information that can be collected on the observation domain.

The band limitation error on $\hat{G}$ for large observation distance is [19, 22]:

$$\Delta\hat{G}(\xi, \mathbf{r}') = \hat{G}_{w'}(\xi, \mathbf{r}') - \hat{G}(\xi, \mathbf{r}') \tag{2.10}$$

and involves integrals whose kernel includes exponential functions with imaginary argument:

$$\exp[j(\gamma(\xi', \mathbf{r}') \pm w'(\xi - \xi'))] \tag{2.11}$$

wherein $\gamma(\xi', \mathbf{r}') = \psi(\xi') - \beta R(\xi', \mathbf{r}')$.

These kinds of integrals are quite common in electromagnetic diffraction theory and are solved using the method of stationary phase [63]. Asymptotically (i.e., increasing the distance between source and observation domains), the integral does not vanish only for ξ' values close to the stationary point of the argument of the exponential function. As a relevant consequence, if there are *no stationary points*, $\Delta\hat{G}$ is (asymptotically) zero, i.e., $\hat{G}_{w'}(\xi, \mathbf{r}') \rightarrow \hat{G}(\xi, \mathbf{r}')$, and the approximation error tends to zero. The presence of stationary points depends on the w' value. For any ξ, there is a minimum value $\bar{w}'$ of w' that assures the absence of stationary points. Any bandwidth value not smaller than $\bar{w}'$ assures an approximation error tending to zero.

As noted above, the value of w' that assures the absence of stationary points depends on ξ. The function $\bar{w}'(\xi)$ is consequently the "local bandwidth" around ξ, while its maximum, $W' = \max_\xi \bar{w}'(\xi)$, is called the "effective bandwidth" of the field [22].

Since the local bandwidth changes along the observation curve, the Nyquist step also changes. To simplify the representation, and to have a constant Nyquist step, it is advantageous to choose the ξ function such that the local bandwidth has a constant value, f.i. W', along the observation curve, imposing [22]:

$$\bar{w}'(\xi) = \beta\Delta(s)\frac{ds}{d\xi} = W' \tag{2.12}$$

[2] w' is bandwidth in the angular spatial frequency. The angular spatial frequency is defined in this chapter as 2π times the spatial frequency. The relationship between these two quantities is the same relationship that exists between angular frequency and frequency in time domain signals. In electromagnetic theory the natural choice is the angular spatial frequency, which is directly related to the physical concept of wave number, while in Information Theory the frequency is usually preferred.

wherein (see Fig. 2.6):

$$\Delta(s) = \frac{\max_{\mathbf{r}'} \hat{\mathbf{R}} \cdot \hat{\mathbf{t}} - \min_{\mathbf{r}'} \hat{\mathbf{R}} \cdot \hat{\mathbf{t}}}{2} \tag{2.13}$$

$\hat{\mathbf{t}}$ is the unit vector tangent to the curve in the point $P(\xi)$, and s is the curvilinear abscissa in the P point. This choice gives a constant sampling step along the observation curve equal to:

$$\Delta\xi = \frac{\pi}{W'} \tag{2.14}$$

The expression of ξ and ψ functions and of the effective bandwidth are reported in [22] for the main source geometries and receiving manifold (including planar, cylindrical, and spherical).

According to the bandlimited function theory, the number of (spatial) samples is given by the space-spatial bandwidth product, i.e., $W'\Xi/\pi$, wherein Ξ is the extension of the observation domain in the ξ variable and W' is called "the effective angular spatial bandwidth of the field" [21].

Equation 2.13 includes the role of the geometry of the source and of the observation domain in the field representation in a simple and elegant formula. In spite of the relative simplicity, this formula allows to subtract the effect of the geometry around the observation point with an amazing efficiency, giving a number of spatial samples almost equal to NDF_ϵ.

It is understood that thanks to the choice of the ξ parameterization the sampling points are placed at a constant distance (equal to $\Delta\xi = \pi/W'$) along ξ, but are *nonuniformly placed* along the curvilinear abscissa s.

In summary, the reduced field admits a bandlimited approximation with an angular bandwidth practically equal to W' and a Nyquist step along the observation curve which is constant in the parameterization $\xi(s)$ but non-constant in the curvilinear abscissa s.

The common application of the bandlimited representation of the functions is the interpolation from the samples. Strictly speaking, in our case prolate spheroidal functions must be used as interpolation functions [67], since $sinc$ functions are orthogonal only on unlimited domains, while the observation domain in ξ is always finite. The solution followed by Bucci and co-workers is to enlarge the bandwidth [22], accepting a slight oversampling able to reduce the truncation error to a negligible value.

However, the problem discussed in this chapter differs from standard interpolation. As previously mentioned, the problem studied by Bucci essentially concerns a *sphere covering problem*, i.e., the minimum number of spheres that allow to cover the set Y, and accepts a degree of overlap among the spheres. This is not acceptable in our case, which concerns a *sphere packing* problem. Consequently, we do not introduce oversampling, and the number of samples in the space is fixed directly by the bandwidth W'.

It is also interesting to note that Eq. 2.12 shows a close relationship between $\xi(s)$ and W'. Indeed, in Eq. 2.12 the value of W' can be chosen at our will. Hence, the spatial bandwidth can assume an arbitrary value. Indeed, the fundamental quantity that characterizes the set of waveforms is the NDF of the bandlimited functions. From this point of view, the bandwidth is the NDF per unit observation interval, and $\xi(s)$ is scaled by the value chosen for W'. For example, in case of time-varying waveforms, the numerical value of the bandwidth B is fixed by the amount of unit time, which is the second in MKSA system of units. In the space domain, the use of a unit quantity dividing the space in uniform steps, as meters or degrees in case of observation angles, is not appropriate since the degrees of freedom are not evenly distributed in the space. A common choice is to scale ξ such that ξ ranges from $-\pi$ to π in all the space, but other intervals can be chosen. The numerical value of W' will be fixed by the specific choice.

Despite the mathematical complexity, the ψ and ξ curves have a simple intuitive explanation. As previously discussed, with reference to the phase of the electromagnetic field at an observation point, its value depends on two different phenomena. The first one relates to information that is encoded in some specific values of the phase of the radiated field. The second one is due to the propagation delay. This second phenomenon does not convey information and must be "discarded" in the sampling representation.

With reference to Fig. 2.7, let us consider two radiating point sources, the source S_a and the source S_b, and an observation point P having distance r_a and r_b from S_a and S_b, respectively. The signals radiated by the two sources reach P with a delay r_a/c and r_b/c, c being the speed of light in vacuum. The average delay in P will be $(r_a + r_b)/(2c)$, while the average difference between the delays is $(r_a - r_b)/2c$. In harmonic regime the delay gives a phase shifting. The average phase in P is $\beta(r_a + r_b)/2$, while the average phase difference is $\beta(r_a - r_b)/2$. The curves on which the average phase is constant are ellipses having foci in S_a and S_b, while the curves on which the average phase difference is constant are hyperbolas having foci in S_a and S_b. The hyperbolas that intersect the source at a distance of $\lambda/2$ from each other give a phase difference between two consecutive curves equal to π. Put simply, if we move P along an ellipse, the field has a phase shift of π each time P

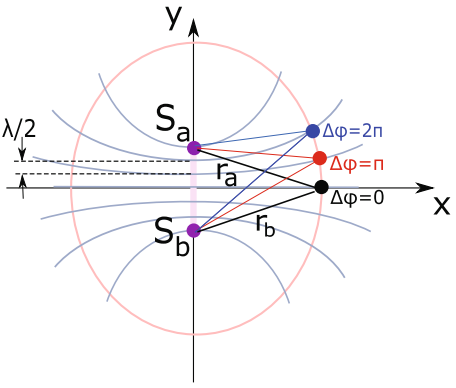


Fig. 2.7 Relevant for the spatial sampling of the field; the curves $\xi = k\pi/W'$ are hyperbolas having foci on the extreme points of the point source S_a and S_b (points "a" and "b"); the difference of phase between the waves radiated by the two extreme points is constant along the $\xi = constant$ curves and is equal to π between two consecutive $\xi_k = k\pi/W'$ curves

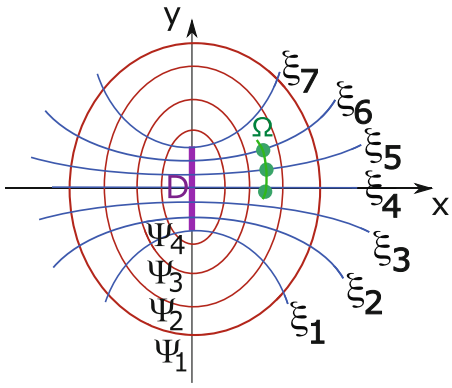


Fig. 2.8 Relevant for the sampling of the field; the radiating line is drawn in violet in the center of the figure; the observation line is drawn in green; the curves at $\xi = \xi_k$ are drawn in blue; the curves $\psi = \psi_n$ are drawn in red; the sampling points on the observation curve B are given by the intersection points between the blue curves and the green curve (green points)

passes over a hyperbola. It is useful to note that the time delay difference evaluated by considering any other pair of points along the segment connecting S_a and S_b changes more slowly in the space and hence would give a slower phase variation in the space.

Now, let us consider a continuous source. In Fig. 2.8 a linear source D (violet thick line) radiates in free space. The field is observed on the observation curve Ω (plotted as thick green curve). This geometry has been analyzed in [22, 72]. The curves $\psi = constant$ are ellipses whose foci are the two extreme points of the source. They are drawn as red curves in the figure. The curves $\xi = \xi_k = k\pi/W'$, for $k = 1, \ldots, SNDF$ are hyperbolas having foci at the extreme points of the source and intersecting the source at a distance of $\lambda/2$ from each other [22, 72]. This gives an intuitive explanation of the spatial sampling procedure. As first step, the average phase associated with the average delay of the field radiated at the extreme points is subtracted at each point of the observation curve. As second step, we collect a sample each time the observation curve intersects a ξ_k curve (green points). These curves are related to the time delay difference evaluated by extreme points of the source and therefore give the fastest phase variation in the space associated with the time delay.

This process "filters out" the artifacts due to the geometry and gives us a simple picture of how information flows from the DPL to the PL. For example, at instant t_k the PL receives the request to send the string "1 0 0" to a receiver that has access to the value of the field on the Ω curve (see Fig. 2.9). The string "1 0 0" is encoded in a current distribution J such that the field received in the three sampling positions takes on different values (amplitude and/or phase) depending on the corresponding bit. In this way it is possible to send 3 bits at the same time. From the PL point of view, we have three parallel communication sub-channels, i.e., a three-channel MIMO communication system.

The maximum number of spatial communication channels is consequently limited by the NDF_ϵ, or, in the spatial bandlimited approach, by the spatial bandwidth and the extension of the observation curve. This number is approximately the number of Nyquist spatial intervals on the observation curve, i.e., $\simeq W'\Xi/\pi = 2W\Xi$ [34], wherein Ξ is the extension of the observation domain measured in ξ, and $W = W'/(2\pi)$ is the spatial bandwidth [34].

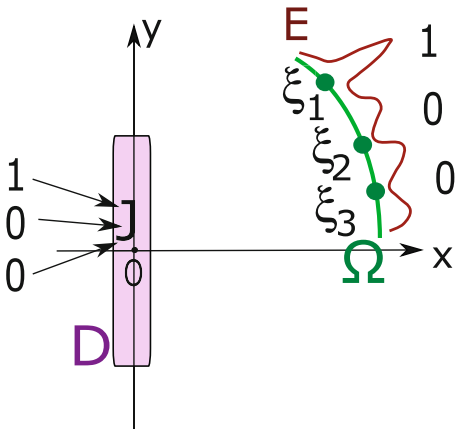


Fig. 2.9 Each spatial sample can be associated with an independent information; up to $NDF_{\epsilon} \simeq 2W\Xi$ information can be transmitted using the space channel, each giving a potential parallel communication MIMO spatial channel at the Physical Layer level

As further example, one of the emerging technologies is Massive MIMO [70], in which antennas with a large number of elements are used to connect significantly fewer users than the number of radiating elements. An important characteristic of Massive MIMO is that it is possible to obtain decorrelation among the signals on the users in non-cooperation condition. The field radiated by MU-MIMO antennas can be modeled as a wide-bandwidth spatial signal that is largely subsampled since the sample positions (i.e., the positions of the users) are placed at distances much greater than the Nyquist spatial step. In a scenario where the number of users is fixed, the greater the spatial bandwidth the greater the decorrelation between the users.

It is understood that the geometrical approach outlined in this section can be extended to include scattering objects, making it possible to apply the theory, for example, to smart controllable surfaces [19].

Hereinafter the Number of Degrees of Freedom associated with the *spatial* variation of a harmonic field will be referred to as SNDF (Spatial Number of Degrees of Freedom).

At the end of this paragraph, it is impossible not to notice that in the DPL spatial information is linked to one of the most fundamental properties of our universe, the finite speed of light. Consequently, with reference to the description of a communication system at a macroscopic level, there are no layers with more fundamental limits than the DPL.

2.4 The Space-Time Polarization Number of Degrees of Freedom (STPNDF)

In previous section the analysis was focused on the spatial channel. Let us consider now the role of time and polarization in the communication process.

Time domain representation was discussed at the beginning of the previous section and will not be repeated. Let us just remember that in case of a signal having effective bandwidth B and observed for a time interval T, we have that the $TNDF$ is given by the time-bandwidth product, i.e., $TNDF \simeq 2BT$, wherein B is the bandwidth and T the observation interval.

In previous section it was also pointed out that there are strong similarities between temporal and spatial communication. Basically, the main difference between the use of temporal and spatial resources in a communication system is the physics behind the set of waveforms that can be used to encode information, which provides stricter limitations for spatial waveforms.

In this perspective, the theorems of electromagnetism can provide us with important information on these limitations. For example, Maxwell's equations involve the vectors **E** and **H**, i.e., six scalar components. However, the uniqueness theorem states that the configuration of the electromagnetic field outside any closed surface surrounding the sources and radiating in free space is uniquely defined by only *two* components of the electric *or* magnetic field on the surface [63]. In terms of information transmission, the uniqueness theorem is a rather tight bottleneck. Even if we have six possible quantities for encoding information, only two of them are able to carry independent information. In other words, it states that polarization increases the Number of Degrees of Freedom available by a factor of up to 2. In the following the Number of Degrees of Freedom associated with polarization will be called $PNDF$.

The uniqueness theorem is a consequence of some subtle relations present in Maxwell's equations in free space: the divergence equations impose a dependence between the components of the electric field and the components of the magnetic field so that one of the three components of **E** and of **H** vectors is related to the other two, while the curl equations impose a dependence between **E** and **H**.

The uniqueness theorem states also that the maximum amount of information can be collected on a surface surrounding the source, and collecting data on multiple surfaces or in volumes does not offer improvements.

However, in practical cases the observation domain Ω does not surround the entire source, and it is often quite small. In this case the estimation of the degrees of freedom of the field is not easy. Loosely speaking, when we consider the full functional dependence of the communication process at the DPL, the Hilbert-Schmidt decomposition is applied to the full electric and magnetic vector fields $(\mathbf{e}(\mathbf{r}, t), \mathbf{h}(\mathbf{r}, t))$ observed on the surface Ω. The number of singular values above the uncertainty level is the *Space-Time-Polarization Number of Degrees of Freedom of the Field* (STPNDF).

In the case of a small observation domain (e.g., using antennas whose size does not exceed a couple of wavelengths), the STPNDF is small and the singular values of the radiation operator tend to decrease more slowly than for large antennas. In this case it is possible to get a couple of "extra" degrees of freedom since we have a couple of singular values above the noise which we would not expect in case of large antennas. These extra degrees of freedom allow to increase the number of

spatial subchannels at the PL level. The effect of "extra" degrees of freedom can be significant in the case of small electrical antennas [73, 74].

In order to obtain an estimation of the maximum STPNDF of a communication channel, let us consider an observation surface covering the entire sphere surrounding the source Σ. In a very narrow bandwidth df, the $STNDF$ (without considering polarization) is $\simeq 2Tdf2W(f)\Xi(f)$, wherein T is the observation interval, $W(f)$ is the spatial bandwidth, and $\Xi(f)$ is the extension of the observation domain. Note that W and Ξ, and hence the $STNDF$, depend on the frequency. In particular, the $SNDF$ is proportional to the electric area of the surface Σ of the electromagnetic source, i.e.,$\simeq A/(\lambda^2/4)$[22].

Consequently, an upper bound for the STPNDF in case of an electromagnetic field carrying passband modulated signals ($f_l \leq f \leq f_u$) with bandwidth $B = f_u - f_l$ is [19]:

$$\int_{f_l}^{f_u} \left[2\left(4\frac{A}{c^2}f^2\right)(2Tdf)\right] = \frac{16}{3}\frac{A}{c^2}TB\left(4f_c^2 - f_l f_u\right)$$
$$= 2(2BT)\left[\left(\frac{4A}{c^2}f_c^2\right)\left(1+\frac{1}{12}\left(\frac{B}{f_c}\right)^2\right)\right] \quad (2.15)$$

wherein $f_c = (f_l + f_u)/2$ is the central frequency, the factor 2 at the beginning of the integrand is the Polarization Number of Degrees of Freedom, and c is the speed of light in the vacuum.

The STPNDF is a fundamental quantity that limits the maximum performance of the system. For example, supposing that the bits transmissible per degree of freedom is $\log_2(1 + SNR)/2$, SNR being the Signal to Noise Ratio at the receiver, the capacity at the DPL in case of No Channel State Information (NCSI), i.e., if the receiver has no feedback information on the channel by the transmitter, turns out to be[3] [19]:

$$C = 8\frac{A}{c^2}f_c^2 B\left(1+\frac{1}{12}\left(\frac{B}{f_c}\right)^2\right)\log_2(1+SNR) \quad (2.16)$$

It is useful to note that any mathematical processing performed at the higher layer of the communication system cannot increase the amount of information set at the lower layer [75].

Accordingly, the maximum amount of information that can be reliably transmitted using the electromagnetic field is fixed in the Deep Physical Layer.

Formula (2.15) shows that in case of a radiating surface the SNDF grows according to two terms, the first increasing with the square of the central frequency

[3] It is understood that in case of Channel State Information (CSI), it is possible to distribute the available power among the subchannels according to the waterfilling algorithm in order to maximize the mutual information [75].

f_c and the second increasing with the square of the bandwidth B. From a practical point of view, Eq. 2.15 suggests that there are two main ways to increase the SNDF: increasing frequency and increasing fractional bandwidth. In particular, for a fixed central frequency, an enlargement of the bandwidth B gives, besides a linear increase of the first addend, a cubic increase of the second addend. This observation could be useful in case of 6G sub-6 GHz band communication systems that will be forced to use the space domain in a more aggressive way compared to the ones using millimeter and THz frequencies [76].

2.5 Control of Information Flows at the DPL Level

The channel operator plays a key role in limiting the amount of information transmissible along the wireless channel. This suggests to modify the channel operator using devices capable to manipulate the electromagnetic field. We will call these devices *Electromagnetic field Processing Devices* (EPDs). Loosely speaking, EPDs modify the shape of the set Y of the fields received on the observation domain.

Consider, for example, the two-channel MIMO communication system with BPSK modulation shown in Fig. 2.10. The channel operator maps the set of currents X into a highly elongated set Y. Due to the geometry of Y, there is an overlapping of the spheres of uncertainty which prevents the correct reception of the 2 bits. Any attempt to use two spatial channels results in incorrect data reception.

The EPD is placed between the TX and RX (Fig. 2.11) modifying the TX-RX radiation operator and consequently the geometry of set Y. In practice, the EPD collects information before it is lost and retransmits it, as shown in Fig. 2.11.

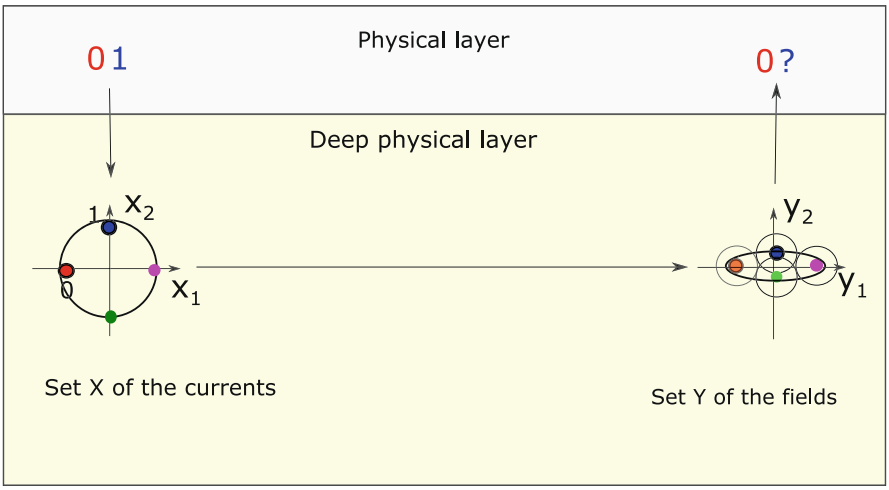


Fig. 2.10 Due to the overlapping of the sphere of uncertainty, the ϵ-channel capacity is equal to 1: any attempt to transmit 2 bits on the spatial channel results in incorrect data reception

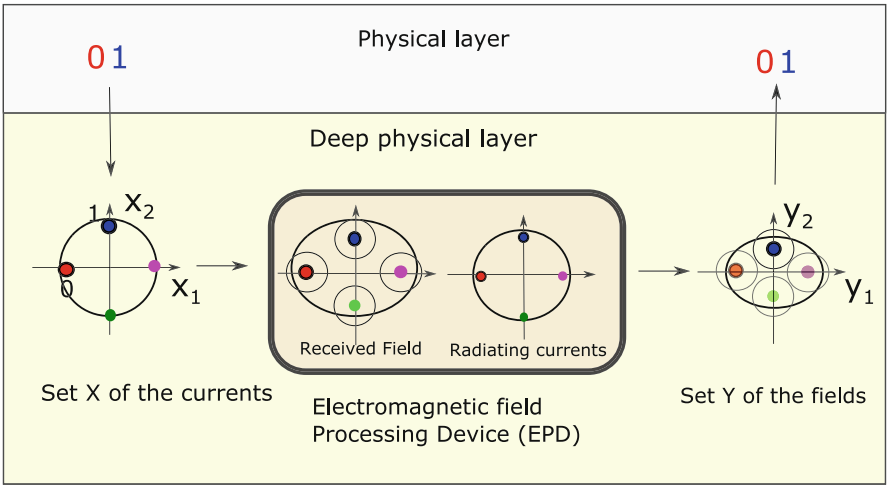


Fig. 2.11 The use of an EPD allows to preserve information obtaining an ϵ-channel capacity equal to 2

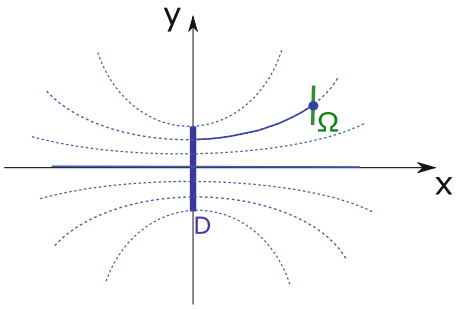


Fig. 2.12 Example of the spatial information flow in absence of EPD; only one ξ_k curve intersects the observation domain Ω, providing only a single spatial channel

The ξ_k curves introduced in Sect. 2.3 allow a simple and intuitive visualization of the information flow. Let us consider the communication problem shown in Fig. 2.12, wherein D is the radiating source and Ω is the observation domain. The curves ξ_k of the source D are drawn in blue. Only one ξ_k curve intersects the observation domain (drawn as solid blue line). Accordingly, we have only one SNDF on Ω and only one spatial subchannel in the PL.

In Fig. 2.13 an EPD is placed between the transmitter and the receiver. The ξ_k curves of the EPD are plotted in red. Two ξ_k curves of D (drawn as solid blue curves) intersect the EPD, indicating that the EPD is able to receive two spatial subchannels. Furthermore, two curves ξ_k of the EPD (drawn as solid red curves) intersect the observation domain Ω, indicating that it is possible to have two spatial subchannels.

However, the geometrical representation does not give a complete view of the problem, since it does not give information on the amplitude of the singular values associated with the channels.

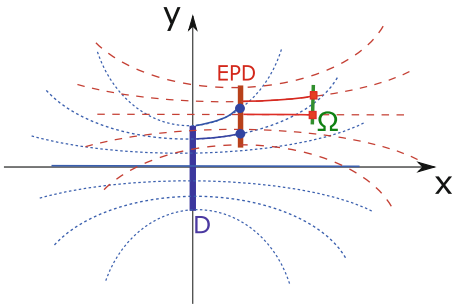


Fig. 2.13 Example of the spatial information flow in presence of EPD; two ξ_k curves intersect the EPD from the source D, providing two spatial channels for D-EPD communication; furthermore, two curves intersect the observation domain Ω from the EPD, providing two spatial channels for EPD-Ω communication; consequently, two spatial channels are available for D-Ω communication

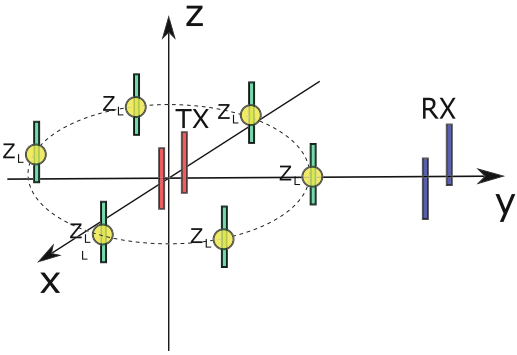


Fig. 2.14 Proof of concept: a number of scatting wire antennas closed on electronically controllable loads are placed around a transmitting MIMO antenna; the loads of the elements are selected in order to optimize the channel capacity of the MIMO system [77]

Further information can be obtained by the analysis of results obtained by ADAM antenna [77], a prototype built in 2006. ADAM consists of a standard MIMO antenna with two radiating elements surrounded by a 6 bits EPD made using wire antennas closed on electronically controllable loads (Fig. 2.14). In this way we have a controllable scattering environment around the antenna. ADAM was tested in a 2×2 MIMO communication system working at 2.5 GHz (center frequency) transmitting 30 tones with 1 MHz spacing [78]. The results were compared with the results obtained using the same communication system in the same environment but without the EPD.

The singular values in case of absence of EPD are plotted in Fig. 2.15 as blue circles. The figure shows that only 30 singular values associated with the 30 tones of the first spatial subchannel are relevant, while the remaining are below -20 dB.

The use of EPD gives the singular value distribution drawn in the same figure as blue circles. The singular values are normalized to the first singular value in absence of EPD to simplify the comparison. The figure shows an increase of the amplitude of all the singular values. In particular, in presence of modest SNR ratio, it is possible to obtain many more space-time channels. For example, with a SNR of 15 dB, it is

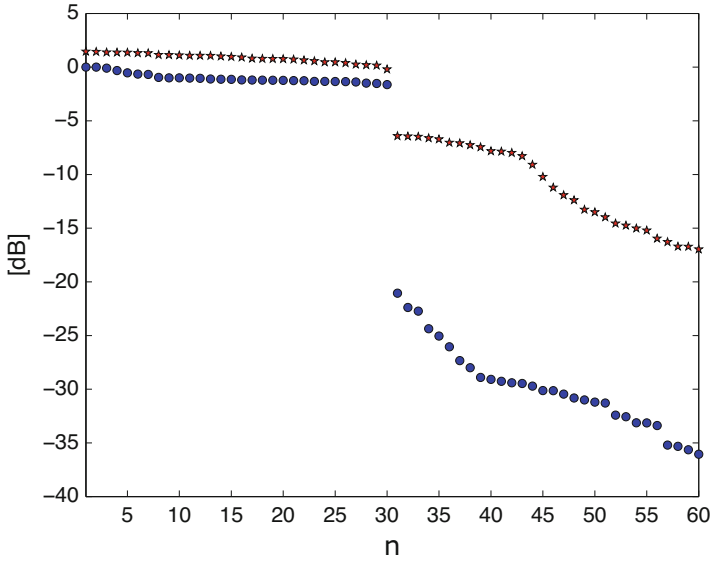


Fig. 2.15 Example of singular values without (blue circle) and with (red stars) control of the environment; the optimization of the scattering environment doubles the STNDF available for the transmission

possible to have more than 50 space-time channels, while a SNR of 20 dB allows the use of all the available space-time channels.

The increase of all the singular values indicates also that the energy on the receiving antenna is higher when using the EPD. This is not a trivial result. EPD requires electronically controlled scattering elements that cause losses in the scattering process. The EPD design is a delicate balance between increasing the number of elements to extend the set of radiated fields and the ability to concentrate the energy toward the receiver and the use of a smaller number of elements to reduce losses avoiding the use of expensive amplifiers. Furthermore, the response of the scattering elements (including the electronic switches) changes with the frequency, making it difficult to select the best scattering response of a broadband EPD.

Despite the difficulties mentioned above, there has recently been a renewed interest in controlling the propagation environment by means of reconfigurable scattering devices. Indeed, one of the main novelties in 6G is the introduction of electromagnetic signal processing directly at the Deep Physical Layer level.

One of the most promising solutions is represented by the Reconfigurable Intelligent Surfaces (RIS) [5]. The main application of RISs consists in extending the cell size in millimeter or THz communication beyond the line of sight The RISs are designed to obtain a reflection of the incident wave in the desired direction by means of an appropriate phase variation on their surface obtained from electronically controlled scattering elements.

The implementation of RIS in communication systems is a great challenge from both a theoretical and a technological point of view. Although the basic structure is similar to that of reflectarrays [79], whose design methods are now well established, their use in non-military systems, where cost is the determining factor for the success or failure, requires new economical, reliable, and large-band technological solutions.

2.6 Conclusions

The communication systems of the next generation will face the enormous challenges of a deeply interconnected society.

Achieving the required performance will increasingly require investigating the boundaries between the information theory, which describes the limits of the telecommunication system in terms of bit rate, and the theory of electromagnetism, which describes the physical limits to which it is subject.

In order to reach this goal, it is necessary a description of the communication process at a DPL compared to Physical Layer of the OSI model that allows to relate the physical processes underlying the communication to the amount of information transmissible by the system. This suggests the introduction of a further layer underneath the physical layer, called the Deep Physical Layer.

In DPL the communication process is modeled in terms of distinguishable configurations of the radiated electromagnetic field. The performance of the communication system in terms of channel capacity is fixed at the level of the Deep Physical Layer. It is also possible to carry out a "processing" of the electromagnetic field at the Deep Physical Layer level which allows to optimize the performance of the telecommunications system, obtaining an "intelligent DPL" environment. The solutions currently being studied for 6G based on the use of reconfigurable intelligent surfaces are an example of this approach.

As last remark, this study is only one aspect of the rather complex and fascinating problem of the relationship between energy and information. In fact, the close connection between "bits and fields" shown by the DPL analysis is an indication of the existence of some important relationships between information and physics. These connections are much deeper than the simple use of electromagnetic waves in wireless communication systems and involve fundamental aspects of physics.

In the famous example discussed in the *Theory of Heat* [80], Maxwell introduced a demon that seems capable of violating the validity of the second law of thermodynamics. Maxwell considers a box divided into two sections, connected by a hole closed by a small door. A demon is able to see single molecules. Opening the door at the right moment traps the higher velocity molecules in one section, say the left one, and the slower velocity ones in the other. In this way the demon is able to increase the temperature difference between the two sections, violating the second thermodynamic principle. Variants of Maxwell's demon were later discussed by Szilard, Brillouin, Rothstein, and other authors [10, 81]. In Szilard's model,

the selection of molecules in thermal equilibrium requires that the demon *sees* the individual molecules, that is, a sort of detection is required, which on the one hand increases the state of knowledge of the demon with respect to the environment; on the other hand, it causes an increase in the entropy of statistical mechanics, which is "transformed" into information by the demon. Overall, the entropy of the system does not decrease, and therefore there is no violation of the second thermodynamic principle. Generally speaking, the many examples reported in literature converge on the conclusion that even for information there is no "free meal" and considering information completely independent of physics can cause the violation of first principles. Information and physics are intimately related, and any method that allows us to transmit an unlimited amount of information with finite energy simply violates the laws of our universe.

What are the implications of this conclusion in communication systems? Communication systems basically encode information in different states of a physical observable. Accordingly, for any ϵ approximation level, the set of the states must admit a *finite* number of ϵ-distinguishable balls. This requires that the set is *totally bounded*[4] [19, 46]. In a physical approach to the communication system, the total boundness of the observed quantity (i.e., the electromagnetic field in our case) on the observation surface is automatically guaranteed by the properties of the radiation operator [9].

Introduction of physics in the communication model allows us to walk on the boundaries of the communication data rate avoiding the risk of obtaining non-physical solutions.

References

1. E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, R. Zhang, Wireless communications through reconfigurable intelligent surfaces. IEEE Access **7**, 116753–116773 (2019)
2. S. Dang, O. Amin, B. Shihada, M.-S. Alouini, What should 6G be? Nat. Electron. **3**(1), 20–29 (2020)
3. E. Björnson, Ö. Özdogan, E.G. Larsson, Reconfigurable intelligent surfaces: three myths and two critical questions. IEEE Commun. Mag. **58**(12), 90–96 (2020)
4. E. Basar, Reconfigurable intelligent surface-based index modulation: a new beyond mimo paradigm for 6G. IEEE Trans. Commun. **68**(5), 3187–3196 (2020)
5. M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, S. Tretyakov, Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research, and the road ahead. IEEE J. Select. Areas Commun. **38**(11), 2450–2525 (2020)
6. A. Pizzo, T.L. Marzetta, L. Sanguinetti, Spatial characterization of holographic mimo channels (2019). Preprint. arXiv:1911.04853

[4] We recall that fixed an $\epsilon > 0$, an ϵ-net of A is a subset A' of a metric space A such that A' is finite and $\forall x \in A,\ \exists\, y \in A' : d(y, x) \leq \epsilon$, wherein $d(\cdot, \cdot)$ is the distance. A totally bounded set in a metric space contains a ϵ-net $\forall \epsilon > 0$.

7. A. Pizzo, T.L. Marzetta, L. Sanguinetti, Degrees of freedom of holographic mimo channels (2019). Preprint. arXiv:1911.07516
8. A. Pizzo, T.L. Marzetta, L. Sanguinetti, Spatially-stationary model for holographic mimo small-scale fading. IEEE J. Select. Areas Commun. **38**(9), 1964–1979 (2020)
9. M.D. Migliore, Horse (electromagnetics) is more important than horseman (information) for wireless transmission. IEEE Trans. Antennas Propag. **67**(4), 2046–2055 (2018)
10. L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1959)
11. D.S. Lebedev, L.B. Levitin, Information transmission by electromagnetic field. Inf. Control. **9**(1), 1–22 (1966)
12. G.T. Di Francia, Degrees of freedom of an image. JOSA **59**(7), 799–804 (1969)
13. F. Gori, G. Guattari, Shannon number and degrees of freedom of an image. Opt. Commun. **7**(2), 163–165 (1973)
14. C.W. Helstrom, Resolvable degrees of freedom in observation of a coherent object. JOSA **67**(6), 833–838 (1977)
15. T. Jannson, Shannon number of an image and structural information capacity in volume holography. Optica Acta Int. J. Optics **27**(9), 1335–1344 (1980)
16. R. Piestun, D.A. Miller, Electromagnetic degrees of freedom of an optical system. JOSA A **17**(5), 892–902 (2000)
17. M. Born, E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Elsevier, Amsterdam, 2013)
18. L. Mandel, E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995)
19. M.D. Migliore, The world beneath the physical layer: an introduction to the deep physical layer. IEEE Access **9**, 77106–77126 (2021)
20. O.M. Bucci, G. Franceschetti, On the degrees of freedom of scattered fields. IEEE Trans. Antennas Propag. **37**(7), 918–926 (1989)
21. O. Bucci, G. Franceschetti, On the spatial bandwidth of scattered fields. IEEE Trans. Antennas Propag. **35**(12), 1445–1455 (1987)
22. O.M. Bucci, C. Gennarelli, C. Savarese, Representation of electromagnetic fields over arbitrary surfaces by a finite and nonredundant number of samples. IEEE Trans. Antennas Propag. **46**(3), 351–359 (1998)
23. O.M. Bucci, C. Gennarelli, C. Savarese, Fast and accurate near-field-far-field transformation by sampling interpolation of plane-polar measurements. IEEE Trans. Antennas Propag. **39**(1), 48–55 (1991)
24. J. Bolomey, O.M. Bucci, L. Casavola, G. D'Elia, M.D. Migliore, A. Ziyyat, 'Reduction of truncation error in near-field measurements of antennas of base-station mobile communication systems. IEEE Trans. Antennas Propag. **52**(2), 593–602 (2004)
25. J.W. Wallace, M.A. Jensen, Intrinsic capacity of the mimo wireless channel, in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 2 (IEEE, Piscataway, 2002), pp. 701–705
26. S. Loyka, Information theory and electromagnetism: Are they related? in *2004 10th International Symposium on Antenna Technology and Applied Electromagnetics and URSI Conference* (IEEE, Piscataway, 2004), pp. 1–5
27. T.K. Sarkar, S. Burintramart, N. Yilmazer, S. Hwang, Y. Zhang, A. De, M. Salazar-Palma, A discussion about some of the principles/practices of wireless communication under a maxwellian framework. IEEE Trans. Antennas Propag. **54**(12), 3727–3745 (2006)
28. M.D. Migliore, Restoring the symmetry between space domain and time domain in the channel capacity of mimo communication systems, in *2006 IEEE Antennas and Propagation Society International Symposium* (IEEE, Piscataway, 2006), pp. 333–336
29. A.S. Poon, R.W. Brodersen, D.N. Tse, Degrees of freedom in multiple-antenna channels: a signal space approach. IEEE Trans. Inform. Theory **51**(2), 523–536 (2005)
30. S.H. Lee, S.Y. Chung, Capacity scaling of wireless ad hoc networks: shannon meets maxwell. IEEE Trans. Inform. Theory **58**(3), 1702–1715 (2012)
31. M. Gustafsson, S. Nordebo, Characterization of mimo antennas using spherical vector waves. IEEE Trans. Antennas Propag. **54**(9), 2679–2682 (2006)

32. M. Gustafsson, S. Nordebo, On the spectral efficiency of a sphere, Progress In Electromagnetics Research, **67**, 275–296 (2007)
33. J. Xu, R. Janaswamy, Electromagnetic degrees of freedom in 2-d scattering environments. IEEE Trans. Antennas Propag. **54**(12), 3882–3894 (2006)
34. M.D. Migliore, On the role of the number of degrees of freedom of the field in mimo channels. IEEE Trans. Antennas Propag. **54**(2), 620–628 (2006)
35. M.D. Migliore, On electromagnetics and information theory. IEEE Trans. Antennas Propag. **56**(10), 3188–3200 (2008)
36. F.K. Gruber, E.A. Marengo, New aspects of electromagnetic information theory for wireless and antenna systems. IEEE Trans. Antennas Propag. **56**(11), 3470–3484 (2008)
37. M.L. Morris, M.A. Jensen, J.W. Wallace, Superdirectivity in mimo systems. IEEE Trans. Antennas Propag. **53**(9), 2850–2857 (2005)
38. A.S. Poon, D.N. Tse, R.W. Brodersen, Impact of scattering on the capacity, diversity, and propagation range of multiple-antenna channels. IEEE Trans. Inform. Theory **52**(3), 1087–1100 (2006)
39. A. Poon, D. Tse, Polarization degrees of freedom, in *2008 IEEE International Symposium on Information Theory* (IEEE, Piscataway, 2008), pp. 1587–1591
40. M. Franceschetti, M.D. Migliore, P. Minero, F. Schettino, The information carried by scattered waves: near-field and nonasymptotic regimes. IEEE Trans. Antennas Propag. **63**(7), 3144–3157 (2015)
41. C. Ehrenborg, M. Gustafsson, Fundamental bounds on mimo antennas. IEEE Antennas Wirel. Propag. Lett. **17**(1), 21–24 (2017)
42. M. Franceschetti, M.D. Migliore, P. Minero, The capacity of wireless networks: information-theoretic and physical limits. IEEE Trans. Inform. Theory **55**(8), 3413–3424 (2009)
43. C. Ehrenborg, M. Gustafsson, Physical bounds and radiation modes for mimo antennas. IEEE Trans. Antennas Propag. **68**(6), 4302–4311 (2020)
44. C. Ehrenborg, M. Gustafsson, M. Capek, Capacity bounds and degrees of freedom for mimo antennas constrained by Q-factor. IEEE Trans. Antennas Propag. **69**(9), 5388–5400 (2021)
45. M.D. Migliore, Who cares about the horse? A gentle introduction to information in electromagnetic theory [wireless corner]. IEEE Antennas Propag. Mag. **62**(5), 126–137 (2020)
46. A. Kolmogorov, S. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Dover Puplications Inc., Mineola, 1999)
47. V. Tikhomirov, ε-entropy and ε-capacity of sets in functional spaces, in *Selected works of AN Kolmogorov* (Springer, Berlin, 1993), pp. 86–170
48. R.T. Prosser, W.L. Root, The ε-entropy and ε-capacity of certain time-invariant channels. J. Math. Anal. Appl. **21**(2), 233–241 (1968)
49. Y. Niu, Y. Li, D. Jin, L. Su, A.V. Vasilakos, A survey of millimeter wave communications (mmwave) for 5G: opportunities and challenges. Wirel. Netw. **21**(8), 2657–2676 (2015)
50. Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G.K. Karagiannidis, P. Fan, 6G wireless networks: vision, requirements, architecture, and key technologies. IEEE Veh. Technol. Mag. **14**(3), 28–41 (2019)
51. À.O. Martínez, E. De Carvalho, J.Ø. Nielsen, Towards very large aperture massive mimo: a measurement based study, in *2014 IEEE Globecom Workshops (GC Wkshps)* (IEEE, Piscataway, 2014), pp. 281–286
52. H.Q. Ngo, A. Ashikhmin, H. Yang, E.G. Larsson, T.L. Marzetta, Cell-free massive mimo versus small cells. IEEE Trans. Wirel. Commun. **16**(3), 1834–1850 (2017)
53. G. Interdonato, E. Björnson, H.Q. Ngo, P. Frenger, E.G. Larsson, Ubiquitous cell-free massive mimo communications. EURASIP J. Wirel. Commun. Netw. **2019**(1), 197 (2019)
54. U. Madhow, D.R. Brown, S. Dasgupta, R. Mudumbai, Distributed massive mimo: algorithms, architectures and concept systems, in *2014 Information Theory and Applications Workshop (ITA)* (IEEE, Piscataway, 2014), pp. 1–7
55. F. Tariq, M.R. Khandaker, K.-K. Wong, M.A. Imran, M. Bennis, M. Debbah, A speculative study on 6G. IEEE Wirel. Commun. **27**(4), 118–125 (2020)

56. C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, M. Debbah, Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends. IEEE Wireless Communications **27**(5), 118–125 (2020)
57. L. Subrt, P. Pechac, Intelligent walls as autonomous parts of smart indoor environments. IET Commun. **6**(8), 1004–1010 (2012)
58. S. Hu, F. Rusek, O. Edfors, Beyond massive mimo: the potential of data transmission with large intelligent surfaces. IEEE Trans. Signal Process. **66**(10), 2746–2758 (2018)
59. S. Sun, T. Wang, F. Yang, J. Song, Z. Han, Intelligent reflecting surface-aided visible light communications: potentials and challenges. IEEE Veh. Technol. Mag. **17**(1), 47–56 (2021)
60. S.V. Hum, J. Perruisseau-Carrier, Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: a review. IEEE Trans. Antennas Propag. **62**(1), 183–198 (2013)
61. C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, I. Akyildiz, A new wireless communication paradigm through software-controlled metasurfaces. IEEE Commun. Mag. **56**(9), 162–169 (2018)
62. G. Oliveri, P. Rocca, M. Salucci, A. Massa, Holographic smart em skins for advanced beam power shaping in next generation wireless environments (2021). Preprint. arXiv:2106.10932
63. D.S. Jones, *Acoustic and Electromagnetic Waves* (Oxford, New York, 1986)
64. G.W. Hanson, A.B. Yakovlev, *Operator Theory for Electromagnetics: An Introduction* (Springer Science & Business Media, Berlin, 2013)
65. A. Paulraj, A.P. Rohit, R. Nabar, D. Gore, *Introduction to Space-Time Wireless Communications* (Cambridge University Press, Cambridge, 2003)
66. A. Pinkus, *N-Widths in Approximation Theory*, vol. 7 (Springer Science & Business Media, Berlin, 2012)
67. D. Slepian, H.O. Pollak, Prolate spheroidal wave functions, fourier analysis and uncertainty—I. Bell Syst. Tech. J. **40**(1), 43–63 (1961)
68. W.H. Fuchs, On the eigenvalues of an integral equation arising in the theory of band-limited signals. J. Math. Anal. Appl. **9**(3), 317–330 (1964)
69. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
70. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010)
71. R.E. Collin, *Antenna Theory*, vol. 1 (McGraw-Hill, New York, 1969)
72. M.D. Migliore, On the sampling of the electromagnetic field radiated by sparse sources. IEEE Trans. Antennas Propag. **63**(2), 553–564 (2014)
73. M.R. Andrews, P.P. Mitra, R. DeCarvalho, Tripling the capacity of wireless communications using electromagnetic polarization. Nature **409**(6818), 316–318 (2001)
74. R. Tian, B.K. Lau, Experimental verification of degrees of freedom for colocated antennas in wireless channels. IEEE Trans. Antennas Propag. **60**(7), 3416–3423 (2012)
75. R.G. Gallager, *Information Theory and Reliable Communication*, vol. 2 (Springer, Berlin, 1968)
76. M.D. Migliore, Big, bigger, biggest. What should we expect from 6G antennas? in *2020 XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science* (IEEE, Piscataway, 2020), pp. 1–4
77. M.D. Migliore, D. Pinchera, F. Schettino, Improving channel capacity using adaptive mimo antennas. IEEE Trans. Antennas Propag. **54**(11), 3481–3489 (2006)
78. D. Pinchera, J.W. Wallace, M.D. Migliore, M.A. Jensen, Experimental analysis of a wideband adaptive-mimo antenna. IEEE Trans. Antennas Propag. **56**(3), 908–913 (2008)
79. J. Huang, J.A. Encinar, Reflectarray Antennas (John Wiley & Sons, Hoboken, 2007)
80. J. C. Maxwell, Theory of heat, Longsman, Green, and Company (1876)
81. L. Szilard, Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen. Z. Phys. **53**(11), 840–856 (1929)

Chapter 3
FBMC: A Waveform Candidate for Beyond 5G

Prem Singh and Ekant Sharma

Recently, offset quadrature amplitude modulation (OQAM)-based filter bank multicarrier (FBMC), due to its reduced out-of-band (OOB) emission, has attracted significant research interests for replacing orthogonal frequency division multiplexing (OFDM) in future wireless communication systems. This chapter analyses and designs FBMC-OQAM waveform-based multiple-input multiple-output (MIMO) and multi-user massive MIMO systems. It begins by describing key features and differences of FBMC waveform over the widely popular OFDM waveform, followed by the discussion over key challenges in designing FBMC-based MIMO and massive MIMO systems. A semi-blind (SB) channel state information (CSI) estimation scheme, which enhances the performance with a limited pilot overhead, is developed for MIMO-FBMC system along with its Cramer-Rao lower bound (CRLB) for benchmarking the performance. To compare the performance of FBMC and OFDM waveforms in the uplink transmission, the achievable sum rates are derived for multi-user (MU) massive MIMO technology relying on FBMC waveform with maximum ratio combining (MRC) and zero-forcing (ZF) receivers. The corresponding power scaling laws for MU massive MIMO-FBMC are also found. It is shown that in practical impairments such as carrier frequency offset, massive MIMO-FBMC systems significantly outperform their OFDM counterparts.

P. Singh (✉)
International Institute of Information Technology, Bangalore, India
e-mail: prem.singh@iiitb.ac.in

E. Sharma
Indian Institute of Technology, Roorkee, India
e-mail: ekant@ece.iitr.ac.in

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_3

3.1 Introduction

From the past two decades, orthogonal frequency division multiplexing (OFDM) has enjoyed its widespread dominance in broadband wired [1] and wireless [2] communication systems. OFDM waveform has been widely embraced in digital subscriber lines (DSL) standards, as well as wireless standards such as the third-generation partnership program long-term evolution (3GPP-LTE), IEEE 802.16, IEEE 802.11, LTE-Advanced and 5G-New Radio (NR). The key advantages of OFDM signalling are (i) orthogonality among subcarriers allows synthesis and analysis of the transmit and receive signals using computationally efficient inverse fast Fourier transform (IFFT) and FFT blocks, respectively; (ii) the use of cyclic prefix (CP) guarantees one-tap equalization; and (iii) trivial amalgamation with multiple-input multiple-output (MIMO) and massive MIMO technologies. Fifth-generation (5G) communications systems are characterised by a wide range of use cases such as enhanced Machine Type Communications (eMTC), Ultra-Reliable Low latency Communications (URLLC) and enhanced Mobile BroadBand (eMBB) [3]. In order to cope with a large number of applications, future communication systems require a flexible time-frequency resource allocation. The 3GPP body has yet again adopted OFDM (with some minor changes) for 5G wireless communication systems. OFDM shapes each of its subcarrier using a rectangular window, which results in a sinc-shaped frequency localization. As soon as the synchronization in OFDM is lost, the time-domain rectangular pulse shape associated with synthesised subcarriers at the transmitter and analysed subcarriers at the receiver, due to its relatively higher out-of-band (OOB) emission, results in significant leakage of signal power among the band of neighbourhood users [4]. Thus, OFDM-based systems are sensitive to practical impairments such as carrier frequency offset and timing offset, especially in vehicular scenarios where tracking Doppler shifts of different user equipments is challenging. For instance, achieving perfect synchronization in the uplink of orthogonal frequency division multiple access (OFDMA) may not be possible. This happens because each of the users transmits independently from different locations. Attaining synchronization in OFDM-based cognitive radio systems is even more challenging, because primary and secondary users transmit independently and may be operating on distinct standards [5, 6].

In view of the above observations, researchers are motivated to investigate new multicarrier modulation schemes for future wireless communication systems. Such schemes, while outperforming existing solutions in terms of OOB emission, should also ensure similar efficiency and robustness as OFDM. In this context, engineers in Alcatel-Lucent Bell Laboratories have investigated a waveform called universal filtered multicarrier (UFMC) [7], wherein a set of subcarriers assigned to a node are processed through a filter to minimize the multi-node interference. This technique, due to degraded orthogonality among the subcarrier in the existence of multipath fading, results in a performance loss [8]. The second popular waveform is generalized frequency division multiplexing (GFDM), proposed in [9]. In GFDM, each subcarrier is shaped using a well frequency-time (FT) localized filter in circular

fashion such that one CP is attached to a block of N_d information bearing symbols that are dispersed over N subcarriers and N_d/N symbol instants [9].

Another potential waveform candidate for future wireless communication systems is based on filter bank multicarrier (FBMC) operation [10]. Contrary to the time-domain rectangular pulse with sinc-shaped frequency localization in OFDM, subcarriers in FBMC are synthesised and analysed using a bank of well FT localized modulated and demodulated filters, respectively. There are several variants of FBMC in the existing literature, namely, staggered multi-tone (SMT), filtered multi-tone (FMT) and cosine-modulated multi-tone (CMT) [10]. FMT, which has been evolved particularly for DSL [11], is designed using the classic frequency division multiplexing (FDM) operation, wherein the subcarriers band are made unconnected by introducing guard bands. Thus, FMT is a bandwidth inefficient signalling scheme. In contrast, both SMT and CMT, by allowing overlapping of adjacent subcarriers, offer maximum bandwidth efficiency [10]. To carry pulse amplitude modulated (PAM) symbols, overlapped vestigial side-band (VSB) modulated signals are staggered in CMT. On the other side, signalling in SMT is designed using staggering of overlapping double side-band modulated format to carry quadrature amplitude modulated (QAM) symbols, whose real and imaginary parts are time-offset by one half of a symbol duration. A formal mathematical relationship between CMT and SMT has been derived in [12]. Since the VSB modulation in CMT requires Hilbert transform, its implementation is more complex than SMT. In the literature, SMT is popularly known as offset quadrature amplitude modulation (OQAM)-based OFDM (OFDM-OQAM) or FBMC-OQAM.

Figure 3.1a,b show a section of OFDM and FBMC-OQAM filter banks, respectively. The subchannel index therein corresponds to the frequency axis with unity subcarrier spacing. The subcarrier orthogonality for the OFDM filter bank in Fig. 3.1a can be observed through the zero crossings at the integer multiple of subcarrier spacing where only one subchannel is non-zero. As shown in Fig. 3.1a,

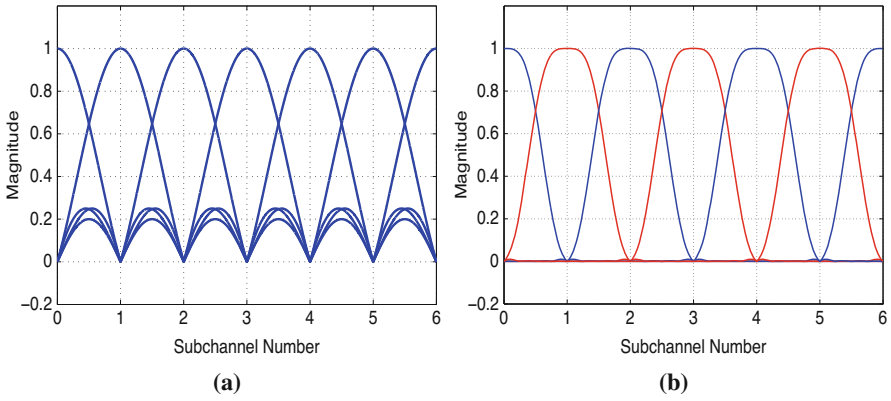


Fig. 3.1 Section of a filter bank. (**a**) OFDM and (**b**) FBMC-OQAM relying on Phydyas filter [13] with a overlapping factor 4

OFDM filter bank has significantly higher OOB emission due to the sinc-shaped spectrum of the rectangular prototype filter. Therefore, performance of OFDM-based systems degrades severely in the presence of synchronization error due to carrier frequency offset (CFO) and timing offset. On the other hand, as can be observed from Fig. 3.1b, only adjacent subcarriers are overlapping in FBMC-OQAM due to the associated well frequency-time localized prototype filter. Thus, one can separate the adjacent bands by inserting an empty subcarrier between them. Furthermore, the sharp prototype filter in FBMC-OQAM significantly lowers down the OOB emission and relaxes the stringent synchronization requirement in such systems [14]. Additionally, these sharp filters also avoid the need for CP that is otherwise required in OFDM to remove inter-symbol-interference (ISI). This increases the spectral efficiency of FBMC-OQAM-based systems. The implementation of FBMC-OQAM can be realised using a computationally efficient polyphase structure [15]. Furthermore, for relatively low frequency-selective channels, FBMC-OQAM can be efficiently coupled to MIMO technology [16]. Additionally, the concept of GFDM can also be extended to FBMC-OQAM [17, 18]. The resulting signalling format from this amalgamation is called circular FBMC (C-FBMC). The references [19] and [4], respectively, show the benefits of FBMC-OQAM waveform over its OFDM counterpart in the uplink of multi-user (MU) networks and cognitive radios. In the light of the above-listed benefits of FBMC-OQAM, and its ability to address the shortcomings of OFDM by using sharp pulse shaping, FBMC-OQAM has recently received significant research interests [16, 20–23], which reflect that FBMC-OQAM is a compelling signalling technique for future wireless communication systems. The aim of this chapter is, therefore, to analyse and design FBMC-OQAM-based multiple-input multiple-output (MIMO) and multi-user massive MIMO systems. For brevity, FBMC-OQAM is referred to as FBMC in the sequel.

The subcarrier in FBMC waveform, unlike OFDM, is orthogonal in the real domain only [15]. The resulting *intrinsic interference* challenges amalgamation of FBMC signalling with future mobile communication systems. Furthermore, the overlapping of FBMC symbols in the time-domain poses additional challenges. For example, FBMC channel state information (CSI) estimation needs the placement of zeros between the adjacent training symbols [24]. Thus, one has to carefully compute the intrinsic interference for constructing the virtual training symbols at the receiver. Moreover, one also needs to find the optimal number of zeros required to avoid overdesign/underdesign of FBMC systems. In light of the above challenges, it is not always feasible to use the existing solutions or corresponding analysis for OFDM-based systems, while designing FBMC aided future wireless systems.

3.2 Organization of Chapter

The next section discusses single-input single-output (SISO) and MIMO-FBMC system models and the key differences between OFDM and FBMC waveforms.

Section 3.4 demonstrates a semi-blind channel estimator for MIMO-FBMC systems. Section 3.5 demonstrates and compares the performance of FBMC and OFDM waveforms in the uplink transmission of MU massive MIMO systems. Section 3.6 concludes the chapter and provides future directions.

3.3 FBMC System Model

The continuous-time baseband FBMC transmitted signal is expressed as [15, 25]

$$s(t) = \sum_{m=0}^{N-1} \sum_{n \in \mathbb{Z}} d_{m,n} \chi_{m,n}(t), \tag{3.1}$$

where $d_{m,n}$ are real-valued OQAM symbols at the mth subcarrier and the nth symbol instant. The parameter N symbolises the number of subcarriers. The FBMC basis function $\chi_{m,n}(t)$ is defined as

$$\chi_{m,n}(t) = p(t - nT/2) e^{j2\pi mFt} e^{\phi_{m,n}}, \tag{3.2}$$

where T is the QAM symbol duration, $F = 1/T$ is the spacing between two consecutive subcarriers, $T/2$ is the offset between the in-phase and quadrature parts of a QAM symbol and $p(t)$ is the symmetrical real-valued pulse, which is different from the rectangular pulse in OFDM. The phase factor $\phi_{m,n}$ is defined as modulo π, for example, $\phi_{m,n} = (\pi/2)(m + n)$ [15]. The real OQAM symbols $d_{m,n}$ are drawn from the spatially and temporally independent and identically distributed (i.i.d.) in-phase and quadrature components of a QAM symbol $c_{m,n}$ as follows:

$$d_{m,2n} = \Phi(c_{m,n}) = \begin{cases} \Re(c_{m,n}), & m \text{ even} \\ \Im(c_{m,n}), & m \text{ odd} \end{cases} \tag{3.3}$$

$$d_{m,2n+1} = \Phi(c_{m,n}) = \begin{cases} \Im(c_{m,n}), & m \text{ even} \\ \Re(c_{m,n}), & m \text{ odd.} \end{cases} \tag{3.4}$$

Each component OQAM symbols $d_{m,2n}$ and $d_{m,2n+1}$ is of duration $T/2$. Let $\mathbb{E}[c_{m,n} c^*_{m,n}] = 2P_d$, which implies that $\mathbb{E}[d_{m,n} d^*_{m,n}] = P_d$. In the presence of ideal channel without noise, the FBMC demodulation at subcarrier index $\bar{m}$ and symbol instant $\bar{n}$ is performed using the matched filtering operation as shown below [15]

$$\hat{d}_{\bar{m},\bar{n}} = \Re\left\{ \int_{t=-\infty}^{+\infty} s(t) \chi^*_{\bar{m},\bar{n}}(t) dt \right\} = \sum_{m=0}^{N-1} \sum_{n \in \mathbb{Z}} d_{m,n} \Re\left\{ \int_{t=-\infty}^{+\infty} \chi_{m,n}(t) \chi^*_{\bar{m},\bar{n}}(t) dt \right\}. \tag{3.5}$$

Thus, in order to recover OQAM symbols $d_{m,n}$ at the receiver, the basis functions $\chi_{m,n}(t)$ satisfy the following real-field orthogonality condition

$$\Re\left\{\int_{-\infty}^{+\infty} \chi_{m,n}(t)\chi^*_{\bar{m},\bar{n}}(t)dt\right\} = \delta_{m,\bar{m}}\delta_{n,\bar{n}}, \tag{3.6}$$

where $\delta_{m,\bar{m}}$ is the Kronecker delta with $\delta_{m,\bar{m}} = 1$ if $m = \bar{m}$ and zero otherwise.

Since N complex valued symbols are transmitted in T time interval, $T_s = T/N$ is the critical sampling interval. The discrete-time FBMC baseband transmitted signal is obtained by sampling $s(t)$ with the sampling rate $1/T_s$. The causal discrete-time prototype pulse $p[k]$ of length L_p is obtained by truncating $p(t)$ from $-(L_p/2)T_s$ to $(L_p/2)T_s$ and delaying it by $((L_p - 1)/2)T_s$. The discrete-time baseband signal is obtained by sampling $s(t)$ at $t = kT_s$ as

$$s[k] = \sum_{m=0}^{N-1}\sum_{n\in\mathbb{Z}} d_{m,n}\chi_{m,n}[k]. \tag{3.7}$$

The discrete-time basis function $\chi_{m,n}[k]$ is defined as

$$\chi_{m,n}[k] = p[k - nN/2]e^{j\frac{2\pi}{N}m\left(k-\frac{L_p-1}{2}\right)}e^{j\phi_{m,n}}. \tag{3.8}$$

The equivalent real-field orthogonality in the discrete domain for reconstructing real OQAM symbols $d_{m,n}$ is given as

$$\Re\left\{\sum_{k=-\infty}^{+\infty} \chi_{m,n}[k]\chi^*_{\bar{m},\bar{n}}[k]\right\} = \delta_{m,\bar{m}}\delta_{n,\bar{n}}. \tag{3.9}$$

Figure 3.2a,b shows the discrete-time baseband model for the FBMC transmitter and receiver, respectively. The main dissimilarities between FBMC and OFDM waveforms lie (i) in the selection of pulse-shaping filter $p[k]$ and (ii) in the property the former accepted OQAM symbols instead of QAM symbols. The function $\Phi(\cdot)$ in Fig. 3.2a performs the operations described in (3.3) and (3.4) for the even and odd indexed subcarriers. This operation is reversed at the receiver using the function $\Phi^{-1}(\cdot)$. The other blocks are self-explanatory. Let the quantity $\xi^{\bar{m},\bar{n}}_{m,n} = \sum_{k=-\infty}^{+\infty} \chi_{m,n}[k]\chi^*_{\bar{m},\bar{n}}[k]$. Thus, it follows from (3.9) that

$$\xi^{\bar{m},\bar{n}}_{m,n} = \begin{cases} 1, & \text{if } (m,n) = (\bar{m},\bar{n}) \\ j\langle\xi\rangle^{\bar{m},\bar{n}}_{m,n}, & \text{if } (m,n) \neq (\bar{m},\bar{n}), \end{cases} \tag{3.10}$$

where $\langle\xi\rangle^{\bar{m},\bar{n}}_{m,n} = \Im\{\sum_{k=-\infty}^{+\infty} \chi_{m,n}[k]\chi^*_{\bar{m},\bar{n}}[k]\}$ extracts the imaginary part of the cross correlation between the basis functions [26]. The signal received at the SISO-FBMC receiver is

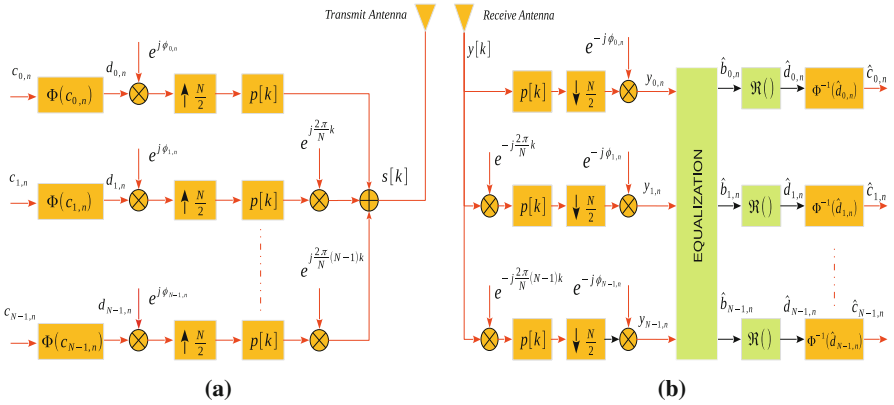


Fig. 3.2 Discrete-time equivalent baseband model for FBMC: (**a**) transmitter and (**b**) receiver

$$y[k] = (s[k] * h[k]) + \eta[k], \tag{3.11}$$

where $h[k]$ denotes impulse response of an L_h-tap channel. The scaler quantity $\eta[k]$ is an additive white Gaussian noise (AWGN) with mean zero and variance σ_η^2. At the receive, the FBMC signal is demodulated at frequency index $\bar{m}$ and symbol instant $\bar{n}$ by matching it with the basis function $\chi_{\bar{m},\bar{n}}[k]$ as [15]

$$y_{\bar{m},\bar{n}} = \sum_{k=-\infty}^{+\infty} y[k]\chi^*_{\bar{m},\bar{n}}[k]. \tag{3.12}$$

By substituting $s[k]$, $\chi_{\bar{m},\bar{n}}[k]$ and $y[k]$ from (3.7), (3.8) and (3.11), respectively, in the above expression, $y_{\bar{m},\bar{n}}$ is expanded as

$$y_{\bar{m},\bar{n}} = \sum_{m=0}^{N-1}\sum_{n\in\mathbb{Z}} d_{m,n} \sum_{l=0}^{L_h-1} h[l]e^{-j2\pi ml/N} \sum_{k} p\big[k-l-nN/2\big]p\big[k-\bar{n}N/2\big]$$
$$\times\, e^{j(\phi_{m,n}-\phi_{\bar{m},\bar{n}})} e^{j\frac{2\pi}{N}(m-\bar{m})\left(k-\frac{L_p-1}{2}\right)} + \eta_{\bar{m},\bar{n}}, \tag{3.13}$$

where $\eta_{\bar{m},\bar{n}} = \sum_{k=-\infty}^{+\infty} \eta[k]\chi^*_{\bar{m},\bar{n}}[k]$ is the demodulated noise which obeys Gaussian distribution with mean

$$\mathbb{E}\big[\eta_{\bar{m},\bar{n}}\big] = \mathbb{E}\Bigg[\sum_{k=-\infty}^{+\infty} \eta[k]\chi^*_{\bar{m},\bar{n}}[k]\Bigg] = 0 \quad \text{and}$$

$$\mathrm{Var}\big[\eta_{\bar{m},\bar{n}}\big] = \sigma_\eta^2 \sum_{k=-\infty}^{+\infty} \chi_{\bar{m},\bar{n}}[k]\chi^*_{\bar{m},\bar{n}}[k] = \sigma_\eta^2.$$

The covariance between $\eta_{\bar{m},\bar{n}}$ and $\eta_{m,n}$ is computed as

$$\mathrm{Cov}\left[\eta_{m,n}, \eta^*_{\bar{m},\bar{n}}\right] = \mathbb{E}\left[\eta_{m,n}\eta^*_{\bar{m},\bar{n}}\right] = \sigma^2_\eta \xi^{\bar{m},\bar{n}}_{m,n}. \tag{3.14}$$

The noise at FT index $(\bar{m}, \bar{n})$ is thus correlated. This correlation, however, is negligible because of the associated sharp prototype filters for pulse shaping in FBMC [24]. The duration of the prototype pulse $p[k]$ in FBMC is typically chosen as an integer multiple of the symbol time T. Thus, the pulse duration is significantly larger than the channel delay spread, i.e. $L_p = k_0 N >> L_h$. For example, references [24, 26] set k_0 to be 4. This implies that the bandwidth of the pulse-shaping filter $p[k]$ is significantly lower than the channel coherence bandwidth. As a result, the impulse response $p[k]$ of prototype filter in time has negligible variations over the channel delay spread. Therefore, as described in [24, 26]

$$p[k - l - nN/2] \approx p[k - nN/2], \quad \text{for } l \in [0,\ L_h]. \tag{3.15}$$

Upon employing the above result, the expression for $y[k]$ in (3.13) can be simplified as

$$y_{\bar{m},\bar{n}} \approx \sum_{m=0}^{N-1} \sum_{n\in\mathbb{Z}} d_{m,n} H_m \xi^{\bar{m},\bar{n}}_{m,n} + \eta_{\bar{m},\bar{n}}, \tag{3.16}$$

where

$$H_m = \sum_{l=0}^{L_h-1} h[l] e^{-j2\pi ml/N} \tag{3.17}$$

symbolises channel frequency response (CFR) for the mth subcarrier. The channel $h[k]$ is considered to be quasi-static throughout the chapter. Separating the desired and undesired terms in (3.16), one obtains

$$y_{\bar{m},\bar{n}} = H_{\bar{m}} \Bigg(d_{\bar{m},\bar{n}} + \underbrace{j \sum_{(m,n)\neq(\bar{m},\bar{n})} d_{m,n} \frac{H_m}{H_{\bar{m}}} \langle \xi \rangle^{\bar{m},\bar{n}}_{m,n}}_{\text{Interference}} \Bigg) + \eta_{\bar{m}\bar{n}}. \tag{3.18}$$

3.3.1 Data Detection

To begin with, let the transmission channel be ideal, i.e. $h[k] = \delta[k]$. Thus, the CFR $H_{\bar{m}} = 1$, and it follows from (3.18) that

$$y_{\bar{m},\bar{n}} = \left(d_{\bar{m},\bar{n}} + j \sum_{(m,n)\neq(\bar{m},\bar{n})} d_{m,n} \langle \xi \rangle^{\bar{m},\bar{n}}_{m,n} \right) + \eta_{\bar{m}\bar{n}}. \tag{3.19}$$

It follows from (3.10) that the quantity $\sum_{(m,n)\neq(\bar{m},\bar{n})} d_{m,n}\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}}$ is real in nature. Consequently, OQAM symbol $d_{\bar{m},\bar{n}}$ at the receiver can be estimated as

$$\hat{d}_{\bar{m},\bar{n}} = \Re\{y_{\bar{m},\bar{n}}\} = d_{\bar{m},\bar{n}} + \Re\{\eta_{\bar{m}\bar{n}}\}. \tag{3.20}$$

In practice, channel $h[k]$ is not ideal. Therefore, as shown in Fig. 3.2b, equalization operation is performed before detecting OQAM symbols. By performing the ZF equalization in (3.18), the estimate of OQAM symbol at the subcarrier $\bar{m}$ and symbol instant $\bar{n}$ is

$$\Re\left\{\frac{y_{\bar{m},\bar{n}}}{H_{\bar{m}}}\right\} = d_{\bar{m},\bar{n}} + \Re\{j\tilde{I}_{\bar{m},\bar{n}}\} + \Re\{\eta_{\bar{m}\bar{n}}\}, \tag{3.21}$$

where the quantity

$$\tilde{I}_{\bar{m},\bar{n}} = \sum_{(m,n)\neq(\bar{m},\bar{n})} d_{m,n}\frac{H_m}{H_{\bar{m}}}\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}} \tag{3.22}$$

is called the *intrinsic interference* in FBMC systems. Since the quantity $\frac{H_m}{H_{\bar{m}}}$ is complex, $\Re\{jI_{\bar{m},\bar{n}}\} \neq 0$. The term $\Re\{jI_{\bar{m},\bar{n}}\}$ characterises the ISI and inter-carrier-interference (ICI) between the transmitted OQAM symbols. This is unlike OFDM systems wherein the former is suppressed using the CP and the latter is mitigated using the orthogonality among the subcarriers. It sounds difficult to obtain a reliable estimate of the OQAM symbol $d_{\bar{m},\bar{n}}$ at this stage. However, introduction of few approximations by exploiting the inherent properties of FBMC systems leads to a reliable estimate of OQAM symbols as follows. The interference evaluation in (3.22) can be recast as

$$\tilde{I}_{\bar{m},\bar{n}} = \sum_{(m,n)\in\Omega_{\bar{m},\bar{n}}} d_{m,n}\frac{H_m}{H_{\bar{m}}}\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}} + \sum_{(m,n)\notin\Omega_{\bar{m},\bar{n}}} d_{m,n}\frac{H_m}{H_{\bar{m}}}\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}}, \tag{3.23}$$

where the symbol $\Omega_{\bar{m},\bar{n}}$ symbolises the neighbourhood of the desired symbol at subcarrier-symbol time index $(\bar{m},\bar{n})$ by excluding the index $(\bar{m},\bar{n})$. Since prototype filter $p[k]$ in FBMC is well localized both in frequency and time, the interference due to the FT points (m,n) outside the neighbourhood $\Omega_{\bar{m},\bar{n}}$ is negligible since $\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}}\big|_{(m,n)\notin\Omega_{\bar{m},\bar{n}}} \approx 0$ [26]. Furthermore, since the subcarrier bandwidth is significantly smaller than the coherence bandwidth of channel, the CFR is well approximated by a constant over the neighbourhood $\Omega_{\bar{m},\bar{n}}$ [26]. Typically, due to the associated sharp pulse-shaping filters in FBMC, the interference mainly arises from the first-order neighbourhood $\Omega_{\bar{m},\bar{n}} = \{(\bar{m}\pm1,\bar{n}\pm1),(\bar{m},\bar{n}\pm1),(\bar{m}\pm1,\bar{n})\}$. Upon using the above inherent properties of FBMC systems, the intrinsic interference in

(3.23) is well approximated as

$$\tilde{I}_{\bar{m},\bar{n}} \approx I_{\bar{m},\bar{n}} = \sum_{(m,n)\in\Omega_{\bar{m},\bar{n}}} d_{m,n}\langle\xi\rangle_{m,n}^{\bar{m},\bar{n}}, \tag{3.24}$$

which also implies that $\Re\{j\tilde{I}_{\bar{m},\bar{n}}\} \approx 0$. Employing the above result in (3.18), the received symbol $y_{\bar{m},\bar{n}}$ at the FT index $(\bar{m},\bar{n})$ can be recast as

$$y_{\bar{m},\bar{n}} \approx H_{\bar{m}} b_{\bar{m},\bar{n}} + \eta_{\bar{m}\bar{n}}. \tag{3.25}$$

Here the term $b_{\bar{m},\bar{n}} = d_{\bar{m},\bar{n}} + jI_{\bar{m},\bar{n}}$ is known as virtual symbol, which is the summation of the desired symbol $d_{\bar{m},\bar{n}}$ and the corresponding interference component $I_{\bar{m},\bar{n}}$. The model in (3.25) is widely popular with the name interference approximation model (IAM) [26], because it utilizes the property that each FBMC symbol interferes with the symbols in its small FT neighbourhood, over which the CFR can be approximated to a constant. With zero mean i.i.d. OQAM symbols, each of power P_d, $I_{\bar{m},\bar{n}}$ has a mean zero and power $\mathbb{E}[|I_{\bar{m},\bar{n}}|^2] \approx P_d$ [27]. Furthermore, the OQAM symbol $d_{\bar{m},\bar{n}}$ and the interference term $I_{\bar{m},\bar{n}}$ are zero mean independent. The virtual symbol $b_{\bar{m},\bar{n}}$ thus has mean zero and power

$$\mathbb{E}[|b_{\bar{m},\bar{n}}|^2] = \mathbb{E}[|d_{\bar{m},\bar{n}}|^2] + \mathbb{E}[|I_{\bar{m},\bar{n}}|^2] \approx 2P_d. \tag{3.26}$$

Following Fig. 3.2b, one can obtain a reliable estimate of OQAM symbols using the model in (3.25) as follows:

$$\hat{d}_{\bar{m},\bar{n}} = \Re\left\{\frac{y_{\bar{m},\bar{n}}}{H_{\bar{m}}}\right\} \approx d_{\bar{m},\bar{n}} + \Re\{\eta_{\bar{m}\bar{n}}\}. \tag{3.27}$$

The function $\Phi^{-1}(\hat{d}_{\bar{m},\bar{n}})$ in Fig. 3.2b at the $\bar{m}$th subcarrier, by combining the estimated OQAM symbols $\hat{d}_{m,2n}$ and $\hat{d}_{m,2n+1}$, obtains the estimate $\hat{c}_{m,n}$ of complex-valued QAM symbol as follows

$$\hat{c}_{m,n} = \begin{cases} \hat{d}_{m,2n} + j\hat{d}_{m,2n+1}, & m \text{ even} \\ \hat{d}_{m,2n+1} + j\hat{d}_{m,2n}, & m \text{ odd.} \end{cases} \tag{3.28}$$

The SISO-FBMC system model discussed above can be easily extended to MIMO with N_t transmit and N_r receive antennas. Similar to (3.7), the FBMC baseband signal $s^t[k]$ at the tth transmit antenna is expressed as

$$s^t[k] = \sum_{m=0}^{N-1}\sum_{n\in\mathbb{Z}} d_{m,n}^t \chi_{m,n}[k], \tag{3.29}$$

where $1 \leq t \leq N_t$, and $d^t_{m,n}$ is an OQAM symbol transmitted by tth antenna at the frequency-time point (m, n). The signal at the rth antenna of FBMC receiver is expressed as

$$y^r[k] = \sum_{t=1}^{N_t} \left(s^t[k] * h^{r,t}[k] \right) + \eta^r[k], \text{ for } 1 \leq r \leq N_r. \tag{3.30}$$

The scaler $h^{r,t}[k]$ above symbolises an L_h-tap channel impulse response between the tth transmit and the rth receive antenna pair. The noise $\eta^r[k]$ at the rth receive antenna obeys $\mathcal{CN}(0, \sigma^2_\eta)$. Following the procedure described in the SISO-FBMC system, the received signal on the rth antenna at subcarrier $\bar{m}$ and symbol instant $\bar{n}$th, after passing through the receive filter bank, is obtained as

$$y^r_{\bar{m},\bar{n}} \approx \sum_{t=1}^{N_t} H^{r,t}_{\bar{m}}\, b^t_{\bar{m},\bar{n}} + \eta^r_{\bar{m},\bar{n}}, \tag{3.31}$$

where the complex quantity $H^{r,t}_{\bar{m}}$ symbolises the CFR between the tth transmit and the rth receive antenna pair at the $\bar{m}$th subcarrier and is computed as $H^{r,t}_{\bar{m}} = \sum_{l=0}^{L_h-1} h^{r,t}[l] e^{-j2\pi\bar{m}l/N}$. The demodulated noise $\eta^r_{\bar{m},\bar{n}} = \sum_{k=-\infty}^{+\infty} \eta^r[k]\ \chi^*_{\bar{m},\bar{n}}[k]$ obeys $\mathcal{CN}(0, \sigma^2_\eta)$. The term $b^t_{\bar{m},\bar{n}} = d^t_{\bar{m},\bar{n}} + jI^t_{\bar{m},\bar{n}}$ is the virtual symbol for the tth transmit antenna, where the associated intrinsic interference $I^t_{\bar{m},\bar{n}}$ is given as

$$I^t_{\bar{m},\bar{n}} = \sum_{(m,n)\in\Omega_{\bar{m},\bar{n}}} d^t_{m,n} \langle \xi \rangle^{\bar{m},\bar{n}}_{m,n}. \tag{3.32}$$

For mathematical ease, (3.31) can be written in vector form as

$$\mathbf{y}_{\bar{m},\bar{n}} = \mathbf{H}_{\bar{m}} \mathbf{b}_{\bar{m},\bar{n}} + \boldsymbol{\eta}_{\bar{m},\bar{n}}, \tag{3.33}$$

where the vector $\boldsymbol{\eta}_{\bar{m},\bar{n}} = [\eta^1_{\bar{m},\bar{n}}, \eta^2_{\bar{m},\bar{n}}, \ldots, \eta^{N_r}_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{N_r \times 1}$ comprises noise with the covariance matrix $\mathbb{E}[\boldsymbol{\eta}_{\bar{m},\bar{n}} \boldsymbol{\eta}^H_{\bar{m},\bar{n}}] = \sigma^2_\eta \mathbf{I}_{N_r}$ and $\mathbf{y}_{\bar{m},\bar{n}} = [y^1_{\bar{m},\bar{n}}, y^2_{\bar{m},\bar{n}}, \ldots, y^{N_r}_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{N_r \times 1}$ is the observation symbol vector across N_r receive antennas. The virtual symbol vector $\mathbf{b}_{\bar{m},\bar{n}} = [b^1_{\bar{m},\bar{n}}, b^2_{\bar{m},\bar{n}}, \ldots, b^{N_t}_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{N_t \times 1}$ has a covariance matrix $\mathbb{E}[\mathbf{b}_{\bar{m},\bar{n}} \mathbf{b}^H_{\bar{m},\bar{n}}] \approx 2P_d \mathbf{I}_{N_t}$. The (r, t)th element of the MIMO CFR matrix $\mathbf{H}_{\bar{m}} \in \mathbb{C}^{N_r \times N_t}$ at the $\bar{m}$th subcarrier is given as $H^{r,t}_{\bar{m}}$. The observed symbols $y^1_{m,n}, y^2_{m,n}, \cdots, y^{N_r}_{m,n}$ are equalized, followed by the operation $\Re(\cdot)$, which pulls out the estimate of OQAM symbols from the estimated virtual data vector $\hat{\mathbf{b}}_{m,n} = [\hat{b}^1_{m,n}, \hat{b}^2_{m,n}, \cdots, \hat{b}^{N_t}_{m,n}]^T$. Finally, the vectors $\hat{\mathbf{d}}_{m,2n} = [\hat{d}^1_{m,2n}, \hat{d}^2_{m,2n}, \cdots, \hat{d}^{N_t}_{m,2n}]^T$ and $\hat{\mathbf{d}}_{m,2n+1} = [\hat{d}^1_{m,2n+1}, \hat{d}^2_{m,2n+1}, \cdots, \hat{d}^{N_t}_{m,2n+1}]^T$ on each subcarrier are clubbed

together for constructing QAM symbol vector as

$$\hat{\mathbf{c}}_{m,n} = \begin{cases} \hat{\mathbf{d}}_{m,2n+1} + j\hat{\mathbf{d}}_{m,2n}, & m \text{ odd} \\ \hat{\mathbf{d}}_{m,2n} + j\hat{\mathbf{d}}_{m,2n+1}, & m \text{ even.} \end{cases} \tag{3.34}$$

3.4 MIMO-FBMC Semi-Blind CSI Estimation

Channel state information (CSI) at receiver is typically acquired using pilot symbols, which do not carry information, and therefore, transmission of pilot symbols results in spectral efficiency loss. Due to the diversity gain offered by MIMO technology, the required signal-to-noise-ratio (SNR) for the desired bit-error-rate performance decreases. In such low SNR regimes, pilot-based CSI estimation methods demand large overheads for providing a reliable CSI estimate. Therefore, this section aims to develop a semi-blind channel estimation scheme for FBMC-based MIMO technology by exploiting the pilot symbols along with statistical properties of the information symbols. This results in a significant reduction in the mean square error (MSE) in comparison to its conventional pilot-based counterpart.

3.4.1 Review of Existing Works

There has been a significant research progress in the area of channel estimation for MIMO-FBMC systems. Reference [28] extended the concept of IAM model-based CFR estimation for FBMC-aided MIMO systems and proposed least squares (LS) estimator. Rottenberg et al. in [29] investigated a linear minimum mean square error (MMSE) CFR estimation approach for the downlink of a distributed MIMO-FBMC technology. Reference [30] described a method for pilot sequence design for the IAM model-based CSI estimation by employing zero-correlation zone sequences in MIMO-FBMC systems. Javaudin et al. designed a scattered pilot-based CSI estimator for MIMO-FBMC systems in [31]. The authors in [32] analysed the performance of the IAM model-based MMSE and LS CSI estimations for MIMO-FBMC waveform in the existence of imperfect channel correlations. Reference [24] provided a comprehensive review on training-based approaches for CSI estimation in FBMC-based SISO and MIMO systems. References [16, 33–35] designed training-based time-domain CSI estimation algorithms for MIMO-FBMC systems over a high frequency selective channel. Lin et al. in [36] designed a pilot-based compressive sensing technique in the time domain for MIMO-FBMC CSI estimation by utilising the generalized approximate message passing algorithm.

A key drawback of the above treatises is to employ only pilot symbols for CSI estimation, which leads to a decrease in the spectral efficiency. To avoid this shortcoming, the FBMC literature has developed blind CSI estimation schemes, which do not need pilot symbols for estimation [37, Chapter 11]. Blind schemes,

however, are usually computationally expansive and affected by poor convergence. For example, the authors in [38] investigated a blind CSI estimation technique for SISO-FBMC systems by utilising the cyclostationarity induced by the overlapping nature of pulse-shaping filters. The proposed technique therein demands a large number of data symbols to provide a reliable CSI estimate, and its performance deteriorates with an increase in the pulse-shaping filters' length. Savaux et al. in [39] designed blind equalization for FBMC-based SISO waveform by utilizing the concept of constant norm. The technique therein is restricted for square QAM constellations, and similar to [38], its convergence needs a large data length.

In presence of the above observations, semi-blind schemes, which significantly improve CSI estimation accuracy by utilising both statistical characteristics of the underlying system and a small training overhead, present a viable alternative to both pilot-based and blind CSI estimation methods. Authors in [40] conceived a semi-blind CSI estimation scheme for FBMC-based SISO systems, wherein the CSI magnitude and phase are estimated blindly by exploiting the subcarrier power and the spatial-sign covariance matrix, respectively, and pilot symbols are utilised to mitigate the sign ambiguity arising from the blind CSI estimation algorithm. The MSE of the blind estimation in [40] is sensitive to frame length, and therefore, its performance deteriorates as the frame length decreases [41]. Kofidis et al. in [42] proposed a tensor-based scheme for semi-blind CSI estimation and data detection for FBMC-based MIMO systems by using canonical polyadic decomposition (CPD). The results presented therein are restricted to a single-input multiple-output (SIMO)-FBMC system, because the CPD model is not always identifiable. Due to the effectiveness and improved performance of semi-blind schemes, this section develops a different semi-blind technique for CSI estimation in FBMC-based MIMO systems by exploiting both pilot and the statistical properties of data symbols. Contrary to the pilot-based CSI estimation schemes in [24, 26, 28–31, 33, 36, 43–46], the semi-blind MIMO-FBMC scheme leverages the pilot symbols along with blind data symbols for estimating the unitary and whitening components of the channel matrix $\mathbf{H}_m$. It is thus offered significantly lower MSE than the existing IAM model-based LS CSI estimation scheme that utilises only pilot symbols. The Cramer-Rao lower bounds (CRLBs) are derived to quantify the MSE gain of the semi-blind method over the pilot-based techniques.

3.4.2 Semi-Blind MIMO-FBMC Channel Estimator

Let the tth antenna transmit L_0 symbols on each of the subcarriers, as shown in the frame in Fig. 3.3. Each frame consists of M pilot symbols to be utilised for estimating CSI. The N_d symbols at the end of the frame carry data. Unlike OFDM, adjacent time-domain FBMC symbols interfere with each other because of the overlapping of the pulse-shaping filters. Thus, z zeros are placed between the adjacent pilot symbols to mitigate the ISI [24, 26]. Due to the inter-frame time gap commonly used in wireless systems, one does not need to insert zeros in the

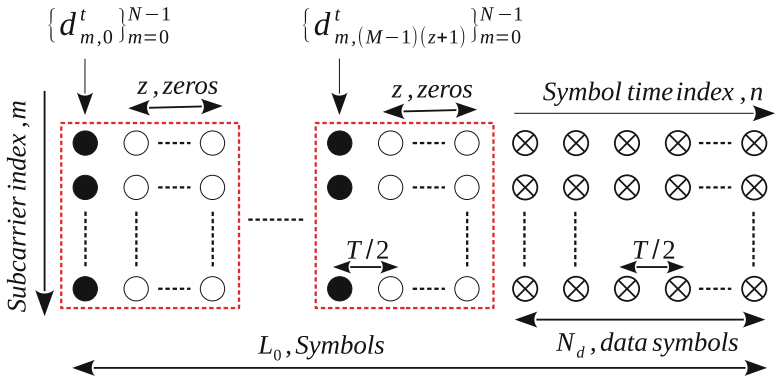


Fig. 3.3 Placement of symbols in the frame for the tth antenna at the FBMC transmitter. Here ●, ○ and ⊗, respectively, represent the pilot, zero and data symbols

beginning of the frame [24]. The effect of varying z on the MSE of the semi-blind and conventional pilot-based CSI estimation schemes is shown later using simulations, which demonstrates that employing $z = 1$ is sufficient for reducing the ISI to a tolerable level. This implies that MIMO-FBMC frame with zeros employs $2M$ OQAM symbols on each of the subcarriers for CSI estimation, which is equivalent to M QAM symbols [24]. The channel estimation pilot overhead in MIMO-FBMC is therefore same as that of MIMO-OFDM systems [47].

As shown in Fig. 3.3, the pilot symbols are located at $n = i(1+z)$ for $0 \leq i \leq M-1$. Evaluating (3.33) at these instants and stacking the resulting observations, one gets

$$\mathbf{Y}_{\bar{m}} = \mathbf{H}_{\bar{m}}\mathbf{B}_{\bar{m}} + \boldsymbol{\eta}_{\bar{m}}. \tag{3.35}$$

Here the virtual pilot matrix $\mathbf{B}_{\bar{m}} = [\mathbf{b}_{\bar{m},0}, \mathbf{b}_{\bar{m},(1+z)}, \ldots, \mathbf{b}_{\bar{m},(M-1)(1+z)}] \in \mathbb{C}^{N_t \times M}$, the observation matrix $\mathbf{Y}_{\bar{m}} = [\mathbf{y}_{\bar{m},0}, \mathbf{y}_{\bar{m},(1+z)}, \ldots, \mathbf{y}_{\bar{m},(M-1)(1+z)}] \in \mathbb{C}^{N_r \times M}$ and the noise matrix $\boldsymbol{\eta}_{\bar{m}} = [\boldsymbol{\eta}_{\bar{m},0}, \boldsymbol{\eta}_{\bar{m},(1+z)}, \ldots, \boldsymbol{\eta}_{\bar{m},(M-1)(1+z)}] \in \mathbb{C}^{N_r \times M}$. Let the elements of the CFR matrix $\mathbf{H}_{\bar{m}}$ obey $\mathcal{CN}(0, \sigma_h^2)$. The tth component of the pilot vector $\mathbf{b}_{\bar{m},i(1+z)} \in \mathbb{C}^{N_t \times 1}$ is obtained as $b^t_{\bar{m},i(1+z)} = d^t_{\bar{m},i(1+z)} + jI^t_{\bar{m},i(1+z)}$, where the interference $I^t_{\bar{m},i(1+z)}$ is obtained as

$$I^t_{\bar{m},i(1+z)} = \sum_{m \neq \bar{m}} d^t_{m,i(1+z)} \Im\Big\{ \sum_{l=-\infty}^{+\infty} p^2[l] e^{j(\phi_{m,0} - \phi_{\bar{m},0})} e^{j2\pi(m-\bar{m})l/N} \Big\}. \tag{3.36}$$

The MIMO-FBMC channel matrix $\mathbf{H}_{\bar{m}}$ of size $N_r \times N_t$ with $N_r \geq N_t$ can be decomposed as

$$\mathbf{H}_{\bar{m}} = \mathbf{W}_{\bar{m}}\mathbf{Q}^H_{\bar{m}}, \tag{3.37}$$

where $\mathbf{Q}_{\bar{m}} \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{W}_{\bar{m}} \in \mathbb{C}^{N_r \times N_t}$, and for $0 \leq \bar{m} \leq N-1$, are referred to as unitary (complex rotation) and the whitening (decorrelating) matrices, respectively.[1] Being a unitary matrix, $\mathbf{Q}_{\bar{m}}$ satisfies the constraint $\mathbf{Q}_{\bar{m}}\mathbf{Q}_{\bar{m}}^H = \mathbf{Q}_{\bar{m}}^H\mathbf{Q}_{\bar{m}} = \mathbf{I}_{N_t}$. Thus, it follows from (3.37) that

$$\mathbf{H}_{\bar{m}}\mathbf{H}_{\bar{m}}^H = \mathbf{W}_{\bar{m}}\mathbf{W}_{\bar{m}}^H. \tag{3.38}$$

The semi-blind CSI estimation scheme works on the principle that the whitening matrix $\mathbf{W}_{\bar{m}}$ can be estimated by exploiting the second-order statistical characteristics of $N_d = L_0 - M(1+z)$ data symbols in Fig. 3.3, and the unitary matrix $\mathbf{Q}_{\bar{m}}$ can be estimated using the M pilot symbols in the frame. The matrix $\mathbf{Q}_{\bar{m}}$ is parameterized by a few parameters. It can therefore be estimated accurately with a limited pilot overhead. On the other side, estimation of $\mathbf{W}_{\bar{m}}$ blindly using statistical properties of data symbols significantly enhances accuracy of the CSI estimation.

The covariance matrix $\mathbf{R}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} \in \mathbb{C}^{N_r \times N_r} = \mathbb{E}\left[\mathbf{y}_{\bar{m},\bar{n}}\mathbf{y}_{\bar{m},\bar{n}}^H\right]$ of the observation vectors $\mathbf{y}_{\bar{m},\bar{n}}$ in (3.33) is calculated as

$$\mathbf{R}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} = 2P_d\mathbf{H}_{\bar{m}}\mathbf{H}_{\bar{m}}^H + \sigma_\eta^2\mathbf{I}_{N_r}. \tag{3.39}$$

The above results follow from the identities $\mathbb{E}[\mathbf{b}_{\bar{m},\bar{n}}\mathbf{b}_{\bar{m},\bar{n}}^H] = 2P_d\mathbf{I}_{N_t}$ and $\mathbb{E}[\boldsymbol{\eta}_{\bar{m},\bar{n}}\boldsymbol{\eta}_{\bar{m},\bar{n}}^H] = \sigma_\eta^2\mathbf{I}_{N_r}$. One can rewrite (3.39) as

$$\mathbf{H}_{\bar{m}}\mathbf{H}_{\bar{m}}^H = \frac{\mathbf{R}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} - \sigma_\eta^2\mathbf{I}_{N_r}}{2P_d} = \mathbf{W}_{\bar{m}}\mathbf{W}_{\bar{m}}^H. \tag{3.40}$$

The whitening matrix $\mathbf{W}_{\bar{m}}$ can now be estimated blindly as

$$\widehat{\mathbf{W}}_{\bar{m}} = \widehat{\mathbf{U}}_{\bar{m}}\widehat{\boldsymbol{\Sigma}}_{\bar{m}}^{1/2}. \tag{3.41}$$

The matrices $\widehat{\boldsymbol{\Sigma}}_{\bar{m}}$ and $\widehat{\mathbf{U}}_{\bar{m}}$ are calculated using SVD as given below:

$$\widehat{\mathbf{U}}_{\bar{m}}\widehat{\boldsymbol{\Sigma}}_{\bar{m}}\widehat{\mathbf{U}}_{\bar{m}}^H = \text{SVD}\left(\frac{\widehat{\mathbf{R}}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} - \sigma_\eta^2\mathbf{I}_{N_r}}{2P_d}\right). \tag{3.42}$$

The matrix $\widehat{\mathbf{R}}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}}$ above, which represents the estimate of $\mathbf{R}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}}$, is obtained by using the observation vectors $\mathbf{y}_{\bar{m},\bar{n}}$ for $M(1+z) \leq n \leq L_0$ as $\widehat{\mathbf{R}}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} = \frac{1}{N_d}\sum_{\bar{n}=M(1+z)}^{L_0}\mathbf{y}_{\bar{m},\bar{n}}\mathbf{y}_{\bar{m},\bar{n}}^H$. Note that the estimate $\widehat{\mathbf{R}}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}} \rightarrow \mathbf{R}_{\mathbf{y}_{\bar{m}}\mathbf{y}_{\bar{m}}}$ with high probability as N_d increases [48]. After obtaining the estimate of $\mathbf{W}_{\bar{m}}$, the estimate

[1] Let the singular value decomposition (SVD) of $\mathbf{H}_{\bar{m}}$ be expressed as $\text{SVD}(\mathbf{H}_{\bar{m}}) = \mathbf{S}_{\bar{m}}\boldsymbol{\Gamma}_{\bar{m}}\mathbf{Q}_{\bar{m}}^H$. It is clear that one possible choice for $\mathbf{W}_{\bar{m}} = \mathbf{S}_{\bar{m}}\boldsymbol{\Gamma}_{\bar{m}}$ and the unitary matrix can be set as $\mathbf{Q}_{\bar{m}}$. It implies that the whitening unitary decomposition in (3.37) is guaranteed to exist.

of the unitary matrix $\mathbf{Q}_{\bar{m}}$ is obtained as a solution of the following optimization problem:

$$\min_{\mathbf{Q}_{\bar{m}}} \left\| \mathbf{Y}_{\bar{m}} - \widehat{\mathbf{W}}_{\bar{m}} \mathbf{Q}_{\bar{m}}^H \mathbf{B}_{\bar{m}} \right\|^2$$
$$\text{s.t. } \mathbf{Q}_{\bar{m}} \mathbf{Q}_{\bar{m}}^H = \mathbf{I}_{N_t}. \tag{3.43}$$

For an orthogonal pilot matrix $\mathbf{B}_{\bar{m}}$ ($\mathbf{B}_{\bar{m}} \mathbf{B}_{\bar{m}}^H = 2P_d M \mathbf{I}_{N_t}$) [49], the solution of the optimization problem above is expressed as [50]

$$\widehat{\mathbf{Q}}_{\bar{m}} = \widehat{\mathbf{V}}_{\mathbf{Q}_{\bar{m}}} \widehat{\mathbf{U}}_{\mathbf{Q}_{\bar{m}}}^H. \tag{3.44}$$

The matrices $\widehat{\mathbf{U}}_{\mathbf{Q}_{\bar{m}}}$ and $\widehat{\mathbf{V}}_{\mathbf{Q}_{\bar{m}}}$ above are obtained using the SVD as follows:

$$\widehat{\mathbf{U}}_{\mathbf{Q}_{\bar{m}}} \widehat{\mathbf{\Sigma}}_{\mathbf{Q}_{\bar{m}}} \widehat{\mathbf{V}}_{\mathbf{Q}_{\bar{m}}}^H = \text{SVD}\left(\widehat{\mathbf{W}}_{\bar{m}}^H \mathbf{Y}_{\bar{m}} \mathbf{B}_{\bar{m}}^H \right). \tag{3.45}$$

Upon employing the estimate of whitening and unitary matrices, the semi-blind estimate of CSI matrix $\mathbf{H}_{\bar{m}}$ is obtained as

$$\widehat{\mathbf{H}}_{\text{SB},\bar{m}} = \widehat{\mathbf{W}}_{\bar{m}} \widehat{\mathbf{Q}}_{\bar{m}}^H = \widehat{\mathbf{U}}_{\bar{m}} \widehat{\mathbf{\Sigma}}_{\bar{m}}^{1/2} \widehat{\mathbf{U}}_{\mathbf{Q}_{\bar{m}}} \widehat{\mathbf{V}}_{\mathbf{Q}_{\bar{m}}}^H. \tag{3.46}$$

Let $\mathbb{H} \in \mathbb{C}^{N_r \times N N_t}$ be the MIMO channel matrix, which is obtained as

$$\mathbb{H} = [\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1}]. \tag{3.47}$$

The semi-blind estimate of $\mathbb{H}$ can be calculated as

$$\widehat{\mathbb{H}}_{\text{SB}} = \left[\widehat{\mathbf{H}}_{\text{SB},0}, \widehat{\mathbf{H}}_{\text{SB},1}, \cdots, \widehat{\mathbf{H}}_{\text{SB},N-1} \right]. \tag{3.48}$$

The algorithmic form of the semi-blind scheme is given in Algorithm 1.

3.4.3 MSE Gain of the Semi-Blind Estimate over the LS Estimate

From (3.35), the pilot-based LS estimate of the CSI matrix $\mathbf{H}_{\bar{m}}$ is given as

$$\widehat{\mathbf{H}}_{\text{LS},\bar{m}} = \arg\min_{\mathbf{H}_{\bar{m}}} \| \mathbf{Y}_{\bar{m}} - \mathbf{H}_{\bar{m}} \mathbf{B}_{\bar{m}} \|^2.$$

One can obtain the LS estimate of the MIMO CSI matrix $\mathbf{H}_{\bar{m}}$ as [24]

$$\widehat{\mathbf{H}}_{\text{LS},\bar{m}} = \mathbf{Y}_{\bar{m}} \mathbf{B}_{\bar{m}}^{\dagger} = \mathbf{H}_{\bar{m}} + \boldsymbol{\eta}_{\bar{m}} \mathbf{B}_{\bar{m}}^{\dagger}. \tag{3.49}$$

Algorithm 1: Algorithmic form of the semi-blind estimator

Input: Virtual pilot matrix $\mathbf{B}_{\tilde{m}}$, Observation matrix $\mathbf{Y}_{\tilde{m}}$, Observation vectors $\mathbf{y}_{\tilde{m},\tilde{n}}$ for $M(1+z) \leq n \leq L_0$, Noise variance σ_η^2 and power P_d

Output: Estimate $\widehat{\mathbf{H}}_{\tilde{m}}$

1 Obtain the estimate $\widehat{\mathbf{R}}_{\mathbf{y}_{\tilde{m}}\mathbf{y}_{\tilde{m}}} = \frac{1}{N_d}\sum_{\tilde{n}=M(1+z)}^{L_0} \mathbf{y}_{\tilde{m},\tilde{n}}\mathbf{y}_{\tilde{m},\tilde{n}}^H$

2 Compute $\widehat{\mathbf{U}}_{\tilde{m}}\widehat{\mathbf{\Sigma}}_{\tilde{m}}\widehat{\mathbf{U}}_{\tilde{m}}^H = \text{SVD}\left(\frac{\widehat{\mathbf{R}}_{\mathbf{y}_{\tilde{m}}\mathbf{y}_{\tilde{m}}} - \sigma_\eta^2\mathbf{I}_{N_r}}{2P_d}\right)$

3 Obtain the estimate $\widehat{\mathbf{W}}_{\tilde{m}} = \widehat{\mathbf{U}}_{\tilde{m}}\widehat{\mathbf{\Sigma}}_{\tilde{m}}^{1/2}$

4 Compute $\widehat{\mathbf{U}}_{\mathbf{Q}_{\tilde{m}}}\widehat{\mathbf{\Sigma}}_{\mathbf{Q}_{\tilde{m}}}\widehat{\mathbf{V}}_{\mathbf{Q}_{\tilde{m}}}^H = \text{SVD}\left(\widehat{\mathbf{W}}_{\tilde{m}}^H\mathbf{Y}_{\tilde{m}}\mathbf{B}_{\tilde{m}}^H\right)$

5 Obtain the estimate $\widehat{\mathbf{Q}}_{\tilde{m}} = \widehat{\mathbf{V}}_{\mathbf{Q}_{\tilde{m}}}\widehat{\mathbf{U}}_{\mathbf{Q}_{\tilde{m}}}^H$

6 Obtain the estimate $\widehat{\mathbf{H}}_{\tilde{m}} = \widehat{\mathbf{W}}_{\tilde{m}}\widehat{\mathbf{Q}}_{\tilde{m}}^H$

7 **return**: The estimate $\widehat{\mathbf{H}}_{\tilde{m}}$

Here the operation $\mathbf{B}_{\tilde{m}}^\dagger = \mathbf{B}_{\tilde{m}}^H\left(\mathbf{B}_{\tilde{m}}\mathbf{B}_{\tilde{m}}^H\right)^{-1}$ gives the pseudo-inverse of $\mathbf{B}_{\tilde{m}}$ [51]. As shown in [52], the MSE of the LS CSI estimator is

$$\mathbb{E}\left[\left\|\widehat{\mathbf{H}}_{\text{LS},\tilde{m}} - \mathbf{H}_{\tilde{m}}\right\|_F^2\right] = \frac{\sigma_\eta^2 N_t N_r}{4P_d M}. \tag{3.50}$$

It is worth mentioning that the CRLB of the LS estimator equals the channel estimation error covariance, because $\widehat{\mathbf{H}}_{\text{LS},\tilde{m}}$ is the minimum variance unbiased estimate. The LS estimate of the MIMO CSI matrix in (3.47) is

$$\widehat{\mathbb{H}}_{\text{LS}} = \left[\widehat{\mathbf{H}}_{\text{LS},0}, \widehat{\mathbf{H}}_{\text{LS},1}, \cdots, \widehat{\mathbf{H}}_{\text{LS},N-1}\right]. \tag{3.51}$$

The MSE in $\mathbb{H}_{\text{LS}}$ is calculated as

$$\mathbb{E}\left[\left\|\widehat{\mathbb{H}}_{\text{LS}} - \mathbb{H}\right\|_F^2\right] = \frac{\sigma_\eta^2 N N_t N_r}{2P_d M}.$$

The CRLB for the channel estimation per parameter using the LS CSI estimator is determined as $\frac{\sigma_\eta^2 N N_t N_r}{2P_d M}\left(\frac{1}{N_t N_r}\right)$, which can be seen to remain unchanged with N_t and N_r.

Coming to the semi-blind estimator, the unitary matrix $\mathbf{Q}_{\tilde{m}}$ constrained as $\mathbf{Q}_{\tilde{m}}^H\mathbf{Q}_{\tilde{m}} = \mathbf{I}_{N_t}$. Thus, one can utilize the constrained CRLB framework in [53] for benchmarking the MSE of the semi-blind estimator. To begin with, the whitening matrix $\mathbf{W}_{\tilde{m}}$ can be assumed to be known at the receiver. This assumption, as shown in [52], holds well with the transmission of a few hundred OQAM data symbols, because the accuracy for estimating the matrix $\mathbf{W}_{\tilde{m}}$ is sufficiently high. It follows from [52] that CRLB for the MSE of the semi-blind estimator for (l,k)th element

$\mathbf{H}_{\bar{m}}(k,l)$ of the CSI matrix $\mathbf{H}_{\bar{m}}$ is expressed as

$$\mathbb{E}\big[|\widehat{\mathbf{H}}_{\mathrm{SB},\bar{m}}(k,l)-\mathbf{H}_{\bar{m}}(k,l)|^2\big] \geq \frac{\sigma_\eta^2}{2P_dM}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t}\frac{\sigma_{\bar{m},i}^2}{\sigma_{\bar{m},j}^2+\sigma_{\bar{m},i}^2}|\mathbf{S}_{\bar{m}}(k,i)|^2|\mathbf{Q}_{\bar{m}}(l,j)|^2, \tag{3.52}$$

where $\mathbf{Q}_{\bar{m}}(l,j)$ and $\mathbf{S}_{\bar{m}}(k,i)$, respectively, represent the (l,j)th element of $\mathbf{Q}_{\bar{m}}$ the (k,i)th element of $\mathbf{S}_{\bar{m}}$. It is important to note that the weighting factor $\sigma_{\bar{m},i}^2/(\sigma_{\bar{m},j}^2+\sigma_{\bar{m},i}^2)$ gives the net reduction in the CSI estimation error in comparison to that of the pilot-aided LS CSI estimator. Furthermore, it follows from [50] that the minimum error for estimating $\mathbf{H}_{\bar{m}}$ by using the semi-blind technique is bounded as

$$\mathbb{E}\left[\left\|\widehat{\mathbf{H}}_{\mathrm{SB},\bar{m}}-\mathbf{H}_{\bar{m}}\right\|_F^2\right] \geq \frac{\sigma_\eta^2}{4P_dM}\mathcal{N}_\theta = \frac{\sigma_\eta^2 N_t^2}{4P_dM}, \tag{3.53}$$

where $N_\theta = N_t^2$ gives the number of real parameters required to parameterize $\mathbf{Q}_{\bar{m}}$ [54]. It also implies that when the estimation of the whitening matrix $\mathbf{W}_{\bar{m}}$ is performed accurately by using the blind information of data symbols, the CSI matrix $\mathbf{H}_{\bar{m}} = \mathbf{W}_{\bar{m}}\mathbf{Q}_{\bar{m}}^H$ also requires estimation of only N_t^2 parameters. These parameters can be estimated with high degree of accuracy by using a limited pilot overhead. The constraint CRLB for the estimation of the CSI matrix in (3.48) is given as

$$\mathbb{E}\left[\left\|\widehat{\mathbb{H}}_{\mathrm{SB}}-\mathbb{H}\right\|_F^2\right] = \sum_{\bar{m}=0}^{N-1}\mathbb{E}\left[\left\|\widehat{\mathbf{H}}_{\mathrm{SB},\bar{m}}-\mathbf{H}_{\bar{m}}\right\|_F^2\right] \geq \frac{\sigma_\eta^2 N N_t^2}{4P_dM}.$$

The CRLB for CSI estimation per parameter is $\geq \frac{\sigma_\eta^2 N N_t^2}{4P_dM}\left(\frac{1}{N_tN_r}\right)$, which reduces with increasing the number of receive antennas N_r. Upon using (3.53) and (3.50), the MSE gain of the semi-blind technique over the pilot-based LS technique is

$$\mathcal{G} \leq \mathbb{E}\left[\left\|\widehat{\mathbf{H}}_{\mathrm{LS},\bar{m}}-\mathbf{H}_{\bar{m}}\right\|_F^2\right] \Big/ \mathbb{E}\left[\left\|\widehat{\mathbf{H}}_{\mathrm{SB},\bar{m}}-\mathbf{H}_{\bar{m}}\right\|_F^2\right] = \frac{2N_r}{N_t}. \tag{3.54}$$

Since $N_r \geq N_t$, it is immediately clear that $\frac{2N_r}{N_t} \geq 2$. It follows from (3.54) that for square MIMO-FBMC system that has $N_r = N_t$, the semi-blind technique outperforms the LS technique up to 3 dB. Furthermore, the MSE gain $\mathcal{G}$ can be increased by increasing the number of receive antennas N_r.

Figure 3.4a shows numerical results to verify MSE gain of the semi-blind (SB) technique over the conventional LS scheme quantified in (3.54). For this study, an $N_t \times N_r$ MIMO-FBMC system with the number of subcarriers $N = 128$ is considered over a type A Rayleigh fading channel for vehicular scenarios that has $L_h = 6$ taps, with the delay profile (in ns) 0, 310, 710, 1090, 1730, 2510 and power profile (in dB) 0.0, −1.0, −9.0, −10.0, −15.0, −20.0 between the (t,r)th

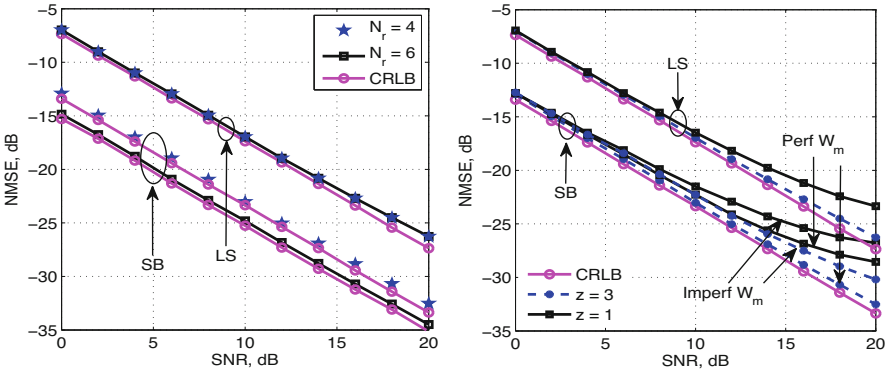


Fig. 3.4 (**a**) MSE gain of the semi-blind (SB) CSI estimator (perfect $\mathbf{W}_{\tilde{m}}$) over the pilot-aided LS CSI estimator with $N = 128$, $z = 3$ and $M = N_t = 2$ and (**b**) MSE comparison of the SB (perfect and imperfect $\mathbf{W}_{\tilde{m}}$) and LS CSI estimators with $N_r = 4$, $M = N_t = 2$ and $N_d = 320$

transmit-receive antenna pair. The isotropic orthogonal transform algorithm (IOTA) prototype pulse [15] of duration $4T$ is used for pulse shaping in FBMC system. Pilot and data symbols are drawn from the real and imaginary parts of 4-QAM symbols. The SNR on each subcarrier is calculated as $2P_d/\sigma_\eta^2$. It can be seen from Fig. 3.4a that the semi-blind CSI estimation technique for 2×4 and 2×6 MIMO-FBMC systems can be observed to provide MSE gain of 6 dB and 7.78 dB over the pilot only LS CSI estimator, respectively. This happens because the MSE per parameter for the latter does not change with N_r, whereas it decreases with N_r for the former. This NMSE trend validates the CRLB analysis presented in Section 3.4.3. It is also seen that both SB and LS techniques achieve their respective CRLBs.

Figure 3.4b shows the NMSE as a function SNR for the SB and LS CSI estimators. The graphs in the presence of perfect and imperfect knowledge of whitening matrix are marked as Perf $\mathbf{W}_m$ and Imperf $\mathbf{W}_m$, respectively. For the Perf $\mathbf{W}_m$ case, $\mathbf{W}_{\tilde{m}} = \mathbf{S}_{\tilde{m}}\mathbf{\Gamma}_{\tilde{m}}$, where $\mathbf{S}_{\tilde{m}}\mathbf{\Gamma}_{\tilde{m}}\mathbf{Q}_{\tilde{m}}^H$ is obtained using the SVD of the CSI matrix $\mathbf{H}_{\tilde{m}}$. For the Imperf $\mathbf{W}_{\tilde{m}}$ case, $\mathbf{W}_{\tilde{m}}$ is estimated using the second-order statistical characteristics of $N_d = 320$ OQAM data symbols. The SB CSI estimator in the presence of imperfect $\mathbf{W}_{\tilde{m}}$ can be seen to perform close to its perfect $\mathbf{W}_{\tilde{m}}$ counterpart. It can also be seen that the NMSE of both the LS and SB techniques with $z = 1$ floors at high SNR. This happens because the ISI between the pilot symbols dominates at the high SNR. As z increases to 3, both the CSI estimation methods receiver NMSE improvement due to the reduced effect of the ISI. Furthermore, the SB CSI estimator in the presence of imperfect $\mathbf{W}_{\tilde{m}}$ achieves a performance close to its CRLB. This shows that fixing $z = 3$ gives the ideal spectral efficiency versus NMSE trade-off, because the CRLB is nothing but the best possible estimation performance of an estimator.

3.5 Performance of FBMC Waveform in Uplink of Massive MIMO

In recent years, massive MIMO technology has received widespread popularity because of its ability to support a large number of users with high throughput [55]. By using a few hundred antennas at the base station, the massive MIMO technology achieves the favourable propagation, which mitigates the co-channel interference by employing linear receivers, namely, ZF, MMSE and maximum ratio combining (MRC). This leads to a significant enhancement in spectral efficiency. OFDM is being used with massive MIMO technology for 5G deployment. However, OFDM-based systems are susceptible to practical impairments associated with carrier and timing offsets, particularly in the uplink wherein it is difficult to track the Doppler spreads experienced by different users [5, 6]. The aim of this section is to analyse the uplink performance of FBMC waveform in the context of MU massive MIMO technology.

3.5.1 Review of Existing Works

Reference [56] showed that the signal-to-interference-plus-noise ratio (SINR) of FBMC waveform-based massive MIMO technology over frequency-selective channel saturates to a deterministic value that depends on the correlation between the channel impulse responses and weights of multi-antenna combining taps. Aminjavaheri et al. in [21] developed an equalizer to mitigate the SINR saturation in [56]. References [57, 58] compared OFDM and FBMC waveforms in the massive MIMO setup and highlight the benefits of the latter over the former in terms of (i) sensitivity to CFO; (ii) peak-to-average power ratio; and (iii) increased bandwidth efficiency. The work in [22] used the FBMC waveform for the uplink transmission in massive MIMO technology and derived the MSE of the estimated symbols for the MMSE, ZF and matched filtering receivers. Reference [49] derived the uplink achievable rate for FBMC-based multi-user multi-cell massive MIMO systems relying on the ZF and MRC receivers. The pertinent power scaling laws in the existence of perfect and imperfect receiver CSIs have also been derived therein. Recently, reference [59] investigated the uplink sum rate performance of multi-cell massive MIMO-FBMC systems over Rician fading channels. The studies reviewed above show that FBMC waveform in the context of massive MIMO technology has got significant attention. This section analyses the performance of FBMC waveform in the uplink of MU massive MIMO technology, in terms of uplink sum rates of the MRC and ZF receivers in the existence of perfect and imperfect CSIs. The lower bounds on the uplink sum rates of the MRC and ZF receivers are also derived, followed by the corresponding power scaling laws. Numerical examples are presented to (i) verify the analysis and (ii) compare the performance of OFDM- and FBMC-based MU massive MIMO technologies.

3.5.2 Massive MIMO-FBMC System Model

An uplink of FBMC waveform-based single-cell MU massive MIMO technology is considered that has N subcarriers, U single-antenna users and a BS comprising L antennas with $1 \ll U \ll L$. The U users communicate to the BS using the same time-frequency resources. An OQAM symbol transmitted by the uth user at FT index (m, n) is denoted by $d^u_{m,n}$. The OQAM symbols are generated from QAM symbols as per the rules given in (3.3) and (3.4). The in-phase and quadrature components of QAM symbol $c^u_{m,n}$ are assumed to be spatially and temporally i.i.d. with power P_d. This implies that $\mathbb{E}\big[d^u_{m,n}\left(d^u_{m,n}\right)^*\big] = P_d$ and $\mathbb{E}\big[c^u_{m,n}\left(c^u_{m,n}\right)^*\big] = 2P_d$. The signal on the lth BS antenna at frequency index $\bar{m}$ and symbol instant $\bar{n}$ is obtained as

$$y^l_{\bar{m},\bar{n}} = \sum_{u=1}^{U} G^{l,u}_{\bar{m}}\, b^u_{\bar{m},\bar{n}} + \eta^l_{\bar{m},\bar{n}}, \tag{3.55}$$

where the complex quantity $G^{l,u}_{\bar{m}} = \sum_{k=0}^{L_h-1} g^{l,u}[k] e^{-j2\pi\bar{m}k/N}$, between the uth user in the cell and the lth antenna of the base station, represents the CFR at subcarrier $\bar{m}$. The term $g^{l,u}[k]$ is the corresponding L_h-tap channel impulse response. The demodulated noise $\eta^l_{\bar{m},\bar{n}}$ at the lth BS antenna obeys $\mathcal{CN}(0, \sigma^2_\eta)$. The quantity $b^u_{\bar{m},\bar{n}} = d^u_{\bar{m},\bar{n}} + jI^u_{\bar{m},\bar{n}}$ is the virtual symbol for the uth user. The interference $I^u_{\bar{m},\bar{n}}$ is expressed by (3.24). For convenience, (3.55) can be expressed in vector form as

$$\mathbf{y}_{\bar{m},\bar{n}} = \mathbf{G}_{\bar{m}}\mathbf{b}_{\bar{m},\bar{n}} + \boldsymbol{\eta}_{\bar{m},\bar{n}}, \tag{3.56}$$

where $\mathbf{y}_{\bar{m},\bar{n}} = [y^1_{\bar{m},\bar{n}}, y^2_{\bar{m},\bar{n}}, \ldots, y^L_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{L\times 1}$ comprises observed symbols across L antennas of the BS, while $\boldsymbol{\eta}_{\bar{m},\bar{n}} = [\eta^1_{\bar{m},\bar{n}}, \eta^2_{\bar{m},\bar{n}}, \ldots, \eta^L_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{L\times 1}$ is the noise vector such that $\mathbb{E}[\boldsymbol{\eta}_{\bar{m},\bar{n}}\boldsymbol{\eta}^H_{\bar{m},\bar{n}}] = \sigma^2_\eta \mathbf{I}_L$. The vector $\mathbf{b}_{\bar{m},\bar{n}} = [b^1_{\bar{m},\bar{n}}, b^2_{\bar{m},\bar{n}}, \ldots, b^U_{\bar{m},\bar{n}}]^T \in \mathbb{C}^{U\times 1}$ consists of virtual symbol of the U users such that $\mathbb{E}[\mathbf{b}_{\bar{m},\bar{n}}\mathbf{b}^H_{\bar{m},\bar{n}}] \approx 2P_d\mathbf{I}_U$. The matrix $\mathbf{G}_{\bar{m}} = [\mathbf{g}^1_{\bar{m}}, \mathbf{g}^2_{\bar{m}}, \ldots, \mathbf{g}^U_{\bar{m}}] \in \mathbb{C}^{L\times U}$ is the CSI matrix between the U users in the cell and the base station. The matrix $\mathbf{G}_{\bar{m}}$ is commonly modelled as [60]

$$\mathbf{G}_{\bar{m}} = \mathbf{H}_{\bar{m}}\left\{\text{diag}[\beta^1, \beta^2, \cdots, \beta^U]\right\}^{1/2} = \mathbf{H}_{\bar{m}}\mathbf{D}^{1/2}, \tag{3.57}$$

where the diagonal matrix $\mathbf{D} = \text{diag}(\beta^1, \beta^2, \cdots, \beta^U) \in \mathbb{R}^{U\times U}$ with β^u being the large-scale fading coefficient for the uth user. The quantity β^u, which remains same for many coherence time intervals, can be assumed to be known at the base station. Also, it is assumed to be independent from BS antennas and subcarriers. The matrix $\mathbf{H}_{\bar{m}} = [\mathbf{h}^1_{\bar{m}}, \mathbf{h}^2_{\bar{m}}, \ldots, \mathbf{h}^U_{\bar{m}}] \in \mathbb{C}^{L\times U}$ consists of small-scale fading factors from the U users in the cell to the base station. The entries of $\mathbf{H}_{\bar{m}}$ are i.i.d. as $\mathcal{CN}(0, 1)$.

3.5.3 Uplink Sum Rate for Massive MIMO-FBMC with Imperfect CSI

The combiner matrix $\mathbf{A}_{\bar{m}}$ for the ZF and MRC receivers, which are widely used due to their low complexity and linear nature, is

$$\mathbf{A}_{\bar{m}} = \begin{cases} \mathbf{G}_{\bar{m}} & \text{for MRC} \\ \mathbf{G}_{\bar{m}} \left(\mathbf{G}_{\bar{m}}^H \mathbf{G}_{\bar{m}} \right)^{-1} & \text{for ZF.} \end{cases} \tag{3.58}$$

OQAM symbols at output of the combiner $\mathbf{A}_{\bar{m}}$ are obtained as $\hat{\mathbf{d}}_{\bar{m},\bar{n}}^u = \Re \left\{ \mathbf{A}_{\bar{m}}^H \mathbf{y}_{\bar{m},\bar{n}} \right\}$. The CSI matrix $\mathbf{G}_{\bar{m}}$, in practice, is estimated at the BS using pilot symbols. Reference [49] derived the uplink achievable rate for FBMC-based multi-user multi-cell massive MIMO systems by considering the effect of pilot contamination. This chapter considers single-cell multi-user massive MIMO-FBMC systems, wherein users transmit orthogonal pilot symbols for channel estimation at the base station. Thus, there is no pilot contamination. Let each user in the cell transmit M pilot symbols on each subcarrier for uplink channel estimation, as shown in Fig. 3.3. As explained in [49], the pilot-aided MMSE estimate of channel at subcarrier $\bar{m}$ from the uth user to the BS is obtained as

$$\hat{\mathbf{g}}_{\bar{m}}^u = \frac{\beta^u}{P_p \beta^u + \sigma_\eta^2} \mathbf{y}_{\bar{m}}^u,$$

where $\mathbf{y}_{\bar{m}}^u$ is the $L \times 1$ observed pilot vector at the subcarrier $\bar{m}$ of the uth user and $P_p = 2P_d M$ represents the pilot power. Let $\mathbf{e}_{\bar{m}}^u = \mathbf{g}_{\bar{m}}^u - \hat{\mathbf{g}}_{\bar{m}}^u$ be the channel estimation error vector for the uth user. One can show that

$$\mathbb{E}[\hat{\mathbf{g}}_{\bar{m}}^u (\hat{\mathbf{g}}_{\bar{m}}^u)^H] = \frac{P_p (\beta^u)^2}{P_p \beta^u + \sigma_\eta^2} \mathbf{I}_L \tag{3.59}$$

$$\mathbb{E}[\mathbf{e}_{\bar{m}}^u (\mathbf{e}_{\bar{m}}^u)^H] = \frac{\beta^u \sigma_\eta^2}{P_p \beta^u + \sigma_\eta^2} \mathbf{I}_L. \tag{3.60}$$

3.5.3.1 MRC Receiver

Employing $\mathbf{g}_{\bar{m}}^u = \hat{\mathbf{g}}_{\bar{m}}^u + \mathbf{e}_{\bar{m}}^u$, the expression in (3.56) can be expanded as

$$\mathbf{y}_{\bar{m},\bar{n}} = \hat{\mathbf{g}}_{\bar{m}}^u b_{\bar{m},\bar{n}}^u + \sum_{j=1, j \neq u}^{U} \hat{\mathbf{g}}_{\bar{m}}^j b_{\bar{m},\bar{n}}^j + \sum_{j=1}^{U} \mathbf{e}_{\bar{m}}^j b_{\bar{m},\bar{n}}^j + \boldsymbol{\eta}_{\bar{m},\bar{n}}.$$

The OQAM symbol estimate at subcarrier $\bar{m}$ and symbol index $\bar{n}$ for the uth user at the output of MRC receiver can be formulated as

$$\hat{d}^{u}_{\bar{m},\bar{n}} = \Re\{(\hat{\mathbf{g}}^{u}_{\bar{m}})^{H}\mathbf{y}_{\bar{m},\bar{n}}\} = \left\|\hat{\mathbf{g}}^{u}_{\bar{m}}\right\|^{2} d^{u}_{\bar{m},\bar{n}} + v^{u,\mathrm{mrc}}_{\bar{m},\bar{n}}.$$

Here the noise-plus-interference $v^{u,\mathrm{mrc}}_{\bar{m},\bar{n}}$ is

$$v^{u,\mathrm{mrc}}_{\bar{m},\bar{n}} = \Re\left\{\sum_{j=1,j\neq u}^{U}(\hat{\mathbf{g}}^{u}_{\bar{m}})^{H}\hat{\mathbf{g}}^{j}_{\bar{m}}b^{j}_{\bar{m},\bar{n}} + \sum_{j=1}^{U}(\hat{\mathbf{g}}^{u}_{\bar{m}})^{H}\mathbf{e}^{j}_{\bar{m}}b^{j}_{\bar{m},\bar{n}} + (\hat{\mathbf{g}}^{u}_{\bar{m}})^{H}\boldsymbol{\eta}_{\bar{m},\bar{n}}\right\}.$$

Upon employing (3.34), the QAM symbol estimate at the MRC receiver output is

$$\hat{c}^{u}_{\bar{m},\bar{n}} = \left\|\hat{\mathbf{g}}^{u}_{\bar{m}}\right\|^{2} c^{u}_{\bar{m},\bar{n}} + \tilde{v}^{u,\mathrm{mrc}}_{\bar{m},\bar{n}}, \tag{3.61}$$

where we have $c^{u}_{\bar{m},\bar{n}} = d^{u}_{\bar{m},2\bar{n}} + jd^{u}_{\bar{m},2\bar{n}+1}$ and $\tilde{v}^{u,\mathrm{mrc}}_{\bar{m},\bar{n}} = v^{u,\mathrm{mrc}}_{\bar{m},2\bar{n}} + jv^{u,\mathrm{mrc}}_{\bar{m},2\bar{n}+1}$ when $\bar{m}$ is even, and for odd $\bar{m}$, $c^{u}_{\bar{m},\bar{n}} = d^{u}_{\bar{m},2\bar{n}+1} + jd^{u}_{\bar{m},2\bar{n}}$ and $\tilde{v}^{u,\mathrm{mrc}}_{\bar{m},\bar{n}} = v^{u,\mathrm{mrc}}_{\bar{m},2\bar{n}+1} + jv^{u,\mathrm{mrc}}_{\bar{m},2\bar{n}}$. As shown in [49], the uth user SINR at subcarrier $\bar{m}$ in the existence of imperfect knowledge of channel is determined as

$$\Upsilon^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}} = \frac{2P_d\left\|\hat{\mathbf{g}}^{u}_{\bar{m}}\right\|^{2}}{2P_d\left(\sum_{j=1,j\neq u}^{U}\left|\tilde{\mathbf{g}}^{j}_{\bar{m}}\right|^{2} + \sum_{j=1}^{U}\frac{\beta^{j}\sigma_\eta^2}{P_p\beta^{j}+\sigma_\eta^2}\right) + \sigma_\eta^2},$$

where the random variable $\tilde{\mathbf{g}}^{j}_{\bar{m}}$ obeys $\tilde{\mathbf{g}}^{j}_{\bar{m}} = (\hat{\mathbf{g}}^{u}_{\bar{m}})^{H}\hat{\mathbf{g}}^{j}_{\bar{m}}/\left\|\hat{\mathbf{g}}^{u}_{\bar{m}}\right\|$. The uplink rate for the uth user is determined as

$$\mathcal{R}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}} = \mathbb{E}\left[\log_2(1+\Upsilon^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}})\right].$$

Upon utilizing Jensen's inequality $\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$ along with the convexity of $\log(1+\frac{1}{x})$, as shown in [49], the uplink rate of the uth user is lower bounded as

$$\mathcal{R}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}} \geq \tilde{\mathcal{R}}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}} = \log_2\left(1 + \frac{P_p(L-1)(\beta^{u})^2}{(P_p\beta^{u}+\sigma_\eta^2)\left(\sum_{j=1,j\neq u}^{U}\beta^{j} + \frac{\sigma_\eta^2}{2P_d}\right) + \beta^{u}\sigma_\eta^2}\right). \tag{3.62}$$

By setting $2P_d = E_u/\sqrt{L}$ for a fixed E_u, and $L \to \infty$, one obtains

$$\tilde{\mathcal{R}}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{IP}} \to \log_2\left(1 + ME_u^2(\beta^{u})^2/\sigma_\eta^4\right). \tag{3.63}$$

3.5.3.2 ZF Receiver

The estimated OQAM symbol vector at the output of ZF receiver in the existence of imperfect channel knowledge is given as

$$\hat{\mathbf{d}}_{\bar{m},\bar{n}} = \Re\{\hat{\mathbf{G}}_{\bar{m}}^{\dagger}\mathbf{y}_{\bar{m},\bar{n}}\} = \mathbf{d}_{\bar{m},\bar{n}} + \mathbf{v}_{\bar{m},\bar{n}}^{\mathrm{zf}}.$$

Here the noise plus interference vector $\mathbf{v}_{\bar{m},\bar{n}}^{\mathrm{zf}} = \Re\{\hat{\mathbf{G}}_{\bar{m}}^{\dagger}\sum_{j=1}^{U}\mathbf{e}_{\bar{m}}^{j}b_{\bar{m},\bar{n}}^{j} + \hat{\mathbf{G}}_{\bar{m}}^{\dagger}\boldsymbol{\eta}_{\bar{m},\bar{n}}\}$. It follows from [49] that the SINR for the uth user can be determined as

$$\Upsilon_{\bar{m},\mathrm{IP}}^{u,\mathrm{zf}} = \frac{2P_d}{\left(2P_d\sum_{j=1}^{U}\frac{\beta^j\sigma_\eta^2}{P_p\beta^j+\sigma_\eta^2} + \sigma_\eta^2\right)\left[\left(\hat{\mathbf{G}}_{\bar{m}}^{H}\hat{\mathbf{G}}_{\bar{m}}\right)^{-1}\right]_{u,u}}.$$

The operation $\left[\left(\hat{\mathbf{G}}_{\bar{m}}^{H}\hat{\mathbf{G}}_{\bar{m}}\right)^{-1}\right]_{u,u}$ above extracts the uth diagonal element of the matrix $\left(\hat{\mathbf{G}}_{\bar{m}}^{H}\hat{\mathbf{G}}_{\bar{m}}\right)^{-1}$. As given in [49], the uplink rate of the ZF receiver for the uth user is lower bound as

$$\mathcal{R}_{\bar{m},\mathrm{IP}}^{u,\mathrm{zf}} \geq \tilde{\mathcal{R}}_{\bar{m},\mathrm{IP}}^{u,\mathrm{zf}} = \log_2\left(1 + \frac{P_p(L-U)(\beta^u)^2}{(P_p\beta^u+\sigma_\eta^2)\left(\sum_{j=1}^{U}\frac{\beta^j\sigma_\eta^2}{P_p\beta^j+\sigma_\eta^2} + \frac{\sigma_\eta^2}{2P_d}\right)}\right). \tag{3.64}$$

Note that for $2P_d = E_u/\sqrt{L}$ and $L \to \infty$, $\tilde{\mathcal{R}}_{\bar{m},\mathrm{IP}}^{u,\mathrm{zf}} \to \mathcal{R}_{\bar{m},\mathrm{IP}}^{u,\mathrm{zf}}$. It is important to note that similar to OFDM-based massive MIMO systems, the uplink power scaling laws also hold for their FBMC counterparts in the existence of imperfect channel knowledge.

3.5.4 Uplink Sum Rate for Massive MIMO-FBMC with Perfect CSI

The uplink rate for the ZF and MRC receiver processing at the BS in the existence of perfect channel knowledge can be derived as special case of their imperfect CSI counterparts as follows.

3.5.4.1 MRC Receiver

The SINR at the output of the MRC receiver for the uth user is

$$\Upsilon_{\bar{m},\mathrm{P}}^{u,\mathrm{mrc}} = \frac{2P_d\left\|\mathbf{g}_{\bar{m}}^{u}\right\|^4}{2P_d\sum_{i=1,i\neq u}^{U}\left|(\mathbf{g}_{\bar{m}}^{u})^H\mathbf{g}_{\bar{m}}^{i}\right|^2 + \sigma_\eta^2\left\|\mathbf{g}_{\bar{m}}^{u}\right\|^2}. \tag{3.65}$$

The lower bound on the uplink rate is

$$\mathcal{R}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{P}} \geq \tilde{\mathcal{R}}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{P}} = \log_2\left(1 + \frac{2P_d(L-1)\beta^u}{2P_d\sum_{i=1,i\neq u}^{U}\beta^i + \sigma_\eta^2}\right).$$

It is easy to verify that for $2P_d = E_u/L$ and $L \to \infty$, the lower-bound $\tilde{\mathcal{R}}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{P}} \to \mathcal{R}^{u,\mathrm{mrc}}_{\bar{m},\mathrm{P}}$.

3.5.4.2 ZF Receiver

The ZF receiver SINR at the $\bar{m}$th subcarrier of the uth user is derived as

$$\Upsilon^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}} = \frac{2P_d}{\sigma_\eta^2\left\{\left(\mathbf{G}^H_{\bar{m}}\mathbf{G}_{\bar{m}}\right)^{-1}\right\}_{u,u}}.$$

The above rate is lower bound as $\mathcal{R}^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}} = \mathbb{E}\left[\log_2(1+\Upsilon^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}})\right]$ is

$$\mathcal{R}^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}} \geq \tilde{\mathcal{R}}^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}} = \log_2\left(1 + \frac{2P_d\beta^u(L-U)}{\sigma_\eta^2}\right). \tag{3.66}$$

If $2P_d = E_u/L$ and L grows large, one obtains $\tilde{\mathcal{R}}^{u,\mathrm{zf}}_{\bar{m},\mathrm{P}} \xrightarrow{L\to\infty} \log_2\left(1 + E_u\beta^u/\sigma_\eta^2\right)$. It is observed that similar to the OFDM-based massive MIMO systems, the uplink power scaling laws also hold for massive MIMO-FBMC systems in the existing of perfect channel knowledge at the base station.

Figure 3.5a,b demonstrate the performance of FBMC and OFDM waveforms in the uplink of massive MIMO technology. FBMC waveform with $N = 128$ subcarriers is considered. Pilot and data symbols are drawn from the in-phase and quadrature components of 4-QAM constellation. Each of the subcarriers in FBMC is shaped using the IOTA filter [15] of length $4T$. The matrix $\mathbf{D}$ comprises large-scale fading coefficients β^u (for $1 \leq u \leq U$), which depend on geographical position of users in the cell as well as radio frequency of EM waves. These coefficients remain constant for multiple coherence intervals [60]. The coefficients β^u, for $1 \leq u \leq U$, are typically modelled as $\beta^u = z^u/(r^u/r_h)^\nu$ [60], where z^u is a log-normal random variable for the uth user with a standard deviation σ_z, r^u is the distance between the uth user and the BS and ν is the path loss exponent. The large-scale fading matrix $\mathbf{D} = \mathrm{diag}[0.749,\ 0.045,\ 0.246,\ 0.121,\ 0.125,\ 0.142,\ 0.635,\ 0.256]$ [61] is a snapshot of the above model with $\sigma_z = 8$ dB, $\nu = 3.8$, $r_h = 100$ metres and $r^u = 1000$ metres. The small-scale fading channel from each user to the base station is considered to be complex Gaussian of length $L_h = 6$ with uniform power delay profile. Each user transmits $M = U$ number of OQAM pilot symbols on each subcarrier for channel estimation. The noise variance σ_η^2 is assumed to be 1.

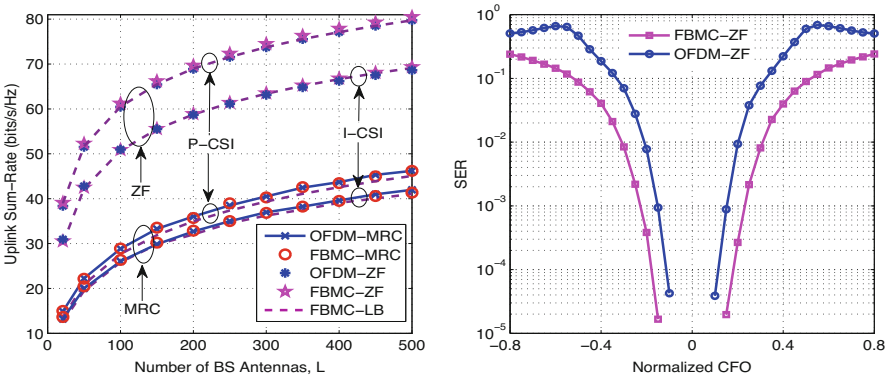


Fig. 3.5 Uplink performance comparison of FBMC- and OFDM-based massive MIMO technologies: (**a**) Uplink sum rate as a function of the number of BS antennas in the existence of imperfect and perfect knowledge of channel with power per user $2P_d = 10$ dB and (**b**) SER versus normalized CFO with perfect knowledge of channel, power per user $2P_d = -5$ dB, $L = 64$, $L_h = 2$ and $\beta^u = 1$ for $1 \leq u \leq U$

With this assumption, the transmit power per user can be interpreted as normalized transmit SNR, and hence it is dimensionless. There are $U = 8$ single antenna users in the cell.

Figure 3.5a shows that the uplink sum rates achieved by the ZF and MRC receivers agree to their respective lower bounds in the existence of imperfect and perfect channel knowledge. It can be seen that that the uplink sum rates of the FBMC-based MU massive MIMO system coincide with their OFDM counterparts. It is also observed that the ZF receiver outperforms the MRC receiver, especially in the high SNR regime. However, in comparison to the latter, the former costs more in terms of computational complexity. The results in Fig. 3.5a are plotted in the existence of perfect synchronization. However, in the existence of synchronization errors due to practical impairments such as CFO, as shown in Fig. 3.5b, the symbol error rate (SER) of an OFDM-based massive MIMO system degrades severely in comparison to its FBMC counterpart. This happens because the OFDM waveform experiences significant ICI due to the sinc-shaped frequency localization of the rectangular pulse shaping, whereas FBMC waveform, due to associated well time-frequency localized pulse-shaping filter, experience significantly lower ICI, which makes FBMC waveform robust against practical impairments.

3.6 Conclusions and Future Directions

This chapter designed and analysed FBMC-OQAM-based MIMO and massive MIMO systems. For improving the accuracy of CSI at the receiver with limited pilot

overhead, a semi-blind (SB) MIMO-FBMC CSI estimation scheme was developed. The SB technique exploited both the pilot symbols and the second-order statistical information of data symbols. The NMSE gain of the SB scheme over the LS increases with the number of receive antennas. It was also demonstrated that the SB scheme achieves its CRLB.

This chapter also analysed the uplink performance of FBMC waveform in multi-user massive MIMO systems in the existence of both perfect and imperfect knowledge of channel. The uplink sum rates and their corresponding lower bounds were derived for massive MIMO-FBMC systems relying on the MRC and ZF receiver processing at the base station with/without perfect CSI. For both the receivers, the derived lower bounds on the achievable uplink sum rate were seen to closely agree with their respective simulated rates. Furthermore, the uplink power scaling laws, similar to OFDM waveform, were seen to exist for FBMC-based massive MIMO systems. It was also shown that OFDM- and FBMC-based massive MIMO systems achieve the same uplink performance in the existence of perfect synchronization. However, in the existence of practical impairments such as CFO, FBMC-based massive MIMO systems were seen to vastly outperform their OFDM counterparts.

The future works may investigate semi-blind scheme for highly frequency-selective and/or time-selective channels. Future research may present an analysis for characterising the performance of FBMC-based massive MIMO systems in time-selective channels. Future research may also analyse performance of FBMC waveform in the downlink of massive MIMO by considering the effect of multi-user precoding, which poses additional challenges. Future lines of this work can also investigate performance of FBMC signalling in other state-of-the-art technologies like millimetre wave, intelligent reflecting surfaces (IRS) and Internet of Things (IoT).

References

1. T. Starr, J.M. Cioffi, P.J. Silverman, *Understanding Digital Subscriber Line Technology* (Prentice Hall PTR, Hoboken, 1999)
2. R.V. Nee, R. Prasad, *OFDM for Wireless Multimedia Communications* (Artech House Inc., Norwood, 2000)
3. S. Schwarz, T. Philosof, M. Rupp, Signal processing challenges in cellular-assisted vehicular communications: efforts and developments within 3GPP LTE and beyond. IEEE Signal Process. Mag. **34**(2), 47–59 (2017)
4. B. Farhang-Boroujeny, R. Kempter, Multicarrier communication techniques for spectrum sensing and communication in cognitive radios. IEEE Commun. Mag. **46**(4), 80–85 (2008)
5. T. Pollet, M.V. Bladel, M. Moeneclaey, BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise. IEEE Trans. Commun. **43**(234), 191–193 (1995)
6. M. Morelli, C.J. Kuo, M. Pun, Synchronization techniques for orthogonal frequency division multiple access (OFDMA): a tutorial review. Proc. IEEE **95**(7), 1394–1427 (2007)
7. V. Vakilian, T. Wild, F. Schaich, S. ten Brink, J. Frigon, Universal-filtered multi-carrier technique for wireless systems beyond LTE, in *Workshops Proceedings of the Global Communications Conference, GLOBECOM*, Atlanta, December 9–13 (2013), pp. 223–228

8. B. Farhang-Boroujeny, H. Moradi, OFDM inspired waveforms for 5G. IEEE Commun. Surv. Tutorials **18**(4), 2474–2492 (2016)
9. N. Michailow, M. Matthe, I.S. Gaspar, A.N. Caldevilla, L.L. Mendes, A. Festag, G.P. Fettweis, Generalized frequency division multiplexing for 5th generation cellular networks. IEEE Trans. Commun. **62**(9), 3045–3061 (2014)
10. B. Farhang-Boroujeny, OFDM versus filter bank multicarrier. IEEE Signal Process. Mag. **28**(3), 92–112 (2011)
11. G. Cherubini, E. Eleftheriou, S. Ölçer, Filtered multitone modulation for very high-speed digital subscriber lines. IEEE J. Select. Areas Commun. **20**(5), 1016–1028 (2002)
12. B. Farhang-Boroujeny, C.H.G. Yuen, Cosine modulated and offset QAM filter bank multicarrier techniques: a continuous-time prospect. EURASIP J. Adv. Sig. Proc. **2010** (2010)
13. M.G. Bellanger, Specification and design of a prototype filter for filter bank based multicarrier transmission, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2001*, 7–11 May, 2001, Salt Palace Convention Center, Salt Lake City (2001), pp. 2417–2420
14. A. Aminjavaheri, A. Farhang, A. RezazadehReyhani, B. Farhang-Boroujeny, Impact of timing and frequency offsets on multicarrier waveform candidates for 5G, in *IEEE Signal Processing and Signal Processing Education Workshop, SP/SPE 2015*, Salt Lake City, August 9–12 (2015), pp. 178–183
15. P. Siohan, C. Siclet, N. Lacaille, Analysis and design of OFDM/OQAM systems based on filterbank theory. IEEE Trans. Signal Proces. **50**(5), 1170–1183 (2002)
16. A.I. Pérez-Neira, M. Caus, R. Zakaria, D.L. Ruyet, E. Kofidis, M. Haardt, X. Mestre, Y. Cheng, MIMO signal processing in offset-QAM based filter bank multicarrier systems. IEEE Trans. Signal Proces. **64**(21), 5733–5762 (2016)
17. M.J. Abdoli, M. Jia, J. Ma, Weighted circularly convolved filtering in OFDM/OQAM, in *24th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2013*, London, September 8–11 (2013), pp. 657–661
18. H. Lin, P. Siohan, Multi-carrier modulation analysis and WCP-COQAM proposal. EURASIP J. Adv. Sig. Proc. **2014**, 79 (2014)
19. H.S. Sourck, Y. Wu, J.W.M. Bergmans, S. Sadri, B. Farhang-Boroujeny, Complexity and performance comparison of filter bank multicarrier and OFDM in uplink of multicarrier multiple access networks. IEEE Trans. Signal Proces. **59**(4), 1907–1912 (2011)
20. R. Nissel, S. Schwarz, M. Rupp, Filter bank multicarrier modulation schemes for future mobile communications. IEEE J. Select. Areas Commun. **35**(8), 1768–1782 (2017)
21. A. Aminjavaheri, A. Farhang, B. Farhang-Boroujeny, Filter bank multicarrier in massive MIMO: analysis and channel equalization. IEEE Trans. Signal Proces. **66**(15), 3987–4000 (2018)
22. F. Rottenberg, X. Mestre, F. Horlin, J. Louveaux, Performance analysis of linear receivers for uplink massive MIMO FBMC-OQAM systems. IEEE Trans. Signal Proces. **66**(3), 830–842 (2018)
23. S. Srivastava, P. Singh, A.K. Jagannatham, A. Karandikar, L. Hanzo, Bayesian learning-based doubly-selective sparse channel estimation for millimeter wave hybrid MIMO-FBMC-OQAM systems. IEEE Trans. Commun. **69**(1), 529–543 (2021)
24. E. Kofidis, D. Katselis, A.A. Rontogiannis, S. Theodoridis, Preamble-based channel estimation in OFDM/OQAM systems: a review. Signal Proces. **93**(7), 2038–2054 (2013)
25. P. Singh, K. Vasudevan, Preamble-based synchronization for OFDM/OQAM systems in AWGN channel, in *4th IEEE International Conference on Signal Processing and Integrated Networks (SPIN)* (2017), pp. 60–65
26. C. Lélé, J. Javaudin, R. Legouable, A. Skrzypczak, P. Siohan, Channel estimation methods for preamble-based OFDM/OQAM modulations. Eur. Trans. Telecommun. **19**(7), 741–750 (2008)
27. P. Singh, R. Budhiraja, K. Vasudevan, SER analysis of MMSE combining for MIMO FBMC-OQAM systems with imperfect CSI. IEEE Commun. Lett. **23**(2), 226–229 (2019)
28. E. Kofidis, D. Katselis, Preamble-based channel estimation in MIMO-OFDM/OQAM systems, in *IEEE International Conference on Signal and Image Processing Applications, ICSIPA*, Kuala Lumpur, 16–18 November (2011), pp. 579–584

29. F. Rottenberg, Y. Medjahdi, E. Kofidis, J. Louveaux, Preamble-based channel estimation in asynchronous FBMC-OQAM distributed MIMO systems, in *International Symposium on Wireless Communication Systems (ISWCS)*, Belgium, 25–28, August (2015), pp. 566–570
30. S. Hu, Z. L. Liu, Y. L. Guan, C. Jin, Y. Huang, J. Wu, Training sequence design for efficient channel estimation in MIMO-FBMC systems. IEEE Access **5**, 4747–4758 (2017)
31. J.-P. Javaudin, Y. Jiang, Channel estimation in MIMO OFDM/OQAM, in *IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, SPAWC* (2008), pp. 266–270
32. P. Singh, K. Vasudevan, MIMO-FBMC channel estimation with limited, and imperfect knowledge of channel correlations, in *National Conference on Communications, NCC*, Bangalore, February 20–23 (2019), pp. 1–6
33. E. Kofidis, Preamble-based estimation of highly frequency selective channels in MIMO-FBMC/OQAM systems, in *Proceedings of the 21th European Wireless Conference* (VDE, Berlin, 2015), pp. 1–6
34. P. Singh, K. Vasudevan, Time domain channel estimation for MIMO-FBMC/OQAM systems. Wirel. Person. Commun. **108**(4), 2159–2178 (2019)
35. P. Singh, E. Sharma, K. Vasudevan, R. Budhiraja, CFO and channel estimation for frequency selective MIMO-FBMC/OQAM systems. IEEE Wirel. Commun. Lett. **7**(5), 844–847 (2018)
36. M. Lin, Y. Li, L. Xiao, J. Wang, A compressive sensing channel estimation for MIMO FBMC/OQAM system. Wirel. Person. Commun., 1–16 (2017)
37. M. Renfors, X. Mestre, E. Kofidis, F. Bader, *Orthogonal Waveforms and Filter Banks for Future Communication Systems* (Academic Press, Cambridge, 2017)
38. H. Bölcskei, P. Duhamel, R. Hleiss, A subspace-based approach to blind channel identification in pulse shaping OFDM/OQAM systems. IEEE Trans. Signal Proces. **49**(7), 1594–1598 (2001)
39. V. Savaux, F. Bader, J. Palicot, OFDM/OQAM blind equalization using CNA approach. IEEE Trans. Signal Proces. **64**(9), 2324–2333 (2016)
40. W. Hou, B. Champagne, Semiblind channel estimation for OFDM/OQAM systems. IEEE Signal Process. Lett. **22**(4), 400–403 (2015)
41. B. Su, Semiblind channel estimation for OFDM/OQAM systems assisted by zero-valued pilots, in *IEEE International Conference on Digital Signal Processing*, Singapore, July 21–24 (2015), pp. 393–397
42. E. Kofidis, C. Chatzichristos, A.L.F. de Almeida, Joint channel estimation/data detection in MIMO-FBMC/OQAM systems – A tensor-based approach, in *25th European Signal Processing Conference, EUSIPCO 2017*, Kos (2017), pp. 420–424
43. J.-M. Choi, Y. Oh, H. Lee, J.-S. Seo, Pilot-aided channel estimation utilizing intrinsic interference for FBMC/OQAM systems. IEEE Trans. Broadcast. **63**(4), 644–655 (2017)
44. H. Lin, P. Siohan, Robust channel estimation for OFDM/OQAM. IEEE Commun. Lett. **13**(10), 724–726 (2009)
45. V. Savaux, F. Bader, Y. Louët, A joint MMSE channel and noise variance estimation for OFDM/OQAM modulation. IEEE Trans. Commun. **63**(11), 4254–4266 (2015)
46. D. Katselis, E. Kofidis, A.A. Rontogiannis, S. Theodoridis, Preamble-based channel estimation for CP-OFDM and OFDM/OQAM systems: a comparative study. IEEE Trans. Signal Proces. **58**(5), 2911–2916 (2010)
47. M.J.F. García, E. Biglieri, G. Taricco, Frequency-domain channel estimation in MIMO-OFDM, in *12th European Signal Processing Conference*, Vienna, September 6–10 (2004), pp. 1869–1872
48. R. Vershynin, How close is the sample covariance matrix to the actual covariance matrix?. J. Theoret. Probab. **25**(3), 655–686 (2012)
49. P. Singh, H.B. Mishra, A.K. Jagannatham, K. Vasudevan, L. Hanzo, Uplink sum-rate and power scaling laws for multi-user massive MIMO-FBMC systems. IEEE Trans. Commun. **68**(1), 161–176 (2020)
50. A.K. Jagannatham, B.D. Rao, Whitening-rotation-based semi-blind MIMO channel estimation. IEEE Trans. Signal Proces. **54**(3), 861–869 (2006)
51. K.B. Petersen, M.S. Pedersen et al., The matrix cookbook. Tech. Univ. Denmark **7**, 15 (2008)

52. P. Singh, H.B. Mishra, A.K. Jagannatham, K. Vasudevan, Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems. IEEE Trans. Signal Proces. **67**(18), 4668–4682 (2019)
53. A.K. Jagannatham, B.D. Rao, Cramér-Rao lower bound for constrained complex parameters. IEEE Signal Proces. Lett. **11**(11), 875–878 (2004)
54. J.C. Roh, B.D. Rao, Efficient feedback methods for MIMO channels based on parameterization. IEEE Trans. Wirel. Commun. **6**(1), 282–292 (2007)
55. T.L. Marzetta, E.G. Larsson, H. Yang, H.Q. Ngo, *Fundamentals of Massive MIMO* (Cambridge University Press, Cambridge, 2016)
56. A. Aminjavaheri, A. Farhang, L.E. Doyle, B. Farhang-Boroujeny, Prototype filter design for FBMC in massive MIMO channels, in *IEEE International Conference on Communications, ICC*, Paris, May 21–25 (2017), pp. 1–6
57. A. Farhang, N. Marchetti, L.E. Doyle, B. Farhang-Boroujeny, Filter bank multicarrier for massive MIMO, in *IEEE 80th Vehicular Technology Conference, VTC Fall*, Vancouver, September 14–17 (2014), pp. 1–7
58. A. Aminjavaheri, A. Farhang, N. Marchetti, L.E. Doyle, B. Farhang-Boroujeny, Frequency spreading equalization in multicarrier massive MIMO, in *Workshop Proceedings of the IEEE International Conference on Communication, ICC*, June 8–12 (2015), pp. 1292–1297
59. S. Singh, P. Singh, S. Sahu, K. Vasudevan, H.B. Mishra, Uplink transmission in MU multicell massive MIMO-FBMC systems over ricean fading, in *94th IEEE Vehicular Technology Conference, VTC Fall 2021*, Norman, September 27–30 (2021), pp. 1–6
60. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)
61. Y. Dai, X. Dong, Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing. IEEE Trans. Wirel. Commun. **15**(9), 5932–5946 (2016)

Chapter 4
Full-Duplex Multi-Hop Communication for Beyond 5G

Ekant Sharma and Prem Singh

4.1 Introduction

With 5G, there exist lots of application, namely, healthcare, self-driving cars, drone operation, machine-to-machine communication, industrial Internet of Things (IoT)/industrial automation and robotics, etc. In many such applications, devices need to cooperate or exchange information with each other, without any requirement of base station. However due to the absence of line-of-sight (LoS) component, the communication suffers. Cooperative communication improves the coverage and increases the data rate, mainly when there is no direct link between transmitter and receiver. Cooperative two-hop communication via relaying is the most commonly used technology to enhance the SE of the network. Relays can operate in either half-duplex or full-duplex mode.

Massive multi-input multi-output technology, when used along with multi-hop communication, further enhances the SE. One question that arises is whether increase in SE also leads to a similar increase in energy consumption. Energy efficiency (EE), which is defined as ratio between the throughput and total power consumption, is an important metric when we talk about 5G [1–3]. With 5G the aim is to improve the SE by three times and EE by ten times simultaneously.

In this chapter, we, therefore, consider both SE and EE as a performance metric. We consider a system model, wherein multiple users want to exchange their information with each other in the absence of the direct link between them.

E. Sharma (✉)
Indian Institute of Technology, Roorkee, India
e-mail: ekant@ece.iitr.ac.in

P. Singh
International institute of Information Technology, Bangalore, India
e-mail: prem.singh@iiitb.ac.in

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_4

4.2 Organization of the Chapter

First we will cover pre-requisite topics which is important to thoroughly understand this chapter content. This will include massive MIMO; relay protocols, namely, amplify-and-forward (AF) and decode-and-forward (DF), among others; one-way and two-way communication; and difference between half-duplex and full-duplex. This chapter will also have a dedicated section to discuss about energy efficiency, which is an important metric for 5G and beyond technologies. Most conventional works optimize the SE of wireless systems, e.g., by using multiple antennas and cooperative relaying. This improvement, however, comes at the cost of higher energy that is consumed by the additional antennas/nodes. The concept of energy-efficient design, commonly defined as information bits per unit transmit energy, has recently attracted attention.

We will next discuss in detail the single-pair and multi-pair half-duplex massive MIMO relaying communication. We will derive SINR expressions for two different precoders maximal-ratio-combiner (MRC)/maximal-ratio-transmission (MRT) and zero-forcing-receiver (ZFR)/ zero-forcing-transmitter (ZFT). Multi-user full-duplex massive MIMO designs are much more challenging due to multiple interferences, namely, due to (1) full-duplex users and relay resulting in self-loop and inter-user interferences, and (2) inter-pair interference, since multi-pair of users are communicating with each other using the same channel, among others. The next section will begin with the literature survey providing the brief overview of the state-of-the-art research on full-duplex multi-hop communication. It will subsequently discuss about the existing self-interference cancellation (active/passive) techniques, which help in bringing down the interference levels to the order of noise power and therefore making the full-duplex study feasible. This allows the device to transmit or receive on the same time/frequency resource, therefore resulting in higher data rate than half-duplex communication. Half-duplex-capable device, on the contrary, can only transmit or receive data at a given time/frequency resource. The detailed mathematical expressions will be provided.

The chapter concludes with a summary of the topics covered and a roadmap for future directions. It will highlight the issues and challenges in massive MIMO full-duplex relay system, which are yet to be investigated. This will open a plethora of research problems which researchers can pursue in future.

4.3 Massive MIMO

5G aims to achieve an order of magnitude increase in throughput of the network. The throughput of a network can improve by (1) using higher spectrum resources resulting in higher bandwidth; (2) increasing the number of cells/km^2; (3) increasing the transmit power; and (4) using large number of antennas at the base station. 5G uses massive MIMO technology, where the base station comprises 64–256 antennas,

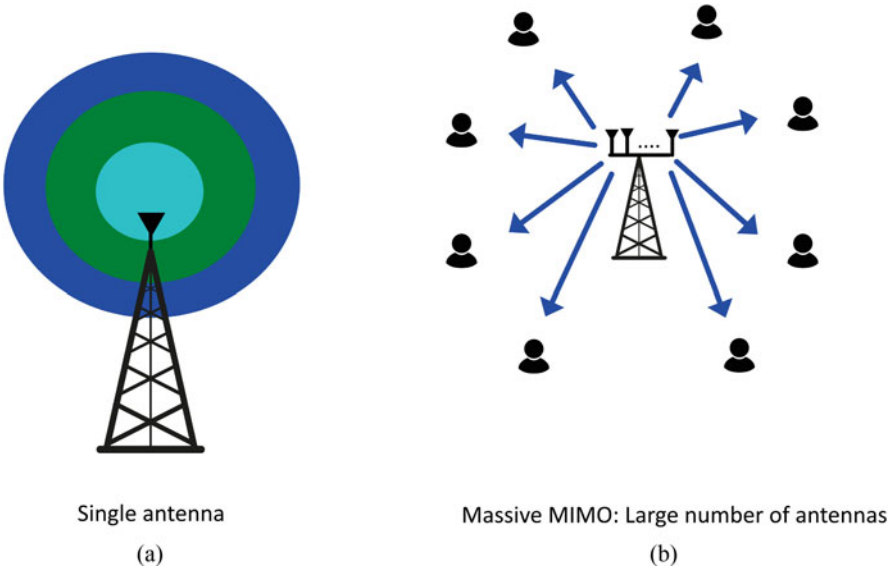


Fig. 4.1 (**a**) Single antenna and (**b**) massive MIMO

which results in large array gain due to increase in the number of antennas, higher spatial multiplexing gain, and higher spatial diversity. The massive MIMO, as shown in Fig. 4.1, by using simple linear transmit processing schemes (e.g., ZFR/ZFT and MRC/MRT) can cancel co-channel interference, which significantly improves their SE with reduced complexity [4–9].

4.4 Multi-hop Communication

Multi-hop communication (commonly called as relay-based communication) has been extensively investigated to increase data rate, improve diversity, and expand coverage [10–14]. Consider a situation, where the user transmits a signal toward an intended receiver. In the absence of LoS, the received signal may experience higher probability of error. In such scenario, a third device known as relay is placed generally somewhere in between the transmitter and receiver. The relay receives the signal from the transmitter and retransmits toward the intended receiver. Figure 4.2 shows some practical relay deployments, resulting in increased coverage and capacity. Most commonly, relay either uses decode-and-forward or amplify-and-forward protocol. In decode-and-forward, the signal received at the relay is decoded first and then retransmitted to the receiver. On the contrary, with amplify-and-forward, the received signal is amplified and then forwarded to the receiver. Decode-and-forward relay has higher complexity due to the additional hardware

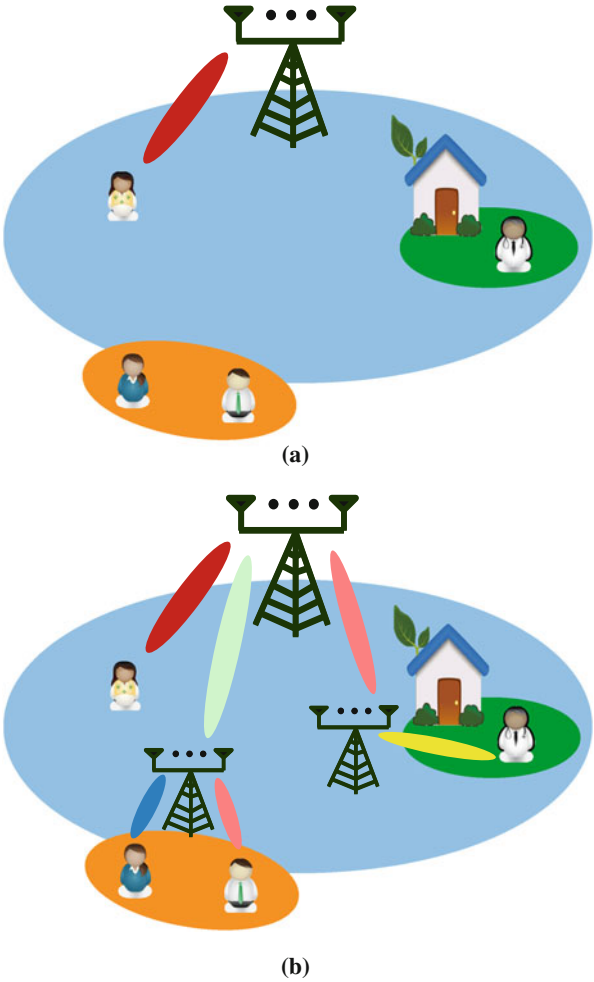


Fig. 4.2 Installation of relays to extend coverage and capacity

required for decoding, whereas amplify-and-forward relay suffers from amplified noise.

4.4.1 Full-Duplex Multi-Hop Communication

Traditional wireless technologies transmit information by employing either frequency-division or time-division approach. This results in half-duplex communication. Full-duplex communication, on the other hand, as shown in Fig.

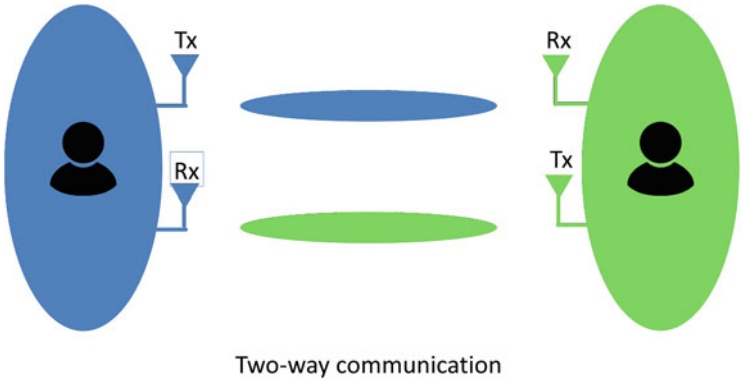


Fig. 4.3 Two-way communication using full-duplex-capable devices

4.3, transmits and receives on the same time/frequency resource, doubling the SE of half-duplex system [15–21]. The full-duplex system, however, suffers from strong self-loop interference (SLI) due to its own transmission. The self-interference cancellation techniques, such as time-domain cancellation, passive isolation, and spatial suppression, help in suppressing the self-interference to a level that FD system results in higher SE than half-duplex system.

4.5 Single-Pair/Multi-Pair One-Way/Two-Way Relay Models

4.5.1 Single User Pair Relay Model

The relay transmission can happen either in one-way or two-way. With one-way, four resources are required for exchanging information between two users. On the contrary, with two-way, we only require two resources, thereby resulting in higher SE [13]. Figure 4.4a and b shows the half-duplex one-way and half-duplex two-way relay transmission, respectively.

A full-duplex relay, as mentioned earlier, simultaneously transmits and receives on the same time/frequency resource. The full-duplex one-way relay, therefore, can double the SE as compared to half-duplex one-way relay, as with full-duplex relay we can transmit using one resource [15]. Two-way full-duplex relaying, on the other hand, allows a user pair to exchange their information with one another via a relay using only one resource [22], resulting in further increase in the system SE. Figure 4.5a and b depicts the full-duplex one-way and full-duplex two-way transmission, respectively.

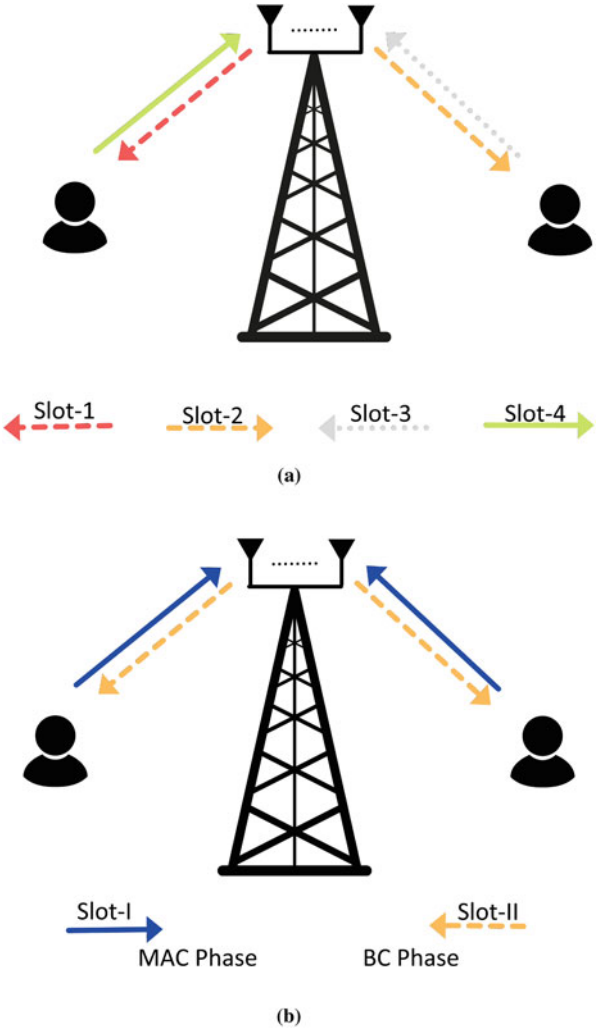


Fig. 4.4 Half-duplex relay. (**a**) One-way relay. (**b**) Two-way relay

4.5.2 Multi-Pair Two-Way Relay Model

The authors in [23–25] extended the single-pair relay model to multi-pair relay model, where multiple user pairs exchange information via relay. Consider a two-way multi-pair relay model as shown in Fig. 4.6, with full-duplex relay and full-duplex users. This architecture will be discussed in more detail in Sect. 4.8.

Due to the full-duplex relay and user, the system will have following interferences: (1) self-interference at the users and at the relay and (2) users on the same

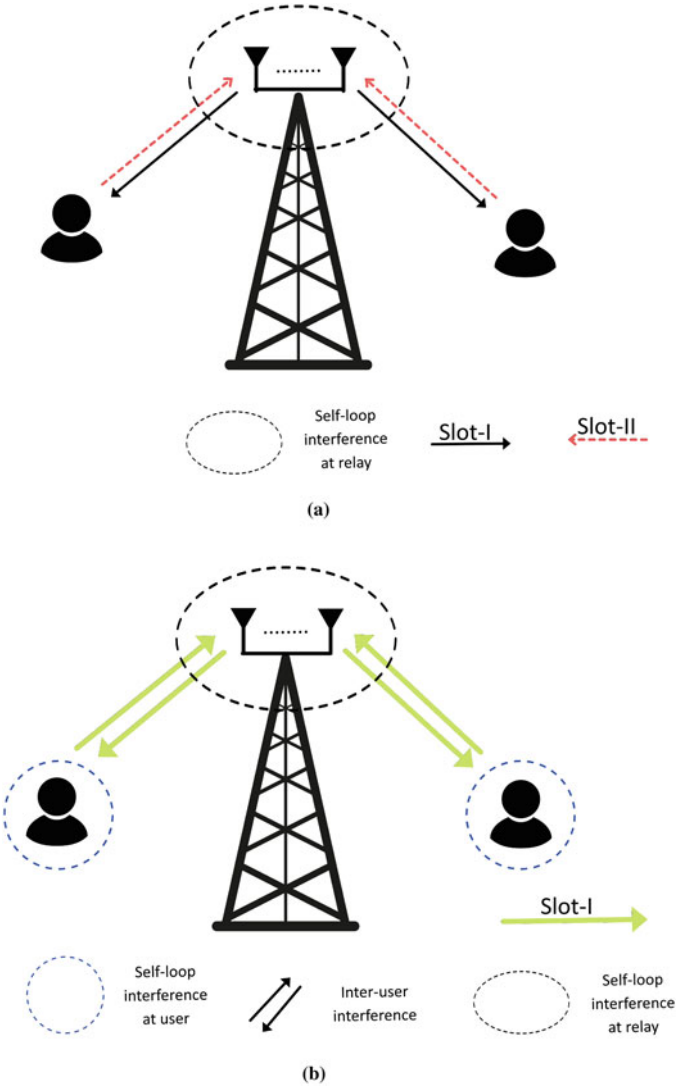


Fig. 4.5 Full-duplex relay. (**a**) One-way relay. (**b**) Two-way relay

side of the relay will interfere with each other, resulting in the inter-user interference (IUI). Also due to multiple pairs of users accessing the same channel, there will be co-channel (inter-pair) interference [24].

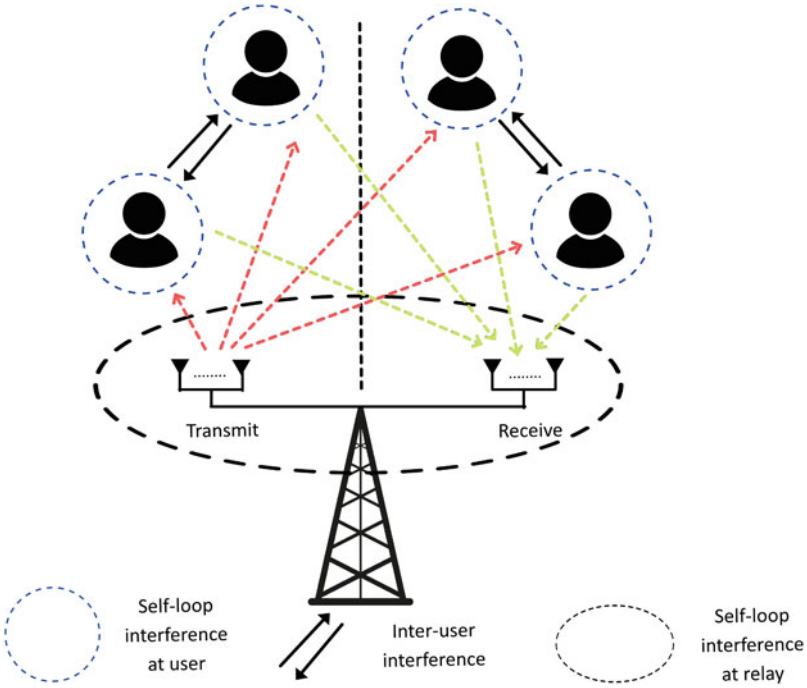


Fig. 4.6 Multi-pair full-duplex two-way relay system

4.6 Single-Pair Half-Duplex Two-Way Massive MIMO Relay: Mathematical Model

Consider a single-pair (i.e., two-user) two-way AF half-duplex relaying system [26–30]. Let N denote the number of antennas at relay, and each user is equipped with a single antenna. The direct link between the two users is absent due to large path loss and heavy shadowing. The channels between the relay and the users are reciprocal; this implies that the users and relay operate in time-division duplex (TDD) mode. We assume that all the users have the same transmit power P_s and relay power is P_R.

The communication takes place in two phases: (1) multiple access (MAC) phase and (2) broadcast (BC) phase. During the MAC phase, the user-1 and user-2 transmit signal $\sqrt{P_s}s_1$ and $\sqrt{P_s}s_2$, respectively, to the relay. The signal received at the relay is expressed as

$$\mathbf{y}_R = \sqrt{P_s}(\mathbf{g_1}s_1 + \mathbf{g_2}s_2) + \mathbf{z}_R = \sqrt{P_s}\mathbf{G}\mathbf{s} + \mathbf{z}_R. \tag{4.1}$$

We define $\mathbf{s} = [s_1,\ s_2]^T$, and $\mathbf{G} = [\mathbf{g}_1,\ \mathbf{g}_2]$, where $\mathbf{g}_k \in \mathbb{C}^{N\times 1}$ denotes the channel between the kth user and the relay. The channel matrix $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$, where $\mathbf{D}$ is the diagonal matrix representing the large-scale fading coefficients as follows:

$\mathbf{D} = \text{diag}\{\sigma_{g,k}^2\}_{k=1}^{2K}$. The matrix $\mathbf{H} \sim \mathcal{CN}(0, \mathbf{I}_N)$ represents the small-scale fading. Remember the elements of the matrix $\mathbf{H} \sim \mathcal{CN}(0, \mathbf{I}_N)$ are all independent and identically distributed (i.i.d.) random variables. The vector $\mathbf{z}_R \sim \mathcal{CN}(0, \sigma_{nr}^2 \mathbf{I}_N)$ represents the additive white Gaussian noise (AWGN) at the relay.

The relay amplifies and beamforms the received signal as follows:

$$\mathbf{x}_R = \alpha \mathbf{W} \mathbf{y}_R. \tag{4.2}$$

Here $\mathbf{W}$ denotes the beamforming matrix (discussed in Sect. 4.6.1), and the scalar α denotes the amplification factor α which is chosen to satisfy the total power constraint at the relay. Using (4.23) and (4.2), the amplification factor α can be written as

$$\alpha = \sqrt{\frac{P_R}{\mathbb{E}\left[\|\mathbf{WGs}\|^2\right] + \mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right]}}. \tag{4.3}$$

In the broadcast phase, the relay transmits the amplified and beamformed signal $\mathbf{x}_R$ to both the users. The signal received at the user-1 and user-2, respectively, is as follows:

$$y_1 = \alpha \mathbf{g}_1^T \mathbf{W} \mathbf{y}_R + z_1 \text{ and} \tag{4.4}$$

$$y_2 = \alpha \mathbf{g}_2^T \mathbf{W} \mathbf{y}_R + z_2. \tag{4.5}$$

Here $\mathbf{g}_k^T \in \mathbb{C}^{1 \times N}$ for $k = 1, \ldots, 2K$ denotes the channel between the relay and the kth user. The scalar $z_k \sim \mathcal{CN}(0, \sigma_n^2)$ represents the AWGN at the kth user. Using (4.23) and (4.2), we can re-express (4.4) and (4.5) as

$$\begin{aligned} y_1 &= \alpha \sqrt{P_s} \mathbf{g}_1^T \mathbf{WGs} + \alpha \mathbf{g}_1^T \mathbf{W} \mathbf{z}_R + z_1 \\ &= \alpha \sqrt{P_s} \mathbf{g}_1^T \mathbf{W} \mathbf{g}_2 s_2 + \alpha \sqrt{P_s} \mathbf{g}_1^T \mathbf{W} \mathbf{g}_1 s_1 + \alpha \mathbf{g}_1^T \mathbf{W} \mathbf{z}_R + z_1 \text{ and} \end{aligned} \tag{4.6}$$

$$\begin{aligned} y_2 &= \alpha \sqrt{P_s} \mathbf{g}_2^T \mathbf{WGs} + \alpha \mathbf{g}_2^T \mathbf{W} \mathbf{z}_R + z_2 \\ &= \alpha \sqrt{P_s} \mathbf{g}_2^T \mathbf{W} \mathbf{g}_1 s_1 + \alpha \sqrt{P_s} \mathbf{g}_2^T \mathbf{W} \mathbf{g}_2 s_2 + \alpha \mathbf{g}_2^T \mathbf{W} \mathbf{z}_R + z_2. \end{aligned} \tag{4.7}$$

Recall that user-1 and user-2 want to exchange their information with each other. In (4.6) and (4.7), the first terms represent the desired signal, the second term represents the self-interference (SI) term, and the third term denotes the amplified noise from the relay.

To further simplify (4.6) and (4.7), we will need the knowledge of beamforming matrix $\mathbf{W}$. We next design beamforming matrix $\mathbf{W}$ based on maximal ratio and zero-forcing principles, which is commonly used in massive MIMO literature.

4.6.1 Relay Beamforming Design

In this chapter, we will consider MRC/MRT and ZFR/ZFT beamformers. Both have their own advantages and disadvantages: the MRC/MRT is simple to implement and has low complexity. In addition it performs better than ZFR/ZFT at low SNR. On the contrary, ZFR/ZFT performs much better than MRC/MRT at high SNR but has higher complexity. There also exist some other linear designs like minimum mean square error (MMSE) beamforming design [31] among others.

4.6.1.1 MRC/MRT Beamformer

The MRC/MRT precoder is formulated as

$$\mathbf{W} = \mathbf{G}^{*}\mathbf{T}\mathbf{G}^{H}, \tag{4.8}$$

where the permutation matrix

$$\mathbf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{4.9}$$

ensures that the users receive their intended data. The MRC/MRT precoder had complexity $O(N^2K)$, where K denotes the number of user pairs. Remember that MRC/MRT cannot completely suppress the self-interference and therefore perform inferior than ZFR/ZFT at high SNR.

4.6.1.2 ZFR/ZFT Precoder

The ZFR/ZFT precoder is formulated as

$$\mathbf{W} = \bar{\mathbf{G}}^{*}\mathbf{T}\bar{\mathbf{G}}^{H}, \tag{4.10}$$

where $\bar{\mathbf{G}} = \mathbf{G}\left(\mathbf{G}^{H}\mathbf{G}\right)^{-1}$. The computational complexity is $O(N^2K + NK^2 + K^3)$, which is higher as compared to the MRC/MRT beamforming design.

4.6.2 Performance Analysis of Single-Pair Relay Model: MRT/MRC Beamforming

In this subsection, we calculate the SE and EE of the single-pair AF two-way half-duplex relay model considering MRC/MRT beamforming matrix. Using (4.8), the

received signal in (4.6) and (4.7) can be further simplified as

$$y_1 = \alpha\sqrt{P_S}\mathbf{g}_1^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_2 s_2 + \alpha\sqrt{P_S}\mathbf{g}_1^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_1 s_1 + \alpha\mathbf{g}_1^T\mathbf{W}\mathbf{z}_R + z_1 \text{ and} \tag{4.11}$$

$$y_2 = \alpha\sqrt{P_S}\mathbf{g}_2^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_1 s_1 + \alpha\sqrt{P_S}\mathbf{g}_2^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_2 s_2 + \alpha\mathbf{g}_2^T\mathbf{W}\mathbf{z}_R + z_2. \tag{4.12}$$

The received instantaneous signal-to-noise ratio (SNR) is given as

$$\text{SINR}_1 = \frac{\alpha^2 P_S|\mathbf{g}_1^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_2|^2}{\alpha^2 P_S|\mathbf{g}_1^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_1 s_1|^2 + \alpha^2\|\mathbf{g}_1^T\mathbf{W}\|^2\sigma_{nr}^2 + \sigma_n^2}, \tag{4.13}$$

$$\text{SINR}_2 = \frac{\alpha^2 P_S|\mathbf{g}_2^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_1|^2}{\alpha^2 P_S|\mathbf{g}_2^T\mathbf{G}^*\mathbf{T}\mathbf{G}^H\mathbf{g}_2 s_2|^2 + \alpha^2\|\mathbf{g}_2^T\mathbf{W}\|^2\sigma_{nr}^2 + \sigma_n^2}. \tag{4.14}$$

The sum SE of the system, in bps/Hz, is therefore

$$\text{SE} = \frac{1}{2}\sum_{k=1}^{2}\mathbb{E}[\log_2(1 + \text{SINR}_k)]. \tag{4.15}$$

We also define the EE of the system, in bits/Joule, as the ratio of the sum SE to the total power consumed by it

$$\text{EE} = \frac{B \times \text{SE}}{\mu_R P_R + 2K\mu_s P_S + P_{c,k}}. \tag{4.16}$$

where B denotes the bandwidth and μ_R and μ_s denote the inverse transmit power amplifier efficiency of the relay and users, respectively. The scalar $P_{c,k}$ denotes the circuit power consumption.

4.6.3 *Performance Analysis of Single-Pair Relay Model: ZFR/ZFT Beamforming*

In this subsection, we calculate the SE and EE of the single-pair AF two-way half-duplex relay model considering ZFR/ZFT beamforming matrix. Using (4.35), we can write

$$\mathbf{G}^T\mathbf{W}\mathbf{G} = \mathbf{G}^T\mathbf{G}^*\left(\mathbf{G}^T\mathbf{G}^*\right)^{-1}\mathbf{T}\left(\mathbf{G}^H\mathbf{G}\right)^{-1}\mathbf{G}^H\mathbf{G} = \mathbf{T}, \tag{4.17}$$

$$\Rightarrow \begin{bmatrix}\mathbf{g}_1^T \\ \mathbf{g}_2^T\end{bmatrix}\mathbf{W}\begin{bmatrix}\mathbf{g}_1 & \mathbf{g}_2\end{bmatrix} = \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix}, \tag{4.18}$$

which implies that $\mathbf{g}_1^T\mathbf{W}\mathbf{g}_2 = \mathbf{g}_2^T\mathbf{W}\mathbf{g}_1 = 1$ and $\mathbf{g}_1^T\mathbf{W}\mathbf{g}_1 = \mathbf{g}_2^T\mathbf{W}\mathbf{g}_2 = 0$. Using (4.17), the received signal in (4.6) and (4.7) can be further simplified as

$$y_1 = \alpha\sqrt{P_S}s_2 + \alpha\mathbf{g}_1^T\mathbf{W}\mathbf{z}_R + z_1 \text{ and} \tag{4.19}$$

$$y_2 = \alpha\sqrt{P_S}s_1 + \alpha\mathbf{g}_2^T\mathbf{W}\mathbf{z}_R + z_2. \tag{4.20}$$

The received instantaneous signal-to-interference-noise ratio (SINR) at the kth user (for $k = 1, 2$) is given as

$$\text{SINR}_k = \frac{\alpha^2 P_S}{\alpha^2\|\mathbf{g}_k^T\mathbf{W}\|^2\sigma_{nr}^2 + \sigma_n^2}. \tag{4.21}$$

The sum SE of the system, in bps/Hz, is therefore

$$\begin{aligned}\text{SE} &= \frac{1}{2}\sum_{k=1}^{2K}\mathbb{E}[\log_2(1+\text{SINR}_k)] \\ &= \frac{1}{2}\sum_{k=1}^{2K}\mathbb{E}\left[\log_2\left(1+\frac{\alpha^2 P_S}{\alpha^2\|\mathbf{g}_k^T\mathbf{W}\|^2\sigma_{nr}^2+\sigma_n^2}\right)\right].\end{aligned} \tag{4.22}$$

The EE of the system can be defined similar to (4.16) by substituting the value of SE of ZFR/ZFT.

4.7 Multi-Pair Half-Duplex Two-Way Relay: Mathematical Model

We now generalize the single-pair model results to its multi-pair counterpart, wherein now multiple pairs of users want to exchange information with each other via a shared relay on the same time-frequency resource [26–30]. The notations and assumptions remain the same. We now denote the user pair exchanging information as $(S_k, S_{k'})$ for $k, k' = 1, \ldots, K$. We have in total K number of user pairs and therefore $2K$ number of users.

During MAC phase, all the $2K$ users transmit their signal to the relay. The signal received at the relay is expressed as

$$\mathbf{y}_R = \sqrt{P_s}\mathbf{G}\mathbf{s} + \mathbf{z}_R, \tag{4.23}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_{2K}]^T$. We define $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_1, \ldots, \mathbf{g}_{2K}]$, where $\mathbf{g}_k \in \mathbb{C}^{N\times 1}$ denotes the channel between the kth user and the relay. The relay then amplifies and beamforms the received signal and then broadcasts it to all the users.

The signal received at the kth user in the multi-pair scenario is as follows

$$y_k = \alpha \mathbf{g}_k^T \mathbf{W} \mathbf{y}_R + z_k \tag{4.24}$$

$$= \alpha \sqrt{P_s} \mathbf{g}_k^T \mathbf{W} \mathbf{G} \mathbf{s} + \alpha \mathbf{g}_k^T \mathbf{W} \mathbf{z}_R + z_k$$

$$= \alpha \sqrt{P_s} \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} s_{k'} + \alpha \sqrt{P_s} \mathbf{g}_k^T \mathbf{W} \mathbf{g}_k s_k + \alpha \sqrt{P_s} \mathbf{g}_k^T \mathbf{W} \sum_{i \neq k,k'}^{2K} \mathbf{g}_i s_i + \alpha \mathbf{g}_k^T \mathbf{W} \mathbf{z}_R + z_k. \tag{4.25}$$

Here (k, k') represents a user pair which exchange information. The first and second terms are the desired and the SI term. The third term, which was absent in the single-pair model, represents the interference coming from other $K-1$ pairs and is termed as inter-pair interference. The fourth term is the amplified noise from the relay.

4.7.1 MRC/MRT Beamforming

The received instantaneous SINR at the kth user, using (4.41), is given by

$$\text{SINR}_k = \frac{\alpha^2 P_S |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'}|^2}{\alpha^2 P_S |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'}|^2 + \alpha^2 \sum_{i \neq k,k'}^{2K} |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_i|^2 + \alpha^2 \|\mathbf{g}_k^T \mathbf{W}\|^2 \sigma_{nr}^2 + \sigma_n^2}, \tag{4.26}$$

where $\mathbf{W} = \mathbf{G}^* \mathbf{T} \mathbf{G}^H$ with $\mathbf{T} = \text{blkdiag}\{\mathbf{T}_1, \ldots, \mathbf{T}_K\}$. Here permutation matrix

$$\mathbf{T}_m = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{ for } 1 \leq m \leq K. \tag{4.27}$$

4.7.2 ZFR/ZFT Beamforming

Using (4.17), we can write $\mathbf{g}_k^T \mathbf{W} \mathbf{g}_i = \delta_{k',i}$, where $\delta_{k',i} = 1$ when $\overline{k} = i$ and 0 otherwise. The SI as well as interference tends to zero. The received signal in (4.41) is, therefore, simplified to

$$y_k = \alpha \sqrt{P_S} s_{k'} + \alpha \mathbf{g}_k^T \mathbf{W} \mathbf{z}_R + z_k. \tag{4.28}$$

The received SINR at the kth user is given as

$$\text{SINR}_k = \frac{\alpha^2 P_S}{\alpha^2 \|\mathbf{g}_k^T \mathbf{W}\|^2 \sigma_{nr}^2 + \sigma_n^2}. \tag{4.29}$$

This is same as that obtained from its single-pair counterpart. The SE and EE, therefore, can be obtained, as explained earlier.

4.8 Full-Duplex Multi-Pair AF Massive MIMO Relay

We now extend our understanding from previous sections to full-duplex multi-pair two-way massive MIMO relay [2, 24, 32]. There are few points to note

- multiple full-duplex user pairs with single transmit and receive antenna exchange their data via a shared relay consisting of large number of antennas.
- system suffers from SLI and IUI, from the transmit antenna to receive antenna of the relay and user, respectively.
- system suffers from the inter-pair interference because of multiple users simultaneously transmitting and receiving in the time/frequency resource.

Before discussing the system model in detail, we briefly digress to discuss about the SLI suppression techniques.

4.8.1 SLI Suppression

Full-duplex systems, due to its ability to double the SE as compared to half-duplex systems, have recently received significant attention [27, 33–36]. The main hurdle is the SLI, which arises due to transmission and reception on the same time-frequency resource. Recent full-duplex studies, however, have shown that the SLI can be significantly suppressed [2, 3]. There exist many SLI cancellation techniques in full-duplex literature; we will next discuss some of them.

4.8.1.1 Passive Cancellation

It was shown in [37, 38] that the passive SLI suppression techniques (cross-polarization, absorptive shielding, and directional isolation) result in significantly reduction in the SLI. In addition if the statistical channel information is known, the authors in [38] showed even an increased level of SI suppression using the passive cancellation techniques.

4.8.1.2 Large Antenna Array at Relay

If we consider a large number of antennas N, the channels between the transmit and receive antenna, i.e., the SLI channel and the desired signal channel, become nearly orthogonal [23]. Therefore, the low complexity linear beamforming techniques, as discussed in earlier sections (MRC/MRT or ZFR/ZFT), project the desired signal to the orthogonal complement space of the SLI, resulting in significant reduction of the SLI.

4.8.1.3 Lower Transmit Power at Relay

Reference [23, 25] showed that the SLI is proportional to the relay power, P_R. The relay transmit power can be reduced while maintaining a desired quality-of-service (QoS). This is because, with large N, the relay can transmit focusing its energy in the direction of the receiver, thereby limiting any signal in the other directions where receivers are not located. In this manner, the SLI can be reduced to a significant level. Therefore, by using large N together with low transmit power, it results in lower SLI, since the SLI is proportional to the transmit power [23].

4.8.1.4 Active Cancellation

There also exist active cancellation (time-domain and spatial suppression) techniques [39]; however, it requires the knowledge of SLI. With massive MIMO, the SLI channel is of large dimension, and therefore it becomes quite complex to estimate the channel.

4.9 Multi-Pair Full-Duplex Relay: Mathematical Model

Consider a system model, as shown in Fig. 4.7, where multiple pair of full-duplex users wants to exchange information with each other via a full-duplex AF relay [2, 24, 40, 41].

- The relay consists of N number of transmit antennas and N number of receive antennas.
- Each user consists of one transmit and one receive antenna.
- The user on the one side of the relay communicates with user on the other side of the relay.
- There is no direct path between the users on the two sides of the relay.
- Relay and the users transmit and receive in the same time-frequency resource resulting in SI.
- IUI, due to the full-duplex users on same side of the relay.

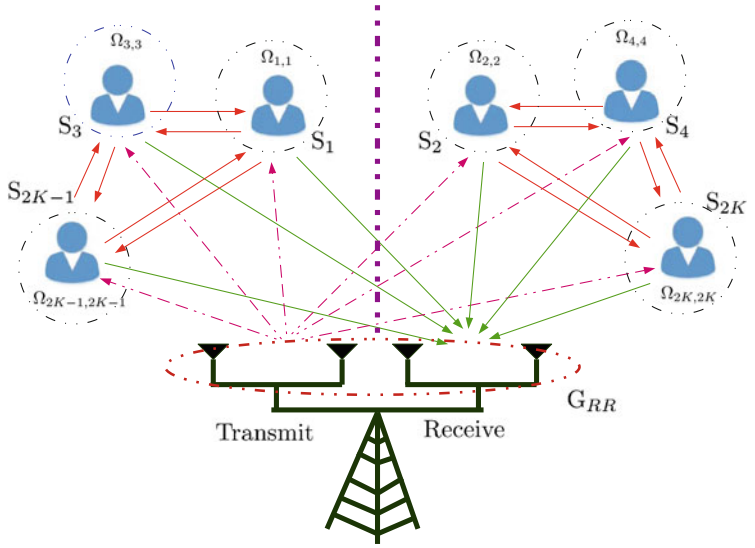


Fig. 4.7 Full-duplex multi-pair relay system [2, 24, 32]

4.9.1 Channel Model

The channel between the transmit antenna of the kth user and the relay receive antenna array is denoted by the vector $\mathbf{g}_k \in \mathbb{C}^{N\times 1}$. Similarly the channels between the relay transmit antenna array and the receive antenna of the kth user are denoted by the vector $\mathbf{f}_k^T \in \mathbb{C}^{1\times N}$. We define $\mathbf{G} = \mathbf{H}_u\mathbf{D}_u^{1/2} = [\mathbf{g}_1, \ldots, \mathbf{g}_{2K}]$ and $\mathbf{F} = \mathbf{H}_d\mathbf{D}_d^{1/2} = [\mathbf{f}_1, \ldots, \mathbf{f}_{2K}]$. Here the matrices $\mathbf{H}_u$ and $\mathbf{H}_d$ account for small-scale fading, with i.i.d. $\mathcal{CN}(0, 1)$ elements. The diagonal matrices $\mathbf{D}_u$ and $\mathbf{D}_d$ account for large-scale fading, with the kth diagonal elements given by $\sigma_{g,k}^2$ and $\sigma_{f,k}^2$, respectively. The SI channel between the transmit and receive antenna of the relay and kth user is denoted as $\mathbf{G}_{RR} \sim \mathcal{CN}(0, \sigma_{LIR}^2\mathbf{I})_N$ and $\Omega_{k,k} \sim \mathcal{CN}(0, \sigma_{k,k}^2)$, respectively. The $\Omega_{k,i} \sim \mathcal{CN}(0, \sigma_{k,i}^2)$ $(k, i \in U_k, i \neq k)$ denotes the IUI channel between the transmit and receive antennas of the users on the same side of the relay. Here for odd k, $U_k = [1, 3, 5, \ldots, 2K-1]$ and for even k, $U_k = [2, 4, 6, \ldots, 2K]$.

4.9.2 Data Transmission

To understand the full-duplex transmission, we consider two time instants: (1) data transmission at time instant $n = 1$—where the relay does not have any signal to transmit and, therefore, only receives from the users and (2) data transmission at time instant $n > 1$—where both relay and users transmit and receive. We now provide details as follows.

- *Time instant* $n = 1$: The received signal at the user S_k and the relay can be written, respectively, as

$$y_k(1) = \sum_{i,k\in U_k} \Omega_{k,i}\sqrt{p_i}x_i(1) + z_k(1) \text{ and}$$

$$\mathbf{y}_R(1) = \sum_{k=1}^{2K} \sqrt{p_k}\mathbf{g}_k x_k(1) + \mathbf{z}_R(1) = \widetilde{\mathbf{G}}\mathbf{x}(1) + \mathbf{z}_R(1). \tag{4.30}$$

Here $\mathbf{x}(n) = [x_1(n), \ldots, x_{2K}(n)]^T \in \mathbb{C}^{2K\times 1}$ with $\mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^H(n)\right] = \mathbf{I}_{2K}$. The scalar $z_k(n) \sim \mathcal{CN}(0, \sigma_n^2)$ and the vector $\mathbf{z}_R(n) \sim \mathcal{CN}(0, \sigma_{nr}^2\mathbf{I}_N)$ denote the AWGN at the user S_k and relay, respectively.

- *Time instant* $n > 1$: The users transmit their respective transmit signal to the relay, and at the same time the relay also broadcasts a vector $\mathbf{x}_R(n) \in \mathbb{C}^{N\times 1}$ to all the users. The signals received at the user S_k and the relay are given, respectively, as

$$y_k(n) = \mathbf{f}_k^T\mathbf{x}_R(n) + \sum_{i,k\in U_k} \Omega_{k,i}\sqrt{p_i}x_i(n) + z_k(n) \text{ and} \tag{4.31}$$

$$\mathbf{y}_R(n) = \widetilde{\mathbf{G}}\mathbf{x}(n) + \mathbf{G}_{RR}\mathbf{x}_R(n) + \mathbf{z}_R(n). \tag{4.32}$$

The relay transmits signal $\mathbf{x}_R(n)$

$$\mathbf{x}_R(n) = \alpha\mathbf{W}\mathbf{y}_R(n-1). \tag{4.33}$$

where the scalar α denotes the amplification factor which is designed to satisfy the transmit power constraint of the relay. The beamforming matrix $\mathbf{W}$ is designed based on either MRC/MRT or ZFR/ZFT as follows:

$$\mathbf{W}_{\text{mr}} = \mathbf{F}^*\mathbf{T}\mathbf{G}^H, \tag{4.34}$$

$$\mathbf{W}_{\text{zf}} = \bar{\mathbf{F}}^*\mathbf{T}\bar{\mathbf{G}}^H. \tag{4.35}$$

Here $\bar{\mathbf{F}} = \mathbf{F}\left(\mathbf{F}^H\mathbf{F}\right)^{-1}$ and $\bar{\mathbf{G}} = \mathbf{G}\left(\mathbf{G}^H\mathbf{G}\right)^{-1}$.

We can observe by iteratively substituting (4.32) into (4.33) that the relay transmit signal is an iterative function of user transmit signal.

$$\mathbf{x}_R(n) = s\left[\mathbf{x}(n-1) + \mathbf{x}(n-2) + \cdots + \mathbf{z}_R(n-1) + \mathbf{z}_R(n-2) + \cdots\right], \tag{4.36}$$

where $s[\cdot]$ is some function. The recent full-duplex studies, e.g., in [18, 23, 24], demonstrate that the SI can be significantly suppressed using the SI suppression techniques mentioned in Sect. 4.8.1. After employing SI suppression techniques,

the strong SI gets significantly suppressed, and the residual SI can be regarded as additional noise [24]. The Gaussian assumption holds due to the various sources of imperfections in the cancellation process (i.e., using central limit theorem). Also this can be considered to be worst-case assumption [25], [42, Prop. 1]. The term $\mathbf{G}_{RR}\mathbf{x}_R$ in (4.32) which represents the SLI at the relay can be rewritten as $\widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R \sim \mathcal{CN}(\mathbf{0}, P_R\sigma^2_{LIR}\mathbf{I}_N)$ be the corresponding residual SLI. We, therefore, have

$$\tilde{\mathbf{y}}_R(n) = \widetilde{\mathbf{G}}\mathbf{x}(n) + \widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R(n) + \mathbf{z}_R(n), \tag{4.37}$$

$$\mathbf{x}_R(n) = \alpha\mathbf{W}\tilde{\mathbf{y}}_R(n-1). \tag{4.38}$$

We observe from (4.38) that the vector $\mathbf{x}_R(n)$ is only a function of $\mathbf{x}(n-1)$, and therefore we drop the time labels as rewritten (4.38) as follows:

$$\mathbf{x}_R = \alpha\mathbf{W}\tilde{\mathbf{y}}_R = \alpha\mathbf{W}\widetilde{\mathbf{G}}\mathbf{x} + \alpha\mathbf{W}\widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R + \alpha\mathbf{W}\mathbf{z}_R. \tag{4.39}$$

Using (4.39) and $P_R = \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{x}_R\mathbf{x}_R^H\right]\right\}$, the amplification factor α can be written as

$$\alpha = \sqrt{\frac{P_R}{\mathbb{E}\left[\|\mathbf{W}\widetilde{\mathbf{G}}\mathbf{x}\|^2\right] + \mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right] + \mathbb{E}\left[\|\mathbf{W}\widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R\|^2\right]}}. \tag{4.40}$$

The received signal at the user S_k in (4.31), using (4.39), is given by

$$\begin{aligned}
y_k &= \mathbf{f}_k^T\mathbf{x}_R + \sum_{i,k\in U_k}\Omega_{k,i}\sqrt{p_k}x_i + z_k \\
&= \mathbf{f}_k^T\alpha\mathbf{W}\widetilde{\mathbf{G}}\mathbf{x} + \mathbf{f}_k^T\alpha\mathbf{W}\widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R + \mathbf{f}_k^T\alpha\mathbf{W}\mathbf{z}_R + \sum_{i,k\in U_k}\Omega_{k,i}\sqrt{p_i}x_i + z_k \\
&= \underbrace{\mathbf{f}_k^T\alpha\mathbf{W}\sqrt{p_{k'}}\mathbf{g}_{k'}x_{k'}}_{\text{desired signal}} + \underbrace{\mathbf{f}_k^T\alpha\mathbf{W}\sqrt{p_k}\mathbf{g}_k x_k}_{\text{self-interference}} + \underbrace{\mathbf{f}_k^T\alpha\mathbf{W}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i}_{\text{inter-pair interference}} \\
&\quad + \underbrace{\mathbf{f}_k^T\alpha\mathbf{W}\widetilde{\mathbf{G}}_{RR}\tilde{\mathbf{x}}_R}_{\text{amplified SI from relay}} + \underbrace{\mathbf{f}_k^T\alpha\mathbf{W}\mathbf{z}_R}_{\text{amplified noise from relay}} + \underbrace{\sum_{i,k\in U_k}\Omega_{k,i}\sqrt{p_i}x_i}_{\text{SI at user and IUI}} + \underbrace{z_k}_{\text{AWGN at } S_k}.
\end{aligned} \tag{4.41}$$

Here (k, k') denotes the user pair exchanging information.

The instantaneous SINR of user S_k can be written using (4.41) as

$$\mathrm{SINR}_k = \frac{p_{k'}|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|^2}{p_k|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_k|^2+\sum\limits_{i\neq k,k'}^{2K} p_i|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_i|^2+\|\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{RR}\|^2\frac{P_R}{N}+\|\mathbf{f}_k^T\mathbf{W}\|^2\sigma_{n_R}^2+\frac{1}{\alpha^2}\sum\limits_{i,k\in U_k}\sigma_{k,i}^2 p_i+\frac{1}{\alpha^2}\sigma_n^2}. \tag{4.42}$$

Recall that $\alpha \in \{\alpha_{\mathrm{mrc}}, \alpha_{\mathrm{zf}}\}$. The sum rate of the considered system model can be written as

$$R = \mathbb{E}\left\{\sum_{k=1}^{2K}\log_2\left(1+\mathrm{SINR}_k\right)\right\}. \tag{4.43}$$

4.10 Performance Analysis

We now provide the SE results for considered multi-pair two-way full-duplex and half-duplex AF relay system model. For this study, we consider the following settings:

- Noise variances as $\sigma_n^2 = \sigma_{nr}^2 = \sigma^2 = 0\,\mathrm{dB}$, and the system SNR is defined as $\mathrm{SNR} = P_R/\sigma^2$.
- $K = 10$ user pairs and $N = 64{,}256$ relay antennas.
- $p_i = P_R/2K$, for $i = 1, \ldots, 2K$, i.e., equal power to all users.
- $\sigma_{g,k}^2 = \sigma_{f,k}^2$, and the SI at the relay σ_{LIR}^2 as well as the SI and IUI at the user $\sigma_{UI}^2 = \sigma_{k,i}^2 = 0\,\mathrm{dB}$ for $k, i = 1, \ldots, 2K$ with respect to σ^2.

Figures 4.8 and 4.9 show the SE versus SNR and SE versus K comparison of MRC/MRT and ZFR/ZFT beamforming for full-duplex multi-pair system. Note that SE saturates for high SNR values, since the relay SI increases proportionally upon increasing system SNR. The SE initially increases, but for large K it starts decreasing. This is because of the increased inter-pair interference coming from increased number of user pairs.

Figure 4.10 compares the SE versus N for half-duplex and full-duplex system with MRC/MRT beamforming. In our framework, SE of half-duplex massive MIMO relaying can be obtained by substituting the SI and IUI to zero. We see that for full-duplex systems, as we increase the value of SI and IUI, i.e., σ_{LIR}^2 and σ_{UI}^2, the SE of full-duplex system decreases. We see from Fig. 4.10 that the full-duplex system has higher SE than a half-duplex system for $\sigma_{LIR}^2 = \sigma_{UI}^2 = 0\,\mathrm{dB}$. For lower SI suppression, i.e., $\sigma_{LIR}^2 = \sigma_{UI}^2 = 10\,\mathrm{dB}$, the half-duplex performs better than full-duplex. Note that with the increase in N, the rate of increase of SE in case of half-duplex relay is lower as compared to its half-duplex counterpart.

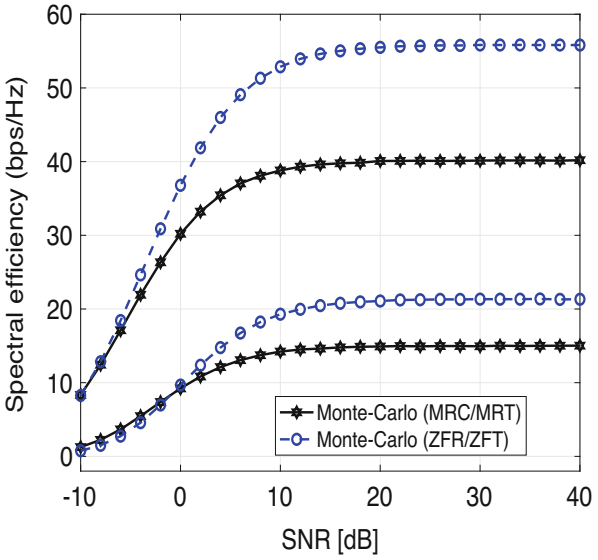


Fig. 4.8 SE versus P_R comparison of MRC/MRT and ZFR/ZFT beamforming, where $K = 10$

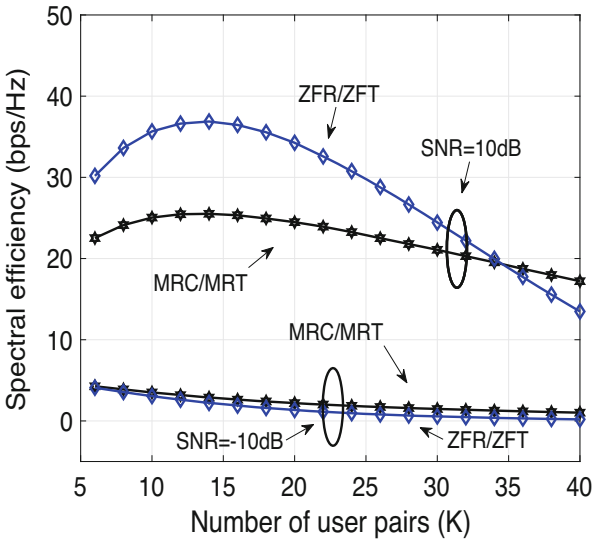


Fig. 4.9 SE versus K with $N = 128$

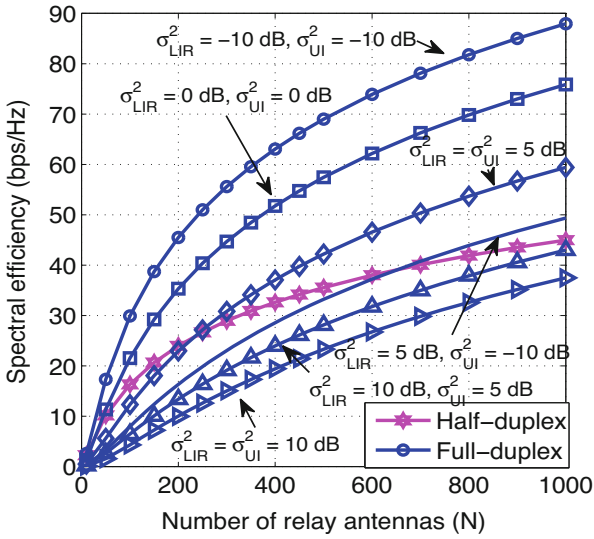


Fig. 4.10 SE versus N for MRC/MRT processing, comparing half-duplex and full-duplex systems. Here SNR = 10 dB

4.11 Summary and Future Works

This chapter considered a system model in which multiple pairs of full-duplex users want to communicate with each other via full-duplex massive MIMO two-way relay. We showed how full-duplex system can help us in achieving better performance than its half-duplex counterpart. The input-output equation derived in this chapter is not only applicable to the considered system model, but the same concept can also be applied to many other 5G and beyond technologies, for example, full-duplex cell-free [43], full-duplex intelligent reflecting surfaces, and full-duplex non-orthogonal multiple access [44], among many others. In this chapter we considered ideal hardware and also ignored channel correlation. The future works can consider imperfect hardware, for example, low-cost low-quality radio-frequency (RF) chains at the relay and the users [45–47]. Due to limited scattering and small antenna spacing, channels between the different antennas of the relay will be correlated [41, 48]. All these practical constraints can be considered to create new research problems for the abovementioned 5G and beyond technologies.

References

1. A. Zappone, E. Jorswieck, Energy efficiency in wireless networks via fractional programming theory. Found. Trends Commun. Inf. Theory **11**(3–4), 185–399 (2014)
2. E. Sharma, R. Budhiraja, K. Vasudevan, L. Hanzo, Full-duplex massive MIMO multi-pair two-way AF relaying: energy efficiency optimization. IEEE Trans. Commun. **66**(8), 3322–3340 (2018)

3. G. Auer, V. Giannini, C. Desset, I. Gódor, P. Skillermark, M. Olsson, M.A. Imran, D. Sabella, M.J. Gonzalez, O. Blume, A.J. Fehske, How much energy is needed to run a wireless network? IEEE Wirel. Commun. **18**(5), 40–49 (2011)
4. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010)
5. E.G. Larsson, O. Edfors, F. Tufvesson, T.L. Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
6. F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
7. L. Lu, G.Y. Li, A.L. Swindlehurst, A.E. Ashikhmin, R. Zhang, An overview of massive MIMO: benefits and challenges. IEEE J. Sel. Top. Signal Process. **8**(5), 742–758 (2014)
8. J. Hoydis, S. ten Brink, M. Debbah, Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013)
9. E. Björnson, E.G. Larsson, T.L. Marzetta, Massive MIMO: ten myths and one critical question. IEEE Commun. Mag. **54**(2), 114–123 (2016)
10. B. Wang, J. Zhang, A. Høst-Madsen, On the capacity of MIMO relay channels. IEEE Trans. Inf. Theory **51**(1), 29–43 (2005)
11. M. Yuksel, E. Erkip, Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective. IEEE Trans. Inf. Theory **53**(10), 3371–3393 (2007)
12. A. Adinoyi, H. Yanikomeroglu, Cooperative relaying in multi-antenna fixed relay networks. IEEE Trans. Wirel. Commun. **6**(2), 533–544 (2007)
13. L. Song, Relay selection for two-way relaying with amplify-and-forward protocols. IEEE Trans. Vehic. Technol. **60**(4), 1954–1959 (2011)
14. Z. Zhang, X. Chai, K. Long, A.V. Vasilakos, L. Hanzo, Full duplex techniques for 5G networks: self-interference cancellation, protocol design, and relay selection. IEEE Commun. Mag. **53**(5), 128–137 (2015)
15. K. Lee, L. Hanzo, Resource-efficient wireless relaying protocols. IEEE Wirel. Commun. **17**(2), 66–72 (2010)
16. A. Sabharwal, P. Schniter, D. Guo, D.W. Bliss, S. Rangarajan, R. Wichman, In-band full-duplex wireless: Challenges and opportunities. IEEE J. Sel. Areas Commun. **32**(9), 1637–1652 (2014)
17. M. Duarte, C. Dick, A. Sabharwal, Experiment-driven characterization of full-duplex wireless systems. IEEE Trans. Wirel. Commun. **11**(12), 4296–4307 (2012)
18. A. Nadh, J. Samuel, A. Sharma, S. Aniruddhan, R.K. Ganti, A taylor series approximation of self-interference channel in full-duplex radios. IEEE Trans. Wirel. Commun. **16**(7), 4304–4316 (2017)
19. T. Riihonen, S. Werner, R. Wichman, Spatial loop interference suppression in full-duplex MIMO relays, in *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers* (IEEE, Piscataway, 2009), pp. 1508–1512
20. Y.Y. Kang, B.-J. Kwak, J.H. Cho, An optimal full-duplex AF relay for joint analog and digital domain self-interference cancellation. IEEE Trans. Commun. **62**(8), 2758–2772 (2014)
21. T. Riihonen, S. Werner, R. Wichman, Mitigation of loopback self-interference in full-duplex MIMO relays. IEEE Trans. Signal Process. **59**(12), 5983–5993 (2011)
22. Z. Zhang, Z. Ma, Z. Ding, M. Xiao, G.K. Karagiannidis, Full-duplex two-way and one-way relaying: Average rate, outage probability, and tradeoffs. IEEE Trans. Wirel. Commun. **15**(6), 3920–3933 (2016)
23. H.Q. Ngo, H.A. Suraweera, M. Matthaiou, E.G. Larsson, Multipair full-duplex relaying with massive arrays and linear processing. IEEE J. Sel. Areas Commun. **32**(9), 1721–1737 (2014)
24. Z. Zhang, Z. Chen, M. Shen, B. Xia, Spectral and energy efficiency of multipair two-way full-duplex relay systems with massive MIMO. IEEE J. Sel. Areas Commun. **34**(4), 848–863 (2016)
25. Z. Zhang, Z. Chen, M. Shen, B. Xia, W. Xie, Y. Zhao, Performance analysis for training-based multipair two-way full-duplex relaying with massive antennas. IEEE Trans. Vehic. Technol. **66**(7), 6130–6145 (2017)

26. E. Sharma, S.S. Chauhan, R. Budhiraja, Weighted sum energy efficiency optimization for massive MIMO two-way half-duplex AF relaying. IEEE Wirel. Commun. Lett. **8**(1), 25–28 (2019)
27. H. Cui, M. Ma, L. Song, B. Jiao, Relay selection for two-way full duplex relay networks with amplify-and-forward protocol. IEEE Trans. Wirel. Commun. **13**(7), 3768–3777 (2014)
28. Y. Dai, X. Dong, Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing. IEEE Trans. Wirel. Commun. **15**(9), 5932–5946 (2016)
29. E. Sharma, A.K. Shukla, R. Budhiraja, Spectral- and energy-efficiency of massive MIMO two-way half-duplex hybrid processing AF relay. IEEE Wirel. Commun. Lett. **7**(5), 876–879 (2018)
30. E. Sharma, N. Gupta, S. Dey, R. Budhiraja, Hybrid massive MIMO two-way relaying with users and relay hardware impairments. IEEE Signal Process. Lett. **27**, 486–490 (2020)
31. E. Björnson, J. Hoydis, L. Sanguinetti, Massive MIMO networks: Spectral, energy, and hardware efficiency. Found. Trends Signal Process. **11**(3–4), 154–655 (2017)
32. E. Sharma, D.N. Amudala, R. Budhiraja, Energy efficiency optimization of massive MIMO FD relay with quadratic transform. IEEE Trans. Wirel. Commun. **19**(2), 1429–1448 (2020)
33. B.P. Day, A.R. Margetts, D.W. Bliss, P. Schniter, Full-duplex MIMO relaying: Achievable rates under limited dynamic range. IEEE J. Sel. Areas Commun. **30**(8), 1541–1553 (2012)
34. I. Krikidis, H.A. Suraweera, P.J. Smith, C. Yuen, Full-duplex relay selection for amplify-and-forward cooperative networks. IEEE Trans. Wirel. Commun. **11**(12), 4381–4393 (2012)
35. T. Riihonen, S. Werner, R. Wichman, Hybrid full-duplex/half-duplex relaying with transmit power adaptation. IEEE Trans. Wirel. Commun. **10**(9), 3074–3085 (2011)
36. J. Lee, Full-duplex relay for enhancing physical layer security in multi-hop relaying systems. IEEE Commun. Lett. **19**(4), 525–528 (2015)
37. E. Everett, A. Sahai, A. Sabharwal, Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans. Wirel. Commun. **13**(2), 680–694 (2014)
38. X. Xia, D. Zhang, K. Xu, W. Ma, Y. Xu, Hardware impairments aware transceiver for full-duplex massive MIMO relaying. IEEE Trans. Signal Process. **63**(24), 6565–6580 (2015)
39. X. Xiong, X. Wang, T. Riihonen, X. You, Channel estimation for full-duplex relay systems with large-scale antenna arrays. IEEE Trans. Wirel. Commun. **15**(10), 6925–6938 (2016)
40. E. Sharma, R. Budhiraja, K. Vasudevan, Multi-pair two way AF full-duplex massive MIMO relaying with ZFR/ZFT processing, in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (2017), pp. 1–7
41. D.N. Amudala, E. Sharma, R. Budhiraja, Spectral efficiency optimization of spatially-correlated multi-pair full-duplex massive MIMO relaying. IEEE Trans. Commun. **67**(12), 8346–8364 (2019)
42. P. Xing, J. Liu, C. Zhai, X. Wang, X. Zhang, Multipair two-way full-duplex relaying with massive array and power allocation. IEEE Trans. Vehic. Technol. **66**(10), 8926–8939 (2017)
43. S. Datta, E. Sharma, D.N. Amudala, R. Budhiraja, S.S. Panwar, Full-duplex cell-free mmimo systems: Analysis and decentralized optimization. CoRR **abs/2010.14110** (2020)
44. V. Mandawaria, E. Sharma, R. Budhiraja, Energy-efficient massive MIMO multi-relay NOMA systems with CSI errors. IEEE Trans. Commun. **68**(12), 7410–7428 (2020)
45. S. Dey, E. Sharma, R. Budhiraja, Impact of user and relay hardware impairments on spectral efficiency of HD massive MIMO relay, in *International Conference on Signal Processing and Communications, SPCOM 2020, Bangalore, July 19–24, 2020* (IEEE, Piscataway, 2020), pp. 1–5
46. S. Dey, E. Sharma, R. Budhiraja, Multi-pair two-way full-duplex massive MIMO relaying with non-ideal hardware, in *2019 IEEE Global Communications Conference, GLOBECOM 2019, Waikoloa, HI, December 9–13, 2019* (IEEE, Piscataway, 2019), pp. 1–6
47. S. Dey, E. Sharma, R. Budhiraja, Scaling analysis of hardware-impaired two-way full-duplex massive MIMO relay. IEEE Commun. Lett. **23**(7), 1249–1253 (2019)
48. N. Gupta, E. Sharma, S. Dey, R. Budhiraja, Spectral efficiency of multi-pair two-way massive MIMO relay with correlated hardware distortion, in *20th IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2019, Cannes, July 2–5, 2019* (IEEE, Piscataway, 2019), pp. 1–5

Chapter 5
NOMA for 5G and Beyond Wireless Networks

Pragya Swami and Vimal Bhatia

5.1 Introduction

The number of users in the wireless network and their data demands are increasing day by day. Also, the expectations of the users connected to the wireless network have gone higher than ever in terms of the services/applications available to them, whether it is for video streaming, online payments, networking, gaming, e-commerce, etc. The Internet of Things (IoT) technology is also developing at a fast pace. Billions of sensors and active devices are installed for different applications, e.g., home security, smart appliances, health monitoring, etc. [1, 2]. The spectrum is constant; however, the number of users, wireless applications, and the associated data usage have increased beyond limits. This has led to spectrum crunch issue [3]. To tackle such problem, either additional spectrum is required or ways to utilize the available spectrum must be searched. For additional spectrum, researchers have proposed to look into the underutilized millimeter-wave (mmWave) frequency range, while to utilize the available spectrum, various techniques are being investigated. One such technique is non-orthogonal multiple access (NOMA), which is the prime focus of this chapter.

The expectations of the fifth generation (5G) and beyond of the wireless communication include spectral efficiency, energy efficiency, and data rate as high as ten times than the one provided by the current fourth generation (4G) of the wireless network [3]. Such breakthrough requirements would also require advanced and groundbreaking wireless technologies. There have been some pioneering

P. Swami (✉) · V. Bhatia
Indian Institute of Technology Indore, Indore, Madhya Pradesh, India
e-mail: vbhatia@iiti.ac.in

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_5

techniques proposed for the 5G and beyond networks, e.g., massive multiple-input multiple-output, millimeter-wave (mmWave) communications, ultra-dense network, visible light communication, cognitive communication, cooperative communication, backscatter communication, NOMA, etc. [3–6]. It is envisioned that NOMA can increase the throughput considerably and lead to massive connectivity owing to its working principle [7, 8]. There have been efforts to include NOMA in 5G and beyond networks completely.

The basic rule of NOMA is to unfollow the traditional orthogonality principle and adopts a new rule where users are served in a non-orthogonal manner. As the name suggests, NOMA serves multiple users using the same frequency and time. This implies that unlike any orthogonal multiple scheme, NOMA is capable of serving higher number of users in the available resource. Therefore, it is fair to say that NOMA will tackle spectrum crunch problem and at the same time will lead to massive connectivity.

5.2 Non-orthogonal Multiple Access

A multiple access scheme by definition means a technique of serving multiple users over the same resource block by applying suitable rules such that the interference is minimum. Traditionally, this can be done by providing orthogonality in frequency, or time, or space, etc. It has been observed that with the advancement of wireless technology from the first generation (1G) of wireless network to the fourth generation of the wireless network, one major change in technology that has occurred is in the multiple access scheme adopted. For instance, 1G used an analog frequency modulation scheme known as frequency division multiple access. Further, the second generation used a digital modulation scheme called as time division multiple access (TDMA). The third generation shifted towards code division multiple access (CDMA), introduced by Qualcomm [9]. However, CDMA was unable to meet high data demands of the 4G networks; therefore, 4G adopted yet another advanced technique, the orthogonal frequency division multiple access (OFDMA) [10]. The orthogonal multiple access (OMA) techniques are widely accepted in the 4G and current 5G (NR2) standards [11, 12].

As the multiple access techniques have advanced from generation to generation, the fifth generation of the wireless networks requires even more advanced method to meet the requirements of massive users and make the system spectrally efficient. Under such circumstances, NOMA is seen as the potential multiple access scheme for the 5G networks. Recently, NOMA has been proposed for the 3GPP long-term evolution (LTE) [13, 14]. It must be noted that in long-term evolution-Advanced (LTE-A), a two-user case of NOMA is implemented [15]. The NOMA technique can be categorized as power domain (PD)-NOMA and code domain (CD)-NOMA (generally referred to as sparse code multiple access). The CD-NOMA technique includes multi-user shared access [16], multiple access solutions relying on low-density spreading [17], successive interference cancellation amenable multiple access [18], sparse code multiple access [19, 20], etc.

The basic and prime difference in the application of PD-NOMA and CD-NOMA is that PD-NOMA exploits power domain while CD-NOMA exploits CODE domain for non-orthogonal resource allocation. This chapter discusses about PD-NOMA owing to its popularity and easier implementation [4]. For employing PD-NOMA, only little or no hardware changes are required in the existing transmitter and receiver operating using orthogonal multiple access technology. In PD-NOMA, multiple users are served in the same resource block, i.e., the same frequency is used at the same time to serve multiple users. This is achieved by having variation in their allocated power. The technique used to make implementation of NOMA successful is, namely, superposition coding and successive interference cancellation (SIC). The messages of multiple users are superimposed using superposition coding and transmitted as shown in Fig. 5.1. One thing to be noted here is that for NOMA, users with different channel conditions are paired together. Application of NOMA considering two users, termed as User 1 (referred to as near user or cell center user) and User 2 (referred to as far user or cell edge user), is illustrated in Fig. 5.1. The messages of User 1 and User 2 are combined using superposition coding. Later, as per the order of users' channel gain conditions, SIC is applied to decode users' messages [7]. Let us assume that the base station (BS) allocates more power to the User 2. This is because User 2 is far away from the BS and hence would require more power as compared to User 1 which is closer to the BS. When User 1 receives the superimposed signal from the BS, it first carries out SIC to detect and remove message of User 2 and then detects its own message. On the other hand, User 2 detects its own message from the superimposed signal by considering the message

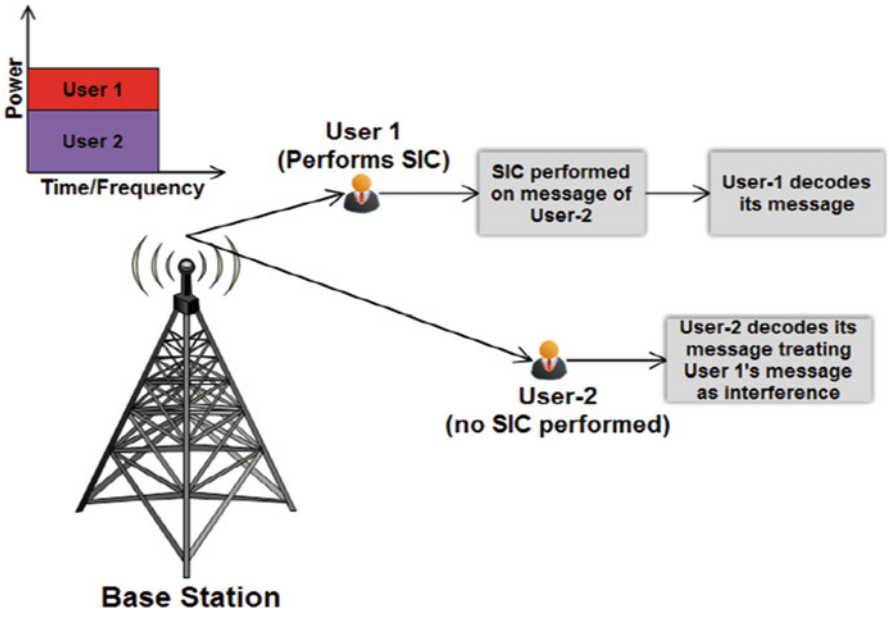


Fig. 5.1 Basic concept of NOMA [21]

of User 1 as interference as described in Fig. 5.1. Therefore, it can be seen that the method of detection of self message differs from User 1 and User 2 after receiving of the same superimposed signal. This difference arises because of the difference in channel gain condition of both the users.

Following this way, NOMA achieves massive connectivity and efficient utilization of spectrum. As per the theoretical research on comparison of NOMA with orthogonal multiple access, it is observed that NOMA achieves higher system throughput [22]. The reason is that although splitting of power takes place in NOMA, however, benefit of larger bandwidth is also present in NOMA. This increases the overall system performance. Furthermore, as the difference in channel gain between the users increases, the system throughput also increases.

5.2.1 Various Applications of NOMA

NOMA provides massive connectivity to support various applications, e.g., for the integrated system such as vehicle-to-everything (V2X) which has been proposed such that vehicles are able to communicate by themselves. For the application of V2X, NOMA has been considered as a key technology to support various simultaneous applications [23]. Furthermore, NOMA with cooperative relaying is a natural extension since the message of far user is also sent to the near user and therefore is capable of transmitting/relaying the message of the far user. This leads to higher reliability [24]. Cooperative relaying only, without NOMA, leads to degradation of spectral efficiency due to transmissions of duplicate messages. With increase in demands by the IoT devices and its underlying challenge of security, NOMA is envisioned as capable of satisfying the secure short-packet communications need in IoT networks [1]. Furthermore, with NOMA, additional security mechanism is also not required by intentionally increasing the co-channel interference at the eavesdropper. Another very interesting application of NOMA is in grant-free transmission. The grant-free transmissions are vulnerable to collisions; however, they achieve low latency which is envisioned for 5G and beyond networks. By exploiting the power domain, and multiplexing the user using different power, NOMA reduces the collisions under grant-free transmission, and this is capable of achieving massive machine-type communication [25, 26].

Another very recent and revolutionizing technique to enable the battery-free IoT is backscatter communication with NOMA. The integration of NOMA and backscatter communication is recognized as spectral and energy-efficient technology for implementing massive IoT networks [27].

One of the additional advantages of NOMA besides massive connectivity and spectral efficiency is its compatibility. Only minimal changes are required to implement NOMA in the existing networks [4]. The popularity and widespread use of superposition coding and SIC (perfect and imperfect SIC [28]) further ease the implementation of NOMA. Therefore, NOMA is considered as a potential multiple access candidate for the 5G and beyond networks.

5.3 Heterogeneous Networks

With increasing data demands in cellular networks and with link efficiency approaching its fundamental limits, there emerges a need to improve spectral efficiency. A possible way of doing this is increasing the number of macro base station (macroBS). However, more macroBSs in an already dense network would imply higher inter-cell interference [29]. Also, deploying additional macroBSs can cost immense expense due to site acquisition in dense areas. An alternate strategy to deal with the ongoing issue is deploying smaller base stations, i.e., base station (BS) with lower power (250 mW to approximately 2 W) and smaller transmission range [30], which also poses lower deployment cost. Such small base stations include pico base stations, femto base station (femtoBS), etc. The small base stations can be overlaid with the macroBS, creating a heterogeneous network (HetNet), which is an accepted design for the LTE-A wireless networks [31]. Figure 5.2 shows a HetNet comprising of macroBS, with transmit power of up to 40 W, overlaid with femtoBSs, which transmit at substantially lower power and are typically deployed in an unplanned manner. Installing small base stations improves coverage and provides capacity gain via higher spatial reuse. Unlike the pico base stations, which are meant for outdoor use, the femtoBSs are designed for indoor users. Since the femtoBSs are particularly user-installed, thereby, they can select the mode of their operation, i.e., which users in the network can access the femtoBSs. The femtoBSs can work in two access modes, open access and closed access [32, 33]. Under

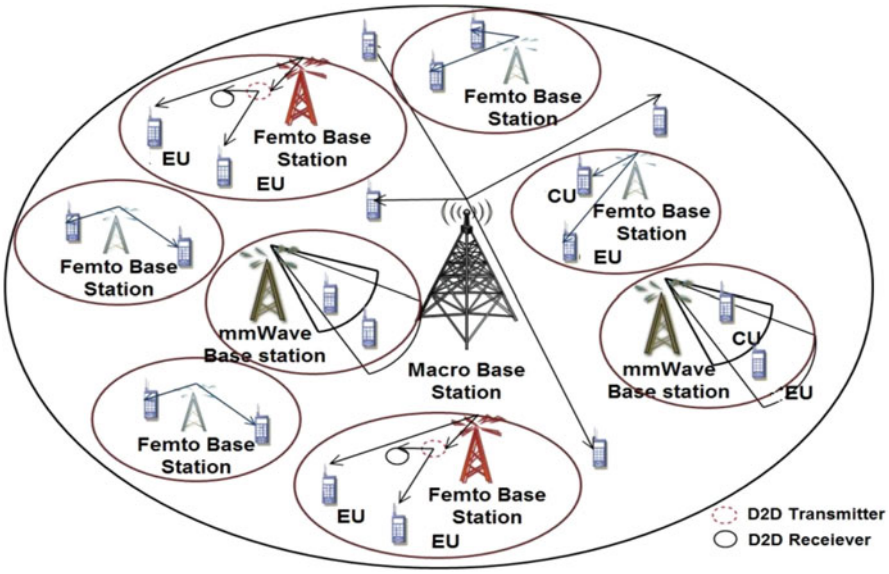


Fig. 5.2 Illustration of NOMA enhanced hybrid HetNet utilizing a mix of macro base stations, femto base stations, D2D transmitters, and mmWave-BSs. CU denote the cell center user and EU represents the cell edge user

closed access, only a closed/fixed group of subscribers (users) have access to the service. However, in open access, any user can connect to the femtoBS, i.e., a user is allowed to connect to any tier without any restrictions.

Since the smaller coverage range leads to lighter loads at the small base stations [34], to relieve the macroBSs of the traffic, the users can be offloaded to the smaller base stations operating under open access. The offloading to small base stations may be biased, i.e., in a manner to have a bias towards admitting users [35], or unbiased. To address the issue of load balancing in HetNets, the contemporary cellular standards, e.g., LTE, use cell range expansion [29] using offloading. Offloading also becomes essential in internet of video things devices [36] due to high computation requirements in visual processing which the Internet of video things are unable to handle.

5.3.1 Tier Selection Probability/Offloading Probability Based on Biased Received Power

The load balancing in the HetNet can be effectively managed using the concept of biasing within tiers leading to flexible association among the tiers. Assuming open access femtoBSs indicates that a user can select either macroBS or femtoBS tier. Considering selection/offloading based on maximum biased received power (BRP) at the user, the offloading probability from the macroBS tier to the femto BS tier may be expressed as [37]

$$\mathcal{P}^{off} = \mathbb{E}_{r_f}\left[\mathcal{P}\left(B_c P_c r_c^{-\alpha_c} < B_f P_f r_f^{-\alpha_f}\right)\right], \tag{5.1}$$

where B_c and B_f represent bias for the macroBS tier and the femtoBS tier, respectively, and α_c and α_f indicate path loss exponent for the macroBS tier and the femtoBS tier, respectively. The r_c and r_f are distance of user from the macroBS tier and the femtoBS tier, respectively. It must be noted that, since the long-term averaged value is considered, fading is averaged out and, thus, is not included. When the bias factors are equal to 1, i.e., $B_m = B_f = 1$, it is termed as an unbiased association and is similar to the conventional association, where the user connects to the BS that offers the strongest average power to the user. However, with $B_f > 1$, the range/coverage of the femtoBS tier is extended.

5.3.2 NOMA-Enabled Heterogeneous Networks

NOMA and HetNets have achieved immense popularity individually. However, the two technologies combined pose more promising advantages due to the reasons mentioned below [38, 39]

- Increased transmit signal-to-interference-and-noise-ratio (SINR): Apart from its several benefits, higher density of BSs in a HetNet results in increased interference, thereby reducing SINR at the user. Using NOMA-enabled BSs enables intelligent tracking of different categories of interference, e.g., inter/intra-tier interference and intra-group interference. This effectively increases the SINR at the users.
- Improved fairness: Fairness among the users is very important for the HetNets, especially when the aim is to study efficient allocation of resource in multi-tier networks. NOMA provides an efficient way to address fairness by giving higher power to user having poor channel condition. Hence, NOMA plays a role of high significance in the HetNets.

The role of macroBS is to serve devices such that the power is required to be split equally among them. However, at the macroBS tier, the role changes. The femtoBS tier serves devices with different data needs and therefore requires different power to achieve them. For example, a femtoBS will serve an Internet of Things (IoT) device along with supporting a high-definition video transmission [40]. Such varied requirements call for appropriate power split as well. Therefore, it becomes suitable to apply NOMA at the femtoBS tier in order to enable power splitting which aids in serving multiple devices with varying requirements, whereas NOMA at macroBS tier is not required.

Furthermore, one additional usage of NOMA at the femtoBS tier will be due to the applied offloading from the macroBS tier to the femtoBS tier. The femtoBS tier is usually installed in crowded places and therefore poses high chances of being fully loaded with its registered users/devices [34]. When offloading occurs, more users are added to the already loaded femtoBS. Hence to support additional users, femtoBS tier can pair the offloaded user with the existing user appropriately and serve both the users using NOMA principle.

5.4 Point Process

The conventional design adopted for the traditional cellular network is grid based. The grid-based model places the BSs (specifically the macroBS) on a grid. Although the Grid-based models are popular and are used extensively, however, they are highly idealized and are not easily tractable. Due to the intractability, complex system-level simulations are used to calculate the outage probability and rate of the network. Furthermore, interference modeling is a challenging task even while considering single-tier cellular networks. The characterization of interference in a single-tier network, assuming the conventional hexagonal grid-based model, requires massive Monte Carlo simulations [9] due to intractability or inaccurate results due to unrealistic assumptions [41]. Thus, a more realistic model with tractable analysis is required, especially for modeling the multi-tier BSs of the HetNets. Practically, due to varied capacity demands in the area of service, the BSs

are far from following a grid-based model. If snapshots of BSs at various locations are checked, the position of the BSs with respect to each other will be random. This makes the grid-based model very idealistic, and its assumptions violate realistic characteristics of a network. Such topological randomness is true for single-tier and multi-tier networks. The design of HetNet requires interference avoidance. Therefore, rigorous yet simple interference models are necessary. Stochastic geometry provides the necessary tools to model the multi-tier networks. Besides capturing the topological randomness of the BSs, stochastic geometry tools are competent in providing tractable analysis [42–44]. In stochastic geometry, the randomness of the BSs is captured using various point processes, appropriate for the given practical scenario. This makes the point process an essential object in stochastic geometry. A point process is defined as a random collection of points in space. The locations of BSs (or users/devices) are the points that are generated using spatial point processes. Detailed discussion on different point processes is given in [43–45]. Some of the popular point processes considered in the existing literature to analyze the cellular networks are Poisson point process [32, 34, 46], hard core point process [47–49], and Poisson cluster process [50–52]. For formal and detailed definitions of the point processes, readers can refer to [43, 53].

5.5 Cooperation Using Device to Device Communication

Device to device (D2D) communication has also received great attention in meeting the booming demands and has appeared to be an essential component in the future generation networks [54, 55]. The key difference of the D2D communication with conventional cellular communication is bypassing the BS, i.e., the communication takes place without traversing the core network. Traditionally, all communication must pass through the core network. Therefore, D2D communication is not transparent to the cellular network. The methodology where the communication takes place through the core network is suitable only for low data rates service, e.g., voice calls, text messages, etc., where the users are not close to each other to have direct communication. However, the advancements of wireless services, for instance, video sharing, gaming, etc., require high data rates, and devices are most possibly in proximity to one another to establish a direct communication (i.e., D2D communication). Therefore, D2D communication is desirable in such scenarios and can potentially increase the spectral efficiency of the network. D2D communication is also capable of improving throughput, energy efficiency, and delay of the network [56]. The D2D communication can occur in two modes, inband (i.e., using the cellular spectrum) and outband (i.e., on the unlicensed spectrum) [56].

Furthermore, to enhance the performance of cellular users, relaying techniques have been explored using D2D communication in cellular networks [54, 57]. Two types of relaying techniques using D2D communications exist: one is relay-assisted D2D communication [58] and other is cooperative D2D communication [54]. In the former, relays assist the transmission to improve D2D user's performance, while in

the latter, D2D users take part in cooperation/relaying for cellular users to improve their performance.

One of the prime utilizations of SIC technique used in NOMA can be for cooperative communication. In NOMA, the superimposed signal transmitted to the users contains messages of both the weak user and the strong user. At the strong user, SIC indicates that the signal of the weak user will be decoded before its own message. Since the strong user has access to the weak user's message, therefore, strong user can be effectively used as a relay to enhance the performance at the weak user [59].

5.6 System Model

This chapter investigates a HetNet with three tiers, as shown in Fig. 5.3. The three tiers are, namely, the macroBS tier, the femtoBS tier, and the D2D tier. The D2D tier aids in cooperation for Scenario-II (explained later in the chapter) wherein an appropriate pairing user for the offloaded user could not be found. The femtoBS tier, macroBS tier, and D2D tier work in underlay fashion. The locations of the BSs of the three tiers are distributed in accordance with the PPP assumption. The densities of the three tiers are fixed as is given in Table 5.1. Furthermore, the BSs of the three tiers transmit using a constant power mentioned in Table 5.1. The coverage range of the three tiers is fixed but differs from each other. Since the macroBS tier is for

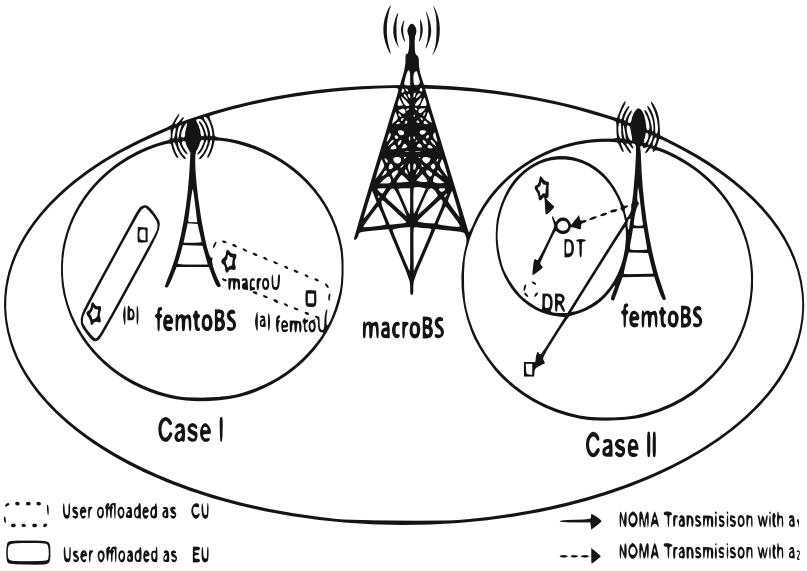


Fig. 5.3 System model [37]

Table 5.1 Simulation parameters

Name	Value taken
Number of simulations	10^3
Power of macroBS, femtoBS, and D2D BSs	40 W, 1 W, 3 mW
Density of macroBS, femtoBS, and D2D tiers	$5 \times 10^{-5}\text{m}^{-2}$, 10^{-4}m^{-2}, and 10^{-4}m^{-2}
Density of users	$5 \times 10^{-4}\text{m}^{-2}$
Bias of macroBS tier and femtoBS tier	1, and 1
Power allocation factor from User 1 and User 2	0.1, and 0.9
Coverage range of macroBS, femtoBS, and D2D BSs	1000 m, 5 m, and 2 m
Channel exponent coefficients from macroBS tier, femtoBS tier, and D2D tier	3, 4, and 3

providing coverage to a larger area as compared to the femtoBS tier and D2D tier, the coverage range of the macroBS tier is largest followed by the coverage range of the femtoBS tier and then the D2D tier. The D2D tier has the smallest coverage range as compared to the macroBS tier and the femtoBS tier. The exact values of the considered coverage range of the macroBS tier, femtoBS tier, and the D2D tier can be found in Table 5.1. It is further assumed that at the D2D tier, the D2D transmitter (DTx) has a fixed D2D receiver (DRx), thereby forming a D2D pair. However, in situations when the D2D tier is involved in cooperation, the D2D pair is converted to a D2D group [60] wherein the DTx serves its DRx and the offloaded macroU using NOMA rule. Otherwise the D2D tier does not employ NOMA. However, the femtoBS tier always operates using NOMA.

Furthermore, it is important to note that since this chapter deals with D2D-Coop, it becomes necessary to assume a sufficiently high D2D density such that a DTx is always available whenever cooperation is required. Therefore, this work considers a high D2D tier density, as given in Table 5.1, to rule the case when DTx is unavailable for cooperation.

The connection policy used at the macroBS tier is the commonly used nearest neighbor (NN) connection policy. Furthermore, bounded path loss model is considered. The bounded path loss model ensures the path loss to be smaller than 1 under circumstances where the distance of the connecting link is very small [61]. A system bandwidth of 1Hz is considered for simplicity. Offloading is performed from the macroBS tier to the femtoBS tier for congestion relief at the macroBS tier. This offloading can be biased in favor of either macroBS tier or femtoBS tier by selecting suitable bias factor, explained in Sect. 5.6.2.1. D2D tiers aid in offloading by providing cooperation using NOMA. The transmission is assumed to be successful under non-outage condition, i.e., when the instantaneous rate is higher than the target rate at the user.

For detailed analysis of the considered HetNet, we consider two scenarios to cover all the possibilities as follows.

- Scenario-I: Under this scenario, it is assumed that when the user is offloaded from the macroBS tier, a pairing user is always available with which the offloaded macroU can be served by the femtoBS tier using NOMA. This is much simpler as compared to the next scenario, since the femtoBS tier does not have to find for a pairing user. The offloaded user can either be served as a cell center user (CU) or cell edge user (EU), depending on the channel gain condition with respect to the available pairing user.
- Scenario-II: This scenario differs from Scenario-I because here the pairing user has to be searched in accordance with the channel gain condition at the offloaded user. Therefore, there is also a possibility that an appropriate pairing user is not available. Hence, D2D-Cooperation (D2D-Coop) is used in this scenario only.

5.6.1 Cooperation from D2D Tier

As already discussed, D2D-Coop is applicable only for Scenario-II. Consider the case when the offloaded user is an EU and the pairing user that the femtoBS tier can find for the offloaded macroU is also an EU. Under such circumstances it becomes difficult for the femtoBS tier to pair the offloaded macroU and serve it. This is because for the application of NOMA, channel difference between the two users paired must be significant. If the channel difference does not exist, it may lead to performance degradation [62]. Therefore, when such situation arises, D2D-Coop plays a vital role. The femtoBS searches for a DRx which is a CU, and is termed as cooperating DTx. Since, it is assumed that the density of the D2D tier is sufficiently high, femtoBS can always find a cooperating DTx (or CDT). The femtoBS then pairs the available EU with the cooperating DTx and sends the information intended for the offloaded macroU to the cooperating DTx. As explained earlier, the D2D pair with the selected cooperating DTx is transformed to D2D group (which comprise the DRx and the offloaded macroU). This D2D group uses NOMA to transmit the intended message received from the femtoBS to the offloaded macroU. This way even when suitable pairing user is unavailable at the femtoBS tier under Scenario-II, the offloaded user does not experience outage. Furthermore, whether the offloaded user is served as a CU or EU in the D2D group is decided by the NC probability given in (5.3).

It must be pointed here that the D2D-Coop is used only when the offloaded macroU is an EU and corresponding CU is unavailable for pairing with the offloaded macroU. Under the situation when the offloaded macroU is CU and the femtoBS tier could not find the appropriate pairing user, D2D-Coop leads to unnecessary

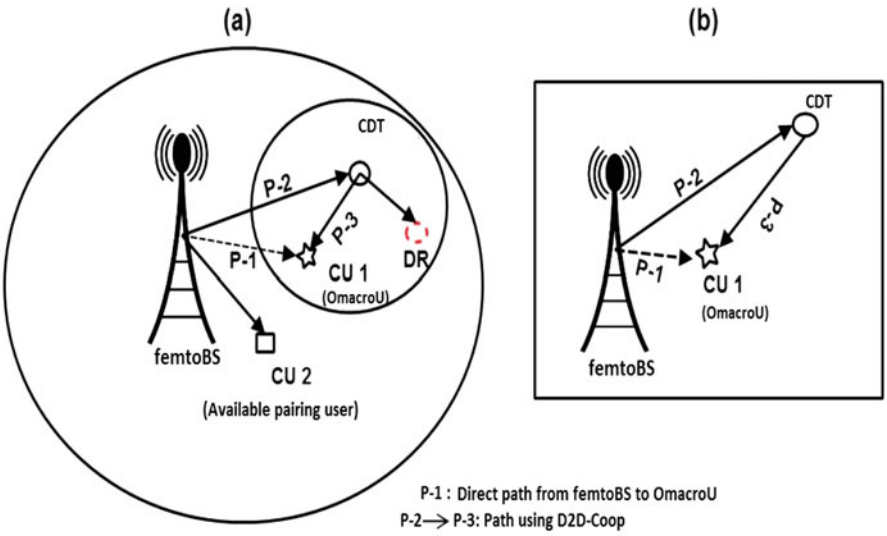


Fig. 5.4 Reason for not using D2D-Coop when the offloaded macroU is CU

power wastage. To make this justification simpler and clearer, we have used figures to explain the underlying concept of not using D2D-Coop when the offloaded macro user (or offloaded user (OmacroU)) is a cell center user (near user) and corresponding pairing user is not available. Figures 5.4 and 5.5 of the chapter show in detail the following case:

- Case-1 (as shown in Fig. 5.4) holds the following conditions:

  1. Offloaded user is a cell center user (near user), denoted by “CU 1."
  2. Available pairing user should be a cell edge user (far user).
  3. However, the available pairing user is another cell center user (and not a cell edge user), denoted by “CU 2." Thus, CU 2 cannot be used for pairing.

  Note: It should be noted that we do not perform D2D-Coop for the given Case-1. In Fig. 5.4, we show graphically why D2D-Coop is not used.

  In Fig. 5.4, P-1 denotes the direct path from femtoBS to the offloaded user (CU 1), and P-2 → P-3 denotes the path using D2D-Coop. As can be seen from Fig. 5.4a, P-2 is a forward path; however, P-3 is the backward path which is unnecessary when the communication can occur directly using path P-1. Hence, we do not perform D2D-Coop for this case.

  To make the visualization simpler, Fig. 5.4b shows an enlarged view of only the paths followed while serving the offloaded user using D2D-Coop for Case-1.
- Case-2 (as shown in Fig. 5.5) holds the following conditions:

  1. Offloaded user is a cell edge user (far user), denoted by “EU 1."
  2. Available pairing user should be a cell center user (near user).

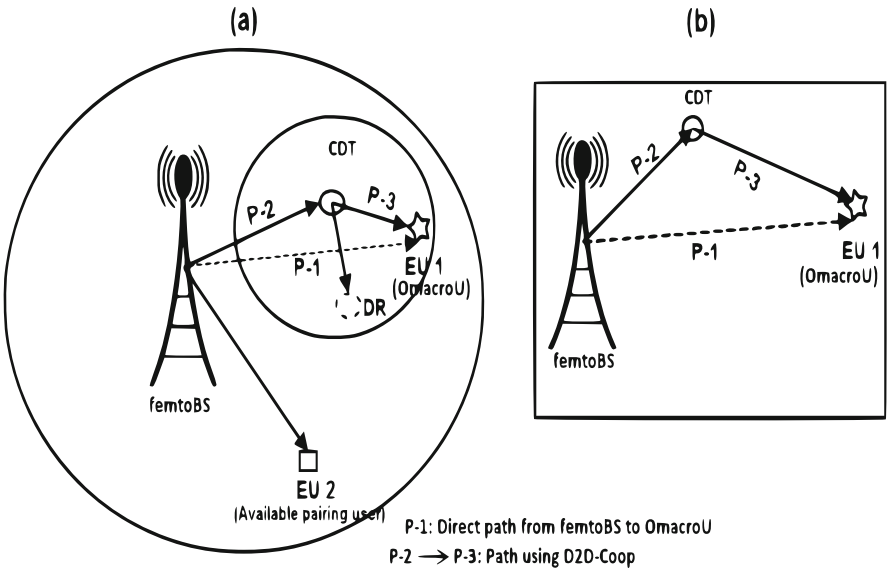


Fig. 5.5 Justification for using D2D-Coop when offloaded user is cell edge user

3. However, available pairing user is another cell edge user (and not a cell center user), denoted by “EU 2." Thus, EU 2 cannot be used for pairing.
4. Hence, D2D-Coop is used.
5. D2D transmitter required for performing D2D-Coop needs to be a cell center user (near user), denoted by “Cooperating DTx."

In Fig. 5.5, P-1 denotes the direct path from femtoBS to the offloaded user (EU 1) and P-2 → P-3 denotes the path using D2D-Coop. It is true that retransmission will take place when the D2D transmitter is cell center user and offloaded user is a cell edge user. However, a forward path will be used, i.e., using P-3 from Fig. 5.5a, unlike a backward path as in Case-1. To make the visualization simpler, Fig. 5.5b shows an enlarged view for only the path followed while serving the offloaded user using D2D-Coop for Case-2. Authors would also like to add that when the offloaded user is a cell edge user, it can also be served using direct path P-1 (as the authors claim for Case-1), since the path using D2D-Coop P-1 + P-3 will be larger than the direct path P-1. It should be noted that the direct path in Case-1, i.e., the path between the femtoBS and CU 1, has a good channel condition, since EU 1 is a center user, which can be utilized efficiently. However, the direct path in Case-2, i.e., the path between the femtoBS and EU 1, has poor channel condition, since EU 1 is a cell edge user. Hence, D2D-Coop is preferred over the direct path.

5.6.2 Some Useful Probabilities

This section highlights probabilities useful in the calculation of the overall system performance.

5.6.2.1 Offloading Probability

Offloading plays a vital role in HetNet to decide whether a macroU will be offloaded from the macroBS tier to the femtoBS tier or will be served by the macroBS tier itself. Such decision-making is based on biased received power rule such that the macroU is served by the tier from which the long-term biased received power is higher [34]. Biasing is done to make the offloading favorable to one particular tier. For instance, if the system requirements are such that the offloading to the femtoBS tier needs to be increased due to severe congestion at the macroBS tier, the bias factor of the femtoBS tier can be chosen to be high. The closed form expression for the offloading probability is expressed in (5.2) as follows [30, 37]

$$\mathcal{P}^{off} = -\frac{3}{8}\mathrm{E}\left(\frac{1}{4}, 2\pi\lambda_m\left(\frac{B_m P_m}{B_f P_f}\right)^{\frac{1}{2}} R_f^{8/3}\right) + \frac{3\Gamma\left(\frac{3}{4}\right)}{8(2\pi)^{3/4} R_f^2\left(\lambda_m\left(\frac{B_m P_m}{B_f P_f}\right)^{\frac{1}{2}}\right)^{3/4}} - \frac{1}{2}e^{-2\pi\lambda_m R_m^2}, \tag{5.2}$$

where B_m and B_f are the bias factor for macroBS and femtoBS tier, respectively. $\mathrm{E}(n, x)$ evaluates the exponential integral as $\mathrm{E}(n, x) = \int_1^\infty e^{-xt}/t^n\, dt$, and $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$ is the complete gamma function.

5.6.2.2 NOMA Compatibility Probability

For NOMA application channel gain condition plays a crucial role. Based on the channel gain condition, the offloaded user can either be treated as a CU or EU at the femtoBS tier. The NOMA compatibility probability defines whether the offloaded macroU is served as a CU or EU by the femtoBS tier with respect to the available pairing user. Let us assume index k to denote the offloaded macroU and index n for the available pairing user. The probability that the offloaded macroU is treated as CU at the femtoBS with respect to the available pairing user is given as $\mathcal{P}_{comp} = \mathcal{P}\left(\frac{|h_n|^2}{|h_k|^2} < q\right)$, where q (satisfying $0 < q < 1$) represents the ratio of channel gain of the PU and the OmacroU. The probability density function for the same is given as [63]

$$f_{\frac{h_n^2}{h_k^2}}(z) = \frac{M_f!}{(n-1)!(-n+k-1)!\,(M_f-k)!} \sum_{j_1=0}^{(n-1)} \sum_{j_2=0}^{(-n+k-1)} \frac{(-1)^{j_1+j_2}\binom{n-1}{j_1}\binom{-n+k-1}{j_2}}{(z\,t_1+t_2)^{\,2}}, \tag{5.3}$$

where $t_1 = j_1 - j_2 + k - n$, and $t_2 = M_f - k + 1 + j_2$. Hence, the probability can be calculated using $\mathcal{P}_{comp} = \int_0^q f_{(h_n^2/h_k^2)}(z)dz$ [37]. NC probability helps us differentiate whether the OmacroU is a CU or EU with respect to the available PU.

5.6.2.3 Corresponding PU Probability

Unlike Scenario-I, where the pairing user is always available, in Scenario-II, the femtoBS is required to search for a corresponding pairing user. When the offloaded macroU is a CU, the femtoBS searched for an EU and when the offloaded macroU is an EU, the femtoBS searches for a CU. Therefore, we derive probability of finding an appropriate pairing user at the FSB tier which is given as follows.

Hence, the probability of at least one user lying in the CU region is given as [37]

$$\mathcal{P}_c = 1 - \frac{\left(1 - e^{\pi(-C)R_m^2}\right)\lambda_m}{C},$$

where $C = \lambda_u \left(\frac{P_f}{P_m}\right)^{\frac{2}{\alpha}} + \lambda_m$.

Similarly, the probability of presence of a pairing user in EU region can be written as [37]

$$\mathcal{P}_e = 1 - \frac{\left(1 - e^{\pi(-C)R_m^2}\right)e^{-\lambda_u \pi R_f^2}\lambda_m}{C}. \tag{5.4}$$

where $C = \lambda_u \left(\frac{P_f}{P_m}\right)^{\frac{2}{\alpha}} + \lambda_m$.

For more details on the derivation, readers may refer to [37].

5.6.3 *Total Outage Probability*

Total outage probability indicates the outage probability at the macroU after the decision of offloading is made. If the decision is in favor of offloading, then the macroU is served by the femtoBS tier; else is served using the macroBS tier which is highlighted in the total outage probability expressions. Furthermore, whether the

offloaded macroU is served as a cell edge user of cell center user at the femtoBS tier is also reflected in the total outage probability. For Scenario-I, it assumed that the pairing user is always available, while for Scenario-II, pairing user is searched by the femtoBS tier; hence, the total outage probability also contains the corresponding PU probability.

5.6.3.1 Total Outage Probability After Offloading

For total outage probability, we combine the outage probability, the offloading probability, and the NC probability and get the expressions as follows [37]:

- When the macroU is offloaded to femtoBS tier and served by the femtoBS tier which does not employ NOMA

$$O_t = (1 - O^{off})O_m + \mathcal{P}^{off} O_f. \tag{5.5}$$

- When macroU is offloaded to the femtoBS tier and served as a CU at femtoBS tier using NOMA

$$O_t^c = (1 - \mathcal{P}^{off})O_m + \mathcal{P}^{off}\mathcal{P}_{comp}O_f^k. \tag{5.6}$$

- When the macroU is offloaded as an EU at femtoBS (with NOMA)

$$O_t^e = (1 - \mathcal{P}^{off})O^m + \mathcal{P}^{off}(1 - \mathcal{P}_{comp})O_f^k. \tag{5.7}$$

Note: Please note that O_m stands for outage probability at macroBS tier, O_f denotes the outage probability at the femtoBS tier when NOMA is not used, and O_f^k is the outage probability at the k^{th} when femtoBS employs NOMA.

Next, for Scenario-II, we calculate the total outage probability when an appropriate pairing user is searched by the femtoBS for the offloaded macroU. The selection of the pairing users is decided on whether the offloaded macroU is an EU or CU. Therefore, the total probability is written as follows [37]:

- When the macroU is offloaded as a CU and the femtoBS tier ed for pairing

$$O_t^c = (1 - \mathcal{P}^{off})O_m + \mathcal{P}^{off}\mathcal{P}_e O_f^k + \mathcal{P}^{off}(1 - \mathcal{P}_e). \tag{5.8}$$

- When the macroU is offloaded as an EU and a CU is searched for pairing

$$\mathcal{P}_t^e = (1 - \mathcal{P}^{off})O_m + \mathcal{P}^{off}\mathcal{P}_c O_f^k + \mathcal{P}^{off}(1 - \mathcal{P}_c). \tag{5.9}$$

5.6.3.2 Total Outage Probability After D2D Cooperation

Under Scenario-I, it is assumed that pairing user is always available. However, for Scenario-II, an appropriate pairing user needs to be searched for. Under situations of unavailability of corresponding pairing user for the offloaded macroU, D2D-Coop becomes useful so that the offloaded user does not fall into outage. As explained in Sect. 5.6.1, D2D-Coop is used only when the offloaded macroU is an EU and a corresponding CU is unavailable. Therefore, the femtoBS searches for a DTx which can act as a CU. This selected DTx is termed as cooperating DTx, and the designated pair converts to a D2D group and performs cooperation as explained in Sect. 5.6.1. The total outage probability at the offloaded macroU after D2D-Coop is given as follows [37].

- When the OmacroU is served as a CU by the cooperating DTx

$$\mathcal{P}^{c}_{coop} = (1 - \mathcal{P}^{off})\mathcal{O}_m + \mathcal{P}^{off}\mathcal{P}_c\mathcal{O}^{k}_{f} + \mathcal{P}^{off}(1 - \mathcal{P}_c)\mathcal{P}_{comp}\left(\mathcal{O}^{k}_{d}\right). \tag{5.10}$$

- When the OmacroU is served as an EU by the cooperating DTx

$$\mathcal{P}^{e}_{coop} = (1-\mathcal{P}^{off})\mathcal{O}_m+\mathcal{P}^{off}\mathcal{P}_c\mathcal{P}^{k}_{f}+\mathcal{P}^{off}(1-\mathcal{P}_c)(1-\mathcal{P}_{comp})\left(\mathcal{P}^{k}_{d}\right). \tag{5.11}$$

Note: $\mathcal{P}^{k}_{d}$ stands for the outage probability at the D2D tier using NOMA.

5.7 Results and Discussions

The outage probability is studied in this section in accordance with the simulation results. The parameters' values for the plots are given in Table 5.1. The outage probability performance for both the considered scenarios, i.e., Scenario-I and Scenario-II, is analyzed.

Firstly, Fig. 5.6 depicts the outage probability with respect to the transmit SNR. The transmit SNR refers to the SNR at the source, and the concept has been used widely, e.g., [64, 65]. The outage probability is plotted individually for all the three considered tiers, i.e., the macroBS tier, the femtoBS tier, and the D2D tier. As discussed, the macroBS tier is deployed for providing coverage to a larger area; therefore, their coverage range is also large. Due to the large coverage range, the path loss experienced at the macroBS tier is also the largest as compared to the femtoBS tier and the D2D tier. Due to high path loss, the outage probability at the macroBS tier is highest as compared to the femtoBS tier and the D2D tier. However, the macroBS tier is capable of serving users covering a large area. Similarly, the outage probability at the D2D tier is the lowest as compared to the femtoBS tier and the macroBS tier. This is because the D2D tier has smallest coverage range. Due

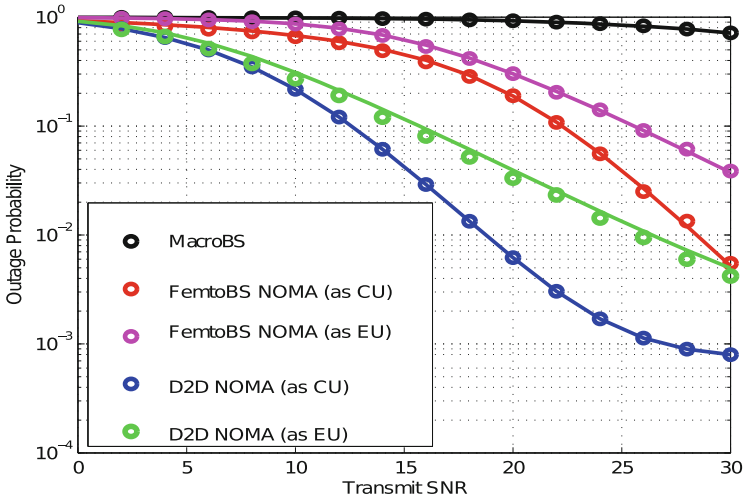
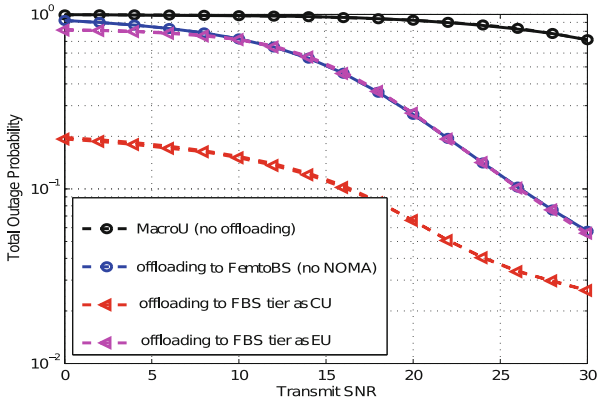


Fig. 5.6 Outage probability at macroBS, femtoBS, and D2D tier [37]

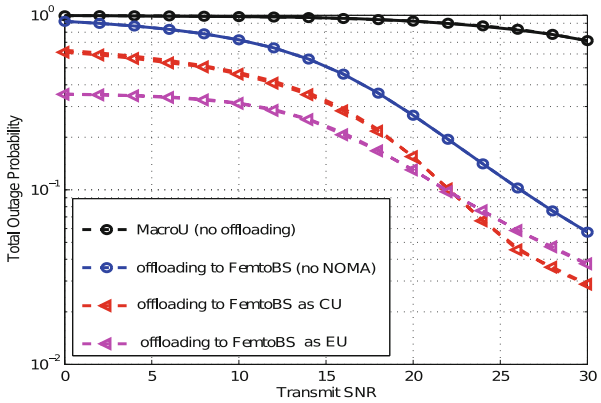
to the small coverage range, the path loss is also low, thereby resulting in lowest outage performance. However, it must be noted that the D2D tier can serve only limited coverage area.

Figure 5.7 shows the total outage probability for Scenario-I wherein offloading takes place from the macroBS tier to the femtoBS tier, and it is pre-assumed that pairing user is available at the femtoBS tier. Also, since femtoBS tier uses NOMA to serve the offloaded macroU, the outage probability of different values of q is plotted to show the impact of channel gain difference between the offloaded macroU and the available PU at the femtoBS on the outage probability of the offloaded macroU. The first thing to observe in all the plots of Fig. 5.7 is that offloading always yields better performance as compared to when no offloading is performed and the macroU is served only by the macroBS tier. Furthermore, Fig. 5.7a reveals the plot for $q = 0.1$, which indicates a large difference in the channel gain difference between the offloaded macroU and the pairing user. A large channel gain difference indicates that the designated CU will have a much better channel gain as compared to the designated EU. As a result, it is evident from Fig. 5.7a that when the offloaded macroU is served as a CU at the femtoBS tier, the offloaded macroU achieves an enhanced outage probability. On the other hand, under the given channel gain difference, if the offloaded user is served as an EU, the outage probability achieved is nearly similar to when the offloading is performed at the femtoBS tier without NOMA.

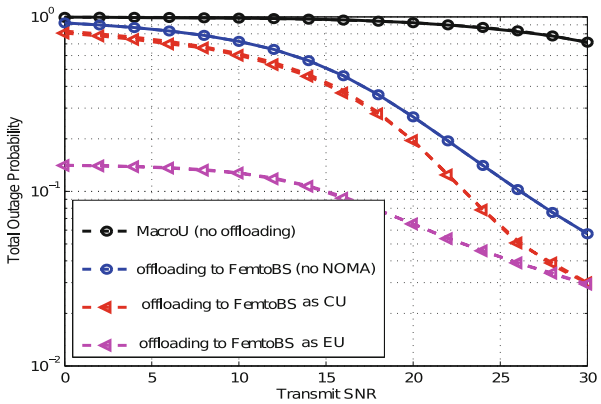
Figure 5.7b is plotted for $q = 0.5$. As the value of q increases, it indicates that the difference in channel gain between the offloaded macroU and the pairing user decreases. It is further observed that whether offloaded macroU is served as CU or EU, the outage performance is better as compared to when offloading is done at the femtoBS tier without NOMA. Under the selected value of q, it is seen that the



(a)



(b)



(c)

Fig. 5.7 Outage probability for various q (Scenario-I) [37]. (**a**) For $q = 0.1$. (**b**) For $q = 0.5$. (**c**) For $q = 0.8$

outage probability at the offloaded user, whether it is served as CU or EU, is nearly the same. This contributes to yet another important aspect of NOMA which is user fairness. It can be noticed that the difference in outage performance between the offloaded macroU as CU and EU is lowest for $q = 0.5$ as compared to the other two selected values of q, i.e., for $q = 0.1$ and $q = 0.8$. This is due to the fact that we have assumed constant power allocation factors for CU and EU. Therefore, although the channel difference gets lower between the CU and EU, the power allocation remains the same. This leads to nearly similar performance between the CU and EU leading to user fairness which is desirable in NOMA application.

Lastly, Fig. 5.7c reveals that the outage probability for $q = 0.8$. It can be noticed that for high values of q, the difference between the CU and EU becomes very low. Therefore, in such case, the best performance is achieved when the offloaded macroU is served as EU at the femtoBS. This is again attributed to the fact that although the channel difference is lowered, the power allocation factors remain constant. Further, it must be noted that whether the offloaded macroU is served as CU or EU, the outage performance achieved is better as compared to when offloading is done to femtoBS tier without NOMA.

Figure 5.8 reveals plots for outage probability under Scenario-II when it is assumed that the femtoBS tier first searched for an appropriate pairing user for the offloaded macroU. It is evident from the figure that the outage probability achieved at the offloaded macroU for Scenario-II is degraded as compared to Scenario-I. The explanation for this degradation is straightforward. Under Scenario-I, the availability of pairing user for the offloaded macroU is guaranteed. However, for Scenario-II, the femtoBS may or may not find corresponding pairing user for the

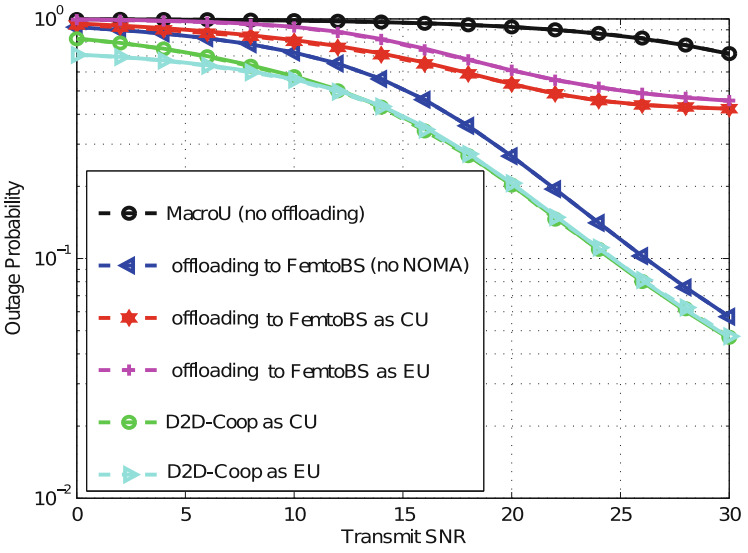


Fig. 5.8 Total outage probability (Scenario-II, $q = 0.5$) [37]

offloaded macroU. Due to the uncertainty of finding a pairing user, the outage probability at the offloaded macroU degrades. To cover this degradation in the outage probability at the offloaded macroU, D2D-Coop is used. It can be noticed that using D2D-Coop achieves significant enhancement in the outage performance, which is an essential requirement for the 5G and beyond networks. Also, whether the offloaded macroU is served as CU or EU by the cooperating DTx, nearly same performance is achieved after transmit SNR of 10 dB. For transmit SNR less than 10 dB, the outage probability achieved by offloaded macroU served as CU by the cooperating DTx is higher as compared when served as EU. The reason is higher interference experienced by the CU due to good channel condition as compared to the EU. Using D2D-Coop in Scenario-II, the offloaded macroU observes a decrease in outage probability by nearly 43%, 55%, 74%, and 87% at transmit SNR of 10 dB, 16 dB, 22 dB, and 28 dB, respectively, when the offloaded macroU is served as an EU by the cooperating DTx, in comparison to when no D2D-Coop is used. Similarly, D2D-Coop outperforms the offloading to femtoBS tier without NOMA and achieves enhancement of 21.54% in the outage probability. The performance enhancement achieved using NOMA and D2D-Coop justifies the proposed system model for the 5G and beyond networks. Moreover, NOMA results in higher connectivity which is very essential for the 5G and beyond networks [5].

5.8 Conclusion

In this chapter, an analytical framework is developed and investigated for a three-tier HetNet with macroBS tier, femtoBS tier, and D2D tier under offloading. The spatial distribution of BSs follows PPP model. NOMA is employed at the femtoBS tier and at the D2D tier. At the D2D tier, the application of NOMA is conditioned on whether D2D-Coop is required by the offloaded macroU. NOMA at the femtoBS tier supports the offloaded users from the macroBS tier, whereas at the D2D tier, NOMA is useful to perform D2D-Coop. It is observed that using the proposed D2D-Coop the performance of offloading in HetNet can be further enhanced and therefore is suitable for the 5G and beyond networks. Furthermore, offloading with NOMA leads to massive connectivity, which is vital for the 5G and beyond networks.

References

1. Z. Xiang, W. Yang, Y. Cai, Z. Ding, Y. Song, Y. Zou, NOMA-assisted secure short-packet communications in IoT. IEEE Wirel. Commun. **27**(4), 8–15 (2020)
2. X. Liu, B. Lin, M. Zhou, M. Jia, NOMA-based cognitive spectrum access for 5G-enabled internet of things. IEEE Network **35**, 290–297 (2021)
3. C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H.M. Aggoune, H. Haas, S. Fletcher, E. Hepsaydir, Cellular architecture and key technologies for 5G wireless communication networks. IEEE Commun. Mag. **52**(2), 122–130 (2014)

4. Z. Wei, J. Yuan, D.W.K. Ng, M. Elkashlan, Z. Ding, A survey of downlink non-orthogonal multiple access for 5G wireless communication networks (2016). Preprint arXiv:1609.01856
5. M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, M. Zorzi, Toward 6G networks: use cases and technologies. IEEE Commun. Mag. **58**(3), 55–61 (2020)
6. R. Mitra, V. Bhatia, Precoded chebyshev-nlms-based pre-distorter for nonlinear led compensation in noma-vlc. IEEE Trans. Commun. **65**(11), 4845–4856 (2017)
7. Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, K. Higuchi, Non-orthogonal multiple access (NOMA) for cellular future radio access, in *IEEE Vehicular Technology Conference (VTC Spring)* (2013), pp. 1–5
8. Z. Ding, P. Fan, H.V. Poor, Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. IEEE Trans. Veh. Technol. **65**(8), 6010–6023 (2016)
9. K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, C.E. Wheatley, On the capacity of a cellular CDMA system. IEEE Trans. Veh. Technol. **40**(2), 303–312 (1991)
10. J. Li, X. Wu, R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach* (Cambridge University Press, Cambridge, 2013)
11. E. LTE, Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN) (3GPP TS 36.300, version 8.11. 0 release 8), December 2009. ETSI TS, vol. 136, no. 300 (2011), p. V8
12. E.U.T.R. Access, Further advancements for E-UTRA physical layer aspects. 3GPP TR 36.814, Technical Report (2010)
13. Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, H.V. Poor, Application of non-orthogonal multiple access in LTE and 5G networks. IEEE Commun. Mag. **55**(2), 185–191 (2017)
14. Y. Saito, A. Benjebbour, Y. Kishiyama, T. Nakamura, System-level performance evaluation of downlink non-orthogonal multiple access (NOMA), in *IEEE 24th Annual International Symposium on Personal, Indoor Mobile Radio Communications (PIMRC)* (2013), pp. 611–615
15. J. Meredith, Study on channel model for frequency spectrum above 6 GHz, 3GPP TR 38.900, Jun, Technical Report (2016)
16. Z. Yuan, G. Yu, W. Li, Multi-user shared access for 5G. Telecommun. Netw. Technol. **5**(5), 28–30 (2015)
17. R. Hoshyar, F.P. Wathan, R. Tafazolli, Novel low-density signature for synchronous CDMA systems over AWGN channel. IEEE Trans. Signal Process. **56**(4), 1616–1626 (2008)
18. X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, J. Xu, Successive interference cancelation amenable multiple access (SAMA) for future wireless communications, in *IEEE International Conference on COMmunication Systems* (IEEE, Piscataway, 2014), pp. 222–226
19. S. Sharma, K. Deka, V. Bhatia, A. Gupta, Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond. IEEE Commun. Lett. **23**(6), 971–974 (2019)
20. R. Mitra, S. Sharma, G. Kaddoum, V. Bhatia, Color-domain SCMA NOMA for visible light communication. IEEE Commun. Lett. **25**(1), 200–204 (2020)
21. V. Bhatia, P. Swami, S. Sharma, R. Mitra, Non-orthogonal multiple access: An enabler for massive connectivity. J. Indian Inst. Sci. **100**(2), 337–348 (2020)
22. L. Lv, J. Chen, Q. Ni, Z. Ding, H. Jiang, Cognitive non-orthogonal multiple access with cooperative relaying: a new wireless frontier for 5G spectrum sharing. IEEE Commun. Mag. **56**(4), 188–195 (2018)
23. B. Di, L. Song, Y. Li, Z. Han, V2X meets NOMA: Non-orthogonal multiple access for 5G-enabled vehicular networks. IEEE Wirel. Commun. **24**(6), 14–21 (2017)
24. M. Zeng, W. Hao, O.A. Dobre, Z. Ding, Cooperative NOMA: state of the art, key techniques, and open challenges. IEEE Netw. **34**(5), 205–211 (2020)
25. C. Zhang, Y. Liu, Z. Ding, Semi-grant-free NOMA: A stochastic geometry model. IEEE Trans. Wirel. Commun. **21**, 1197–1213 (2021)
26. J. Liu, G. Wu, X. Zhang, S. Fang, S. Li, Modeling, Analysis, and Optimization of Grant-Free NOMA in Massive MTC via Stochastic Geometry. IEEE Int. Things J. **8**(6), 4389–4402 (2020)
27. Q. Zhang, L. Zhang, Y.-C. Liang, P.-Y. Kam, Backscatter-NOMA: A symbiotic system of cellular and Internet-of-Things networks. IEEE Access **7**, 20000–20013 (2019)

28. J. Jose, P. Shaik, V. Bhatia, VFD-NOMA under imperfect SIC and residual inter-Relay interference over generalized nakagami-m fading channels. IEEE Commun. Lett. **25**(2), 646–650 (2020)
29. A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, D. Malladi, A survey on 3GPP heterogeneous networks. IEEE Wirel. Commun. **18**(3), 10–21 (2011)
30. S. Singh, H.S. Dhillon, J.G. Andrews, Offloading in heterogeneous networks: Modeling, analysis, and design insights. IEEE Trans. Wirel. Commun. **12**(5), 2484–2497 (2013)
31. Qualcomm, *LTE Advanced: Heterogeneous networks* (2010). http://www.qualcomm.com/common/documents/white papers/LTE Heterogeneous Networks.pdf
32. H.S. Dhillon, R.K. Ganti, F. Baccelli, J.G. Andrews, Modeling and analysis of K-tier downlink heterogeneous cellular networks. IEEE J. Sel. Areas Commun. **30**(3), 550–560 (2012)
33. W. Bao, B. Liang, Stochastic analysis of uplink interference in two-tier femtocell networks: open versus closed access. IEEE Trans. Wirel. Commun. **14**(11), 6200–6215 (2015)
34. H.-S. Jo, Y.J. Sang, P. Xia, J.G. Andrews, Heterogeneous cellular networks with flexible cell association: a comprehensive downlink SINR analysis. IEEE Trans. Wireless Commun. **11**(10), 3484–3495 (2012)
35. S. Parkvall, A. Furuskär, E. Dahlman, Evolution of LTE toward IMT-advanced. IEEE Commun. Mag. **49**(2), 84–91 (2011)
36. F. Guo, H. Lu, B. Li, D. Li, C.W. Chen, Noma-assisted multi-mec offloading for iovt networks. IEEE Wirel. Commun. **28**(4), 26–33 (2021)
37. P. Swami, V. Bhatia, S. Vuppala, T. Ratnarajah, A cooperation scheme for user fairness and performance enhancement in NOMA-HCN. IEEE Trans. Veh. Technol. **67**(12), 11965–11978 (2018)
38. Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, J.A. McCann, Non-orthogonal multiple access in large-scale heterogeneous networks. IEEE J. Sel. Areas Commun. **35**(12), 2667–2680 (2017)
39. A.S. Parihar, P. Swami, V. Bhatia, Z. Ding, Performance analysis of SWIPT enabled cooperative-NOMA in heterogeneous networks using carrier sensing. IEEE Trans. Vehic. Technol. **70**(10), 10646–10656 (2021)
40. Z. Ding, L. Dai, H.V. Poor, MIMO-NOMA design for small packet transmission in the internet of things. IEEE Access **4**, 1393–1405 (2016)
41. J. Xu, J. Zhang, J.G. Andrews, On the accuracy of the Wyner model in cellular networks. IEEE Trans. Wirel. Commun. **10**(9), 3098–3109 (2011)
42. P. Cardieri, Modeling interference in wireless ad hoc networks. IEEE Commun. Surveys Tut. **12**(4), 551–572 (2010)
43. M. Haenggi, *Stochastic Geometry for Wireless Networks* (Cambridge University Press, Cambridge, 2012)
44. H. ElSawy, E. Hossain, M. Haenggi, Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey. IEEE Commun. Surveys Tut. **15**(3), 996–1019 (2013)
45. S.N. Chiu, D. Stoyan, W.S. Kendall, J. Mecke, *Stochastic Geometry and Its Applications* (Wiley, Hoboken, 2013)
46. J.G. Andrews, F. Baccelli, R.K. Ganti, A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
47. H. ElSawy, E. Hossain, A modified hard-core point process for analysis of random CSMA wireless networks in general fading environments. IEEE Trans. Commun. **61**(4), 1520–1534 (2013)
48. G. Alfano, M. Garetto, E. Leonardi, New insights into the stochastic geometry analysis of dense CSMA networks, in *IEEE INFOCOM* (2011), pp. 2642–2650
49. H.Q. Nguyen, F. Baccelli, D. Kofman, A stochastic geometry analysis of dense IEEE 802.11 networks, in *International Conference on Computer Communications (INFOCOM)* (2007), pp. 1199–1207
50. M. Afshang, H.S. Dhillon, Poisson cluster process based analysis of HetNets with correlated user and base station locations. IEEE Trans. Wirel. Commun. **17**(4), 2417–2431 (2018)

51. C. Saha, H.S. Dhillon, N. Miyoshi, J.G. Andrews, Unified analysis of HetNets using Poisson cluster processes under max-power association. IEEE Trans. Wirel. Commun. **18**(8), 3797–3812 (2019)
52. H. Tabassum, E. Hossain, J. Hossain, Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using poisson cluster processes. IEEE Trans. Commun. **65**(8), 3555–3570 (2017)
53. M. Haenggi, R. Ganti, *Interference in Large Wireless Networks* (Now Publishers, Breda, 2008)
54. D. Feng, L. Lu, Y. Yuan-Wu, G.Y. Li, S. Li, G. Feng, Device-to-device communications in cellular networks. IEEE Commun. Mag. **52**(4), 49–55 (2014)
55. H. ElSawy, E. Hossain, M.-S. Alouini, Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks. IEEE Trans. Commun. **62**(11), 4147–4161 (2014)
56. A. Asadi, Q. Wang, V. Mancuso, A survey on device-to-device communication in cellular networks. IEEE Commun. Surveys Tut. **16**(4), 1801–1819 (2014)
57. A. Tang, X. Wang, C. Zhang, Cooperative full duplex device to device communication underlaying cellular networks. IEEE Trans. Wirel. Commun. **16**(12), 7800–7815 (2017)
58. J.-F. Shi, L. Tao, M. Chen, Z.-H. Yang, Power control for relay-assisted device-to-device communication underlaying cellular networks, in *IEEE nternational Conference on Wireless Communications Signal Processing (WCSP)* (2015), pp. 1–6
59. Z. Ding, M. Peng, H.V. Poor, Cooperative non-orthogonal multiple access in 5G systems. IEEE Commun. Lett. **19**(8), 1462–1465 (2015)
60. J. Zhao, Y. Liu, K.K. Chai, Y. Chen, M. Elkashlan, J. Alonso-Zarate, NOMA-based D2D communications: Towards 5G, in *IEEE Global Communications Conference (GLOBECOM)* (2016), pp. 1–6
61. J. Venkataraman, M. Haenggi, O. Collins, Shot noise models for outage and throughput analyses in wireless ad hoc networks, in *IEEE Military Communications Conference MILCOM 2006* (2006), pp. 1–7
62. Z. Ding, P. Fan, H.V. Poor, User pairing in non-orthogonal multiple access downlink transmissions, in *IEEE Global Communications Conference (GLOBECOM)* (2015), pp. 1–5
63. K. Subrahmaniam, On some applications of Mellin transforms to statistics: dependent random variables. SIAM J. Appl. Math. **19**(4), 658–662 (1970)
64. Y. Liu, Z. Ding, M. Elkashlan, J. Yuan, Non-orthogonal multiple access in large-scale underlay cognitive radio networks. IEEE Trans. Veh. Technol. **65**(12), 10152–10157 (2016)
65. L. Lv, H. Jiang, Z. Ding, Q. Ye, N. Al-Dhahir, J. Chen, Secure non-orthogonal multiple access: An interference engineering perspective. IEEE Netw. **35**, 278–285 (2020)

Chapter 6
Energy Harvested Device-to-Device MIMO Systems for Beyond 5G Communication

Parvez Shaik and Vimal Bhatia

6.1 Introduction

Fifth generation (5G) wireless communications through massive machine-type communication (mMTC) enables billions of low-power low-complexity devices to connect [1, 2]. Performance of these energy-constrained devices is limited by their power source especially rechargeable batteries. The vested interest of the researchers toward the motive of automation in 5G communications and green communications has drawn significant attention in industry and academia. To address the energy requirement of the plethora of internet-of-things (IoT) devices, the devices should become self-sustainable and reliable by gleaning energy from the environment. Further, it is standardized by 5G standards and battery life performance indicators [3, 4]. Thus, simultaneous wireless information and power transfer (SWIPT)-enabled communications have drawn prominent research attention beyond 5G communications for providing energy wirelessly.

Energy can be harvested from various sources such as mechanical energy, thermoelectric energy, solar/light energy, and electromagnetic (EM) energy [5]. Energy can be harvested by converting mechanical motion of the device such as vibration and displacement into electricity [6]. By converting the temperature gradient between human bodies and environment, energy can be harvested resulting in thermoelectric energy [6]. However, the energy harvested through these methods is insignificant and cannot enable communication. One of the most amiable resources is solar energy where the incident photons are converted into electricity through

P. Shaik (✉)
Department of Electrical and Communication Engineering, Indian Institute of Sciences (IISc), Bangalore, India

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

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_6

photovoltaic cells. Solar energy can be harvested at a rate of 100 mW/cm2. Energy harvesting can be performed through EM energy which is based on near-field and far-field EM waves. In near-field, energy can be harvested through electromagnetic induction and magnetic resonance, whereas in far-field region, energy can be harvested through radio frequency (RF) signals. In this chapter, the analysis is carried out to address the realization of green communications through EH. Among the available ambient resources, RF electromagnetic signal is the promising resource to harvest the energy.

In the literature there are several seminal works on EH illustrating the potential of EH through SWIPT and presenting the methods to harvest the energy [7–10]. In [7], authors proposed the dedicated sources to transfer the energy wirelessly. Some of the notable works in dual-hop communications is presented to harvested the energy [11–14]. In [11], authors have proposed energy harvesting protocols for delay-limited and delay-tolerant transmission modes. In [15], authors investigated the performance of cooperative D2D multiple-input and multiple-output (MIMO) relay system over Nakagami-m fading channels, optimized the relay location, and performed ASER analysis for hexagonal quadrature amplitude modulation (HQAM), rectangular QAM (RQAM), and cross QAM (XQAM) schemes. In [15], authors investigated the performance of D2D system model without considering the SWIPT techniques and channel imperfections. In [16], authors analyzed cooperative MIMO relay systems performance over Nakagami-m fading channels through the system metrics such as outage probability and asymptotic outage probability and considered the impact of imperfect CSI in the analysis. Further, ASER analysis is performed for HQAM, RQAM, and 32-XQAM schemes. However, the analysis confined to non-SWIPT system model. In [17], authors evaluated the performance of MIMO system model along with SWIPT in terms of outage probability, average capacity, and SEP by considering imperfect CSI. To bridge the gap, authors in [13, 14] analyzed SWIPT-enabled MIMO relay networks through outage probability and asymptotic outage probability and throughput in the presence of channel estimation errors (CEEs) considering both the imperfect CSI and outdated CSI.

In this chapter, we first discuss about the wireless energy harvesting techniques and energy harvesting models. To develop better understanding of the concepts, half-duplex relay-assisted MIMO system model is analyzed over generalized fading Nakagami-m channels. In addition, AF relaying protocol is used at the relay to leverage its low complexity with easy deployment [18]. To attain deeper insights into the EH system model, the impact of EH under practical scenarios such as imperfect channel state information (CSI) and feedback errors are discussed. Analysis of CEEs is presented in terms of system performance metrics such as outage probability and asymptotic outage probability and throughput. The impact of imperfect CSI along with the results is presented first, and later, the impact of feedback errors is discussed.

6.1.1 Wireless Energy Transmission Techniques

The wireless energy transmission schemes are classified as follows [19]:

- **Wireless Power Transfer (WPT):** In the WPT scheme, only power is transmitted through a transmitter to charge their batteries, without establishing a communication link [7].
- **Wireless Powered Communication Network (WPCN):** IN WPCN, energy is harvested from RF signals radiated by an energy transmitter at the wireless receivers, and the harvested energy is used for information exchange between the devices [8, 9]. It is also known as "harvest-then-transmit."
- **SWIPT:** In the SWIPT protocol, a hybrid transmitter is used for wireless energy, and information signals are transmitted over the same waveform to various EH circuits and information decoders. SWIPT has gained research attention due to the high information-energy transmission efficiency [10]. Ideally, the receiver circuits should perform both EH and information processing (IP) simultaneously. However, in the architectural prospective, due to practical circuit limitations, it cannot be realized [13, 19–21]. Hence, the limitation can be relaxed by splitting the signal into two phases: In the first phase, energy is harvested, and in the other phase, communication link is established. In SWIPT, EH can be performed by using time switching (TS) protocol and power splitting (PS) protocol [11] and are represented in Figs. 6.1 and 6.2, respectively. In TS protocol, dedicated

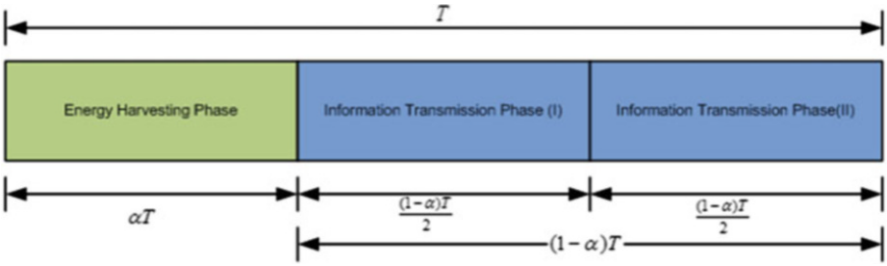


Fig. 6.1 SWIPT: Time switching protocol

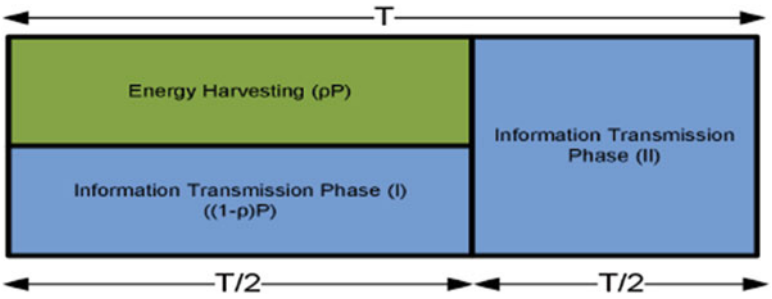


Fig. 6.2 SWIPT: Power splitting protocol

time is allocated for harvesting energy at the energy-constraint node, whereas the information transfer is performed for the rest of the time. In PS protocol, power is divided for both EH and information transmission [11].

6.2 Energy Harvesting Models

EH models are basically classified as linear and nonlinear based on the relation between the RF signal power and the power of the energy harvester.

6.2.1 Linear Energy Harvesting

Linear EH is a de facto standard in various seminal works [8, 11]. In linear EH, the power of the energy harvester is directly proportional to the RF signal power which increases linearly with the input RF signal power. In practice, the devices are nonlinear devices which include active elements such as diodes and transistors, and hence, EH circuits exhibit saturation for a long duration of input power exposure. Since the RF signal is random, thus power harvested is dynamic. If the input signal power falls below the threshold, EH drops to zero and affects the sensitivity of the circuit [22].

6.2.2 Nonlinear Energy Harvesting

There are various seminal works on nonlinear harvesting models to characterize the practical EH circuits based on piece-wise linear function, rational function, a polynomial function, and a sigmoid functions. In piece-wise linear model, linear response is assumed up to the saturation level to capture the practical EH circuit saturation effect [23]. In rational model, seven parameters are used to characterize the nonlinear effect which are determined by curve fitting with measured data [24]. This model is analytically non-tractable. In polynomial model, the diode output is approximated with truncated Taylor expansion [25]. In sigmoid mode, the relation between input RF signal power and output energy harvested power is modeled with a sigmoid function to capture the saturation effect. However, this model fails to retain the sensitivity [26].

6.3 Impact of Imperfect CSI

6.3.1 Mathematical Modeling

Consider a three-terminal device which includes a source (S), relay (R), and the destination (D), where each of these terminal nodes is equipped with multiple antennas. At R node, amplify-and-forward relaying protocol is employed to broadcast the signal to D with the harvested energy [15]. S is considered as a base station, whereas the user equipment (UE) is taken as R and D, respectively. A channel matrix is considered from node P to node Q as $\mathbf{H}_{PQ}$ with dimensions $N_Q \times N_P$, where N_P and N_Q represent the antenna count at the P and Q nodes, respectively. It is highly deterrent to use multiple active antennas due to the requirement for active RF chains, which are cost, power, and size in efficient. Hence, to retain the MIMO gains, transmit antenna selection (TAS) is used at S and R nodes. Hence, the hardware complexity and cost are reduced; however, performance is retained. In TAS, at the node P, an ith transmit antenna is selected which maximizes the received signal-to-noise ratio (SNR) for broadcasting the information through the channel vector $\mathbf{h}_{PQ}^{(i)}$ to the receiver and is given as

$$||\hat{\boldsymbol{h}}_{Q\times 1}|| = \max_{1\le i\le N_P} \{||\hat{\boldsymbol{h}}_{Q\times i}||\}, \tag{6.1}$$

where $P \in \{S, R\}$, $Q \in \{R, D\}$, and $P \neq Q$. Channel matrix $\mathbf{H}_{PQ}$ has the N_P channel vectors ($\mathbf{h}_{PQ}^{(i)}$) with dimensions $N_Q \times 1$ of each. Minimum mean square error (MMSE)-based estimator is used at the receiver for channel estimation. To visualize the versatility of the channel in the system performance, all are assumed to be modeled by the generalize complex Nakagami-m frequency flat fading channels, $Nak(M_L, \hat{\sigma}_{\mathbf{H}}^2)$. All the noise vectors are assumed to be modeled by additive white Gaussian noise (AWGN) vectors with mean 0 and variance $\sigma_{PQ}^2 \mathbf{I}_{N_Q}$. CEEs in channel estimation arises due to the improper pilot pattern. Improper pilot pattern results in irreducible error floor in channel estimation. To avoid CEEs, pilot pattern should be performed as a function of coherence time and frequency [27]. The relation between the actual channel ($\mathbf{H}$) and the estimated channel ($\hat{\mathbf{H}}$) in the presence of CEE ($\boldsymbol{\delta}_{\mathbf{H}}$) is given as

$$\mathbf{H} = \hat{\mathbf{H}} + \boldsymbol{\delta}_{\mathbf{H}}, \tag{6.2}$$

where $\boldsymbol{\delta}_{\mathbf{H}}$ is modeled as $\mathcal{CN}(0, \sigma_{\delta}^2 I)$ and the variance σ_{δ}^2 is given as

$$\sigma_{\delta}^2 = (1-\rho)\sigma_{\mathbf{H}}^2, \tag{6.3}$$

$$\rho = \sigma_{\hat{\mathbf{H}}}^2/\sigma_{\mathbf{H}}^2, \quad 0 < \rho < 1 \tag{6.4}$$

where ρ is the channel correlation coefficient [28]. Variance of the estimated channel is given as $\sigma^2_{\hat{\mathbf{H}}} = \sigma^2_{\mathbf{H}} - \sigma^2_{\delta} = \rho\sigma^2_{\mathbf{H}}$. For harvesting the energy, we consider TS protocol at the R. The system operation completes in two phases. Energy is harvested at R in the first phase, from the received S signals, whereas IP takes place in the second phase.

6.3.1.1 Time Switching Protocol

As discussed, we employ TS protocol to harvest the energy at the R due to its reduced complexity. Entire communication between S and D takes place over a period of (T) time in two phases as shown in Fig. 6.1. The complete time T is divided into αT time and $(1-\alpha)T$ for energy harvesting and IP. α denotes the fraction of time, $0 < \alpha < 1$.

6.3.1.2 Phase 1

In this phase, energy is harvested at R for αT time.

6.3.1.3 Phase 2

In the second phase, communication link is established for the rest over $(1-\alpha)T$ time. $(1-\alpha)T$ time is further divided into two $\frac{(1-\alpha)T}{2}$ equal halves for a dual-hop communication. S broadcasts the information signal to both R and D, simultaneously, in the first $\frac{(1-\alpha)T}{2}$ time, whereas in the other $\frac{(1-\alpha)T}{2}$ time, R broadcast S information signal to the D by utilizing the energy harvested in EH phase. The amount of energy harvested relies on the α value which is allocated to harvest the energy. There is a trade-off between the amount of the throughput achieved and energy harvested. Signal received at R and D is given as

$$\mathbf{y}_{SR} = \sqrt{P_s}(\mathbf{h}^i_{SR} + \boldsymbol{\delta}_{\mathbf{h}^i_{SR}})x + \mathbf{n}_{SR}, \tag{6.5}$$

$$\mathbf{y}_{SD} = \sqrt{P_s}(\mathbf{h}^i_{SD} + \boldsymbol{\delta}_{\mathbf{h}^i_{SD}})x + \mathbf{n}_{SD}, \tag{6.6}$$

respectively, where P_s is the transmit power at the S and $\mathbf{n}_{SR}$ and $\mathbf{n}_{SD}$ are the AWGN vectors corresponding to the respective SR and SD links. It is assumed that $\mathbb{E}\{x^2\} = 1$ [29, 30]. Energy harvested at the R during EH phase is given as [11],

$$E_h = \eta\alpha T P_S ||\mathbf{h}^i_{SR}||^2, \quad 0 < \eta < 1 \tag{6.7}$$

where η is the energy conversion efficiency which is defined by the energy harvesting circuitry and the rectification process [11]. R amplifies the signal received from S to transmit to D through the kth transmit antenna at R with a gain G and is given as [13, 31, 32]

$$G \leq \sqrt{P_R/(P_S||\mathbf{h}_{SR}^i||^4 + \sigma_N^2)} \approx \sqrt{P_R/(P_S||\mathbf{h}_{SR}^i||^4)}, \tag{6.8}$$

$$P_R = \frac{2E_h}{(1-\alpha)T} = \frac{2\eta\alpha P_S||\mathbf{h}_{SR}^i||^2}{(1-\alpha)}, \tag{6.9}$$

where P_R is the transmit power obtained from the energy harvested phase at the R. At D, signal received is given as

$$\mathbf{y}_{RD} = \sqrt{P_R}G\big(\mathbf{h}_{RD}^j + \boldsymbol{\delta}_{\mathbf{h}_{RD}^j}\big)\big(\mathbf{h}_{SR}^i\big)^H\mathbf{y}_{SR} + \mathbf{n}_{RD}, \tag{6.10}$$

where $\mathbf{n}_{RD}$ is the AWGN vector corresponding to the RD link. At D, maximum ratio combiner (MRC)-based optimal receiver filter in the MMSE sense is used to combine the signals received in both the hops[32]. At D, the e2e SNR is given as

$$\Lambda_{e2e}^{(i,j)} = \hat{\Lambda}_{SD}^{(i)} + \hat{\Lambda}_{SRD}^{(i,j)}, \tag{6.11}$$

$$= \frac{\Lambda_{SD}^{(i)}}{\Omega_{SD}} + \hat{\Lambda}_{SRD}^{(i,j)}, \tag{6.12}$$

$$\hat{\Lambda}_{SRD}^{(i,j)} = \frac{\Lambda_{SR}^{(i)}\Lambda_{RD}^{(j)}}{\frac{1}{P_s}\Lambda_{SR}^{(i)} + \Lambda_{RD}^{(j)} + \Lambda_{SR}^{(i)}\hat{\sigma}_{\delta RD}^2 + \Lambda_{RD}^{(j)}\hat{\sigma}_{\delta SR}^2 + \hat{\sigma}_{\delta RD}^2 + \hat{\sigma}_{\delta SR}^2\hat{\sigma}_{\delta RD}^2}. \tag{6.13}$$

where $\hat{\Lambda}_{SD}^{(i)} =$, $\Omega_{SD} = (1 + \hat{\sigma}_{\delta SD}^2)$, $\hat{\sigma}_{\delta SD}^2 = \Lambda_0\sigma_{\delta SD}^2$, $\Lambda_0 = P/\sigma_N^2$ with P transmit power and noise variance (σ_N^2). $\hat{\Lambda}_{SRD}^{(i,k)}$ is approximated as [28]

$$\hat{\Lambda}_{SRD}^{(i,j)} \approx \frac{\frac{\Lambda_{SR}^{(i)}}{\Omega_{SR}}\frac{\Lambda_{RD}^{(j)}}{\Omega_{RD}}}{\left(\frac{\Lambda_{SR}^{(i)}}{\Omega_{SR}} + \frac{\Lambda_{RD}^{(j)}}{\Omega_{RD}}\right)}, \tag{6.14}$$

where $\Omega_{SR} = (1+\hat{\sigma}_{\delta SR}^2)$, $\Omega_{RD} = (\frac{1}{P_s} + \hat{\sigma}_{\delta RD}^2)$, $\hat{\sigma}_{\delta SR}^2 = \Lambda_0\sigma_{\delta SR}^2$, $\hat{\sigma}_{\delta RD}^2 = \frac{K_1}{\sigma_N^2}\sigma_{\delta RD}^2$, $\Lambda_{SD}^{(i)} = \overline{\Lambda}_{SD}||\mathbf{h}_{SD}^i||^2$, $\Lambda_{SR}^{(i)} = \overline{\Lambda}_{SR}||\mathbf{h}_{SR}^i||^2$, $\Lambda_{RD}^{(j)} = \frac{K_1}{\sigma_N^2}||\mathbf{h}_{RD}^j||^2$, and $K_1 = \frac{2\eta\alpha}{(1-\alpha)}$. $\Lambda_{PQ}^{(i)}$ and $\overline{\Lambda}_{PQ}$ are the instantaneous and average SNRs of PQ link, respectively.

6.3.2 System Performance Metrics

6.3.2.1 Outage Probability

A system is said to be in outage when the desired rate of transmission is greater than the maximum possible error-free rate of transmission [33, 34]. Rate of transmission quantifies the threshold. For a rate, $R_{th} = \frac{1-\alpha}{2}\log_2(1+\Lambda_{e2e})$ bits/sec/Hz, the closed-form upper-bound expression for the outage probability is given as [13]

$$\begin{aligned}
\mathrm{P}_{out}^{(UB)}(\Lambda_{th}) &= \Pr\Big[\max_{1\le j\le N_s}\{\hat{\Lambda}_{SD}^{(i)}+\hat{\Lambda}_{SRD}^{(i,j)}\}\le\Lambda_{th}\Big],\\
&\le \Pr\Big[\max_{1\le j\le N_s}\hat{\Lambda}_{SD}^{(i)}\le\Lambda_{th}\Big]\Pr\Big[\max_{1\le j\le N_s}\hat{\Lambda}_{SRD}^{(i,j)}\le\Lambda_{th}\Big],\\
&= F_{\hat{\Lambda}_{SD}^{(i)}}(\Lambda_{th})F_{\hat{\Lambda}_{SRD}^{(i,j)}}(\Lambda_{th}),
\end{aligned}\tag{6.15}$$

The closed-form expressions of $F_{\hat{\Lambda}_{SD}^{(i)}}(\Lambda_{th})$ and $F_{\hat{\Lambda}_{SRD}^{(i,j)}}(\Lambda_{th})$ are given in (6.29) which are derived by following the procedure as in [32]. In (6.29), $\Delta_1 = \frac{p}{\lambda_{SR}}(\frac{\Omega_{SR}}{\Omega_{RD}})+\frac{t+1}{\lambda_{RD}}$, $\Delta_2 = \frac{\Omega_{SR}p(t+1)}{\Omega_{RD}\lambda_{SR}\lambda_{RD}}$, $\Delta = q+n+M_{RD}$, $\vartheta = z-q+1$, $M_{SD} = m_{SD}N_D$, $M_{SR} = m_{SR}N_R$, and $M_{RD} = m_{RD}N_D$. Further, $\lambda_{SD} = \frac{\overline{\Lambda}_{SD}}{m_{SD}}$, $\lambda_{SR} = \frac{\overline{\Lambda}_{SR}}{m_{SR}}$, and $\lambda_{RD} = \frac{K_1}{\sigma_N^2}\mathrm{E}||\mathbf{h}_{RD}^{j}||^2/m_{RD}$. $\phi_{a,b,c}$ is the multinomial theorem coefficient [15].

$$F_{\hat{\Lambda}_{SD}^{(i)}}(\Lambda_{th}) = \sum_{m=0}^{N_S}\sum_{n=0}^{m(M_{SD}-1)}\binom{N_S}{m}(-1)^m\Big(\frac{\Omega_{SD}}{\lambda_{SD}}\Big)^n\phi_{n,m,M_{SD}}\Lambda_{th}^n e^{\frac{-m\Omega_{SD}\Lambda_{th}}{\lambda_{SD}}},$$

$$\begin{aligned}
F_{\hat{\Lambda}_{SRD}^{(i,k)}}(\Lambda_{th}) = 1 &+ \sum_{p=1}^{N_S}\sum_{q=0}^{p(M_{SR}-1)}\sum_{t=0}^{(N_R-1)t(M_{RD}-1)}\sum_{s=0}^{\Delta-1}\sum_{z=0}^{}\frac{(-1)^{p+t}2N_R\phi_{q,p,M_{SR}}\phi_{s,t,M_{RD}}}{\Lambda(M_{RD})\lambda_{RD}^{\frac{2M_{RD}+2s+q-z-1}{2}}}\\
&\times\frac{\Omega_{SR}^{(\frac{2\Delta+z+q+1}{2})}}{(\Omega_{RD}\lambda_{SR})^{\frac{z+q+1}{2}}}\Big(\frac{p}{t+1}\Big)^{\frac{z-q+1}{2}}\binom{N_S}{p}\binom{N_R-1}{t}\binom{\Delta}{z}\Lambda_{th}^{\Delta+1}e^{-\Omega_{SR}\Delta_1\Lambda_{th}}\\
&\times K_{\vartheta}\big(2\Omega_{SR}\sqrt{\Delta_2}\Lambda_{th}\big).
\end{aligned}\tag{6.16}$$

6.3.2.2 Asymptotic Analysis

Asymptotic analysis provides the system deep insights which are useful in modeling the system. It provides information regarding system parameters such as diversity and coding gain which are vital in evaluating the performance of the system. By taking the high SNR approximation of e^x and $K_{\vartheta}(z)$ [35, eq. (1.211.1), Eq. (8.446)], the closed-form asymptotic outage probability expression is derived at $\overline{\Lambda}\to\infty$ as [13]

$$\mathrm{P}_{out}^{\infty}(\Lambda_{th}) = \begin{cases} f_1 f_2 \left(\frac{\Lambda_{th}}{\bar{\Lambda}}\right)^{O_1}, m_{SR}N_S < m_{RD}N_D \\ f_1 f_3 \left(\frac{\Lambda_{th}}{\bar{\Lambda}}\right)^{O_2}, m_{SR}N_S > m_{RD}N_D \\ f_1 (f_2 + f_3) \left(\frac{\Lambda_{th}}{\bar{\Lambda}}\right)^{O_3}, m_{SR}N_S = m_{RD}N_D \end{cases} \tag{6.17}$$

$$f_i = \left(\left(\frac{\Omega_{PQ} m_{SD}}{k_{PQ}} \right)^{M_{PQ}} / (M_{PQ})! \right)^{N_P}, \tag{6.18}$$

For $i = 1$, $PQ = SD$, for $i = 2$, $PQ = SR$, and for $i = 3$, $PQ = RD$. Further, $O_1 = M_{SD}N_S + M_{SR}N_S$, $O_2 = M_{SD}N_S + M_{RD}N_R$, $O_3 = M_{SD}N_S + MN_R$, $M = m_{SR}N_S = m_{RD}N_D$, and $k_{PQ} = \frac{\bar{\Lambda}_{PQ}}{\bar{\Lambda}}$.

6.3.2.3 Throughput Analysis

For a fixed transmission rate, throughput is evaluated in delay-limited transmission mode from the outage probability as [11]

$$\tau = \frac{(1 - \mathrm{P}_{out}) R_{th} (1 - \alpha)}{2}, \tag{6.19}$$

where $\Lambda_{th} = 2^{\frac{2R_{th}}{1-\alpha}} - 1$ is the threshold SNR for a fixed rate R_{th}.

6.3.3 Results and Discussion

In this section, modeled system performance is validated and discussed through the derived analytical expressions of outage probability and asymptotic outage probability and the Monte Carlo simulations. In the system performance analysis, the parameters are considered as follows: channel correlation coefficients are taken as $\{\rho_{SD}, \rho_{SR}, \rho_{RD}\}$, and CSI parameters are considered as $\rho = 0.99$ for perfect case and $\rho = \{0.95, 0.9\}$ for imperfect cases. Further, EH efficiency factor ($\eta = 1$) and different values of α are considered.

In Fig. 6.3, outage probability results are presented for the $2 \times 2 \times 2$ AC with FP $\{1, 1, 1\}$ with $\alpha = 0.2$. In this analysis, three different scenarios are considered such as $\{0.99, 0.99, 0.99\}$, $\{0.95, 0.95, 0.95\}$, and $\{0.9, 0.9, 0.9\}$. Results illustrate the well agreement of the derived analytical results of outage probability and asymptotic outage probability and verified through Monte Carlo simulations. Curves illustrate that the performance of the system is better even after harvesting the energy for $\alpha = 0.2$. Curves at high SNR regime illustrate the loss in diversity order of the

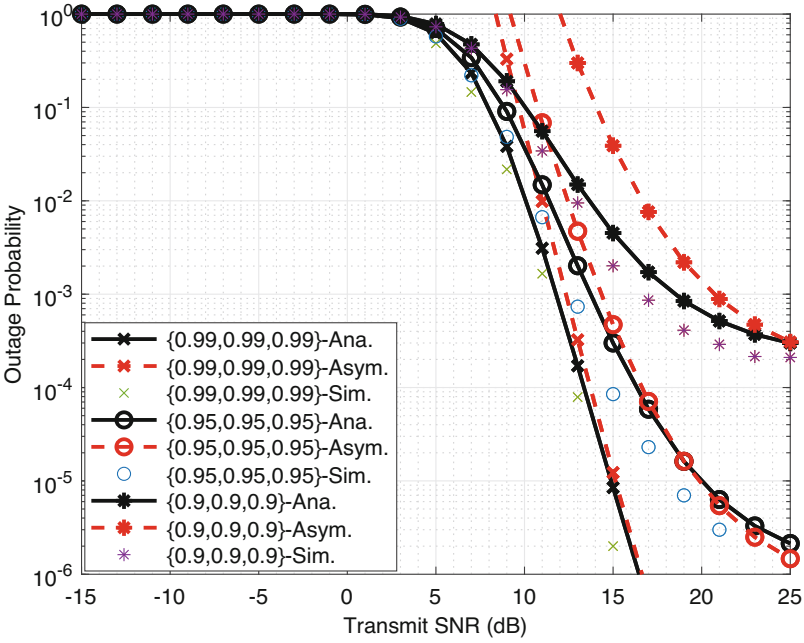


Fig. 6.3 $2 \times 2 \times 2$ MIMO AC: Outage probability vs SNR for different CEE conditions [13]

system with a constant error floor for a minor change in CEE values. This is due to SNR independent CEE variance.

In Fig. 6.4, for $2 \times 1 \times 2$ AC with $\{1, 2, 1\}$ FPs, outage probability analysis is presented to visualize the impact of α under different symmetric CEE conditions. In Fig. 6.4, analytical results of outage probability are presented. The curves demonstrate the degradation of system performance with the increase in α value. With the increase in EH time, error floors increase, and it still increases with the reduction in ρ_{PQ} values. As α increases, EH time increases resulting in a lower IP time, and thus outage probability increases. For an outage probability of 10^{-3} with $\alpha = 0.2$, SNR gain of 6 dB is observed for the system with $\{0.99, 0.99, 0.99\}$ over $\{0.95, 0.95, 0.95\}$ case. As α value increase to 0.5, SNR gain increases to 10 dB for an outage probability of 10^{-2}. It escalates to ∞ gain for $\alpha = 0.7$ over $\{0.95, 0.95, 0.95\}$ case with a constant error floor.

In Fig. 6.5, $2 \times 2 \times 1$-$\{1, 1, 2\}$ AC throughput is attained by the system which is presented for different CEE values at different rates. Curves illustrate that for $R_{th} = 1$ bps/Hz, the system throughput attains saturation with $\tau \sim 0.4$ at 12 dB and 15 dB, respectively, for $\{0.99, 0.99, 0.99\}$ and $\{0.9, 0.9, 0.9\}$ cases. For $R_{th} = 2$ bps/Hz, the throughput of the system saturates to $\sim$0.7 at 22.5 dB approximately for $\{0.99, 0.99, 0.99\}$ case, whereas for imperfect CSI case of $\{0.9, 0.9, 0.9\}$, system throughput saturates to $<$0.3 at $\approx$30 dB. It illustrates the adverse effect of both imperfect CSI and the high data rates over the system performance.

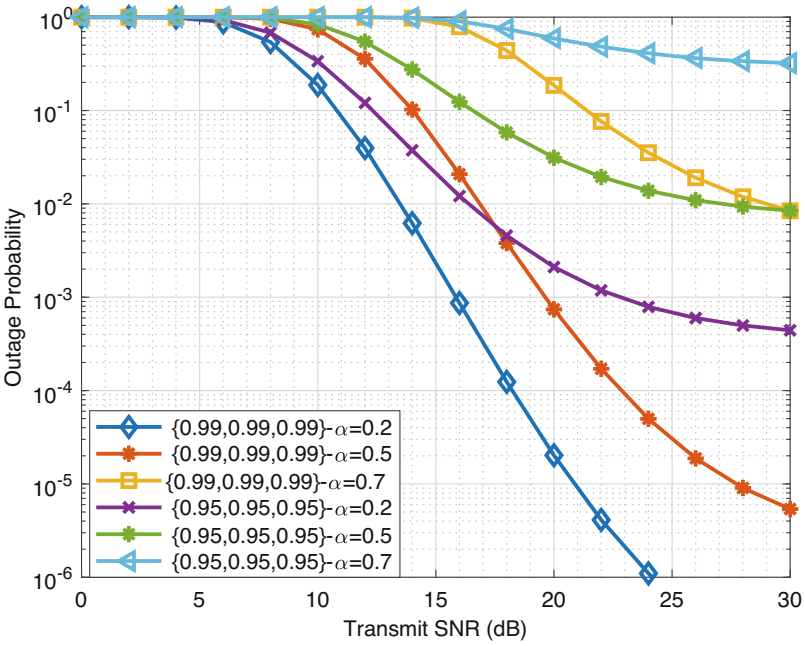


Fig. 6.4 $2 \times 1 \times 2$ MIMO ACs: Outage probability vs SNR for different α values [13]

6.4 Impact of Outdated CSI

In the earlier section, the impact of imperfect CSI over the SWIPT-enabled D2D communications is analyzed. In this section, the outdated CSI and its affects are considered in the analysis. As all the nodes are multi-antenna nodes, the effects of outdated CSI over antenna selection and the EH are important metrics of concern. SWIPT-enabled D2D communications in the presence of feedback error are also highly motivated. Hence, in this section, we present the effects of feedback delays and EH in the analysis of an overlay non-regenerative D2D MIMO relay system over generalized Nakagami-m fading channels. System performance is investigated through outage probability and asymptotic outage probability expressions. For a delay-limited transmission mode, throughput of the system is presented.

6.4.1 Mathematical Modeling

To model the feedback errors, consider a scenario where the CSI received at the transmitter is outdated due to the time-varying nature of the channel when fed back from the receiver. Thus, the resultant CSI at the transmitter is outdated and results non-zero feedback link delay. Let's consider $\hat{\mathbf{h}}^{u}_{PQ}$ an estimated channel vector to be

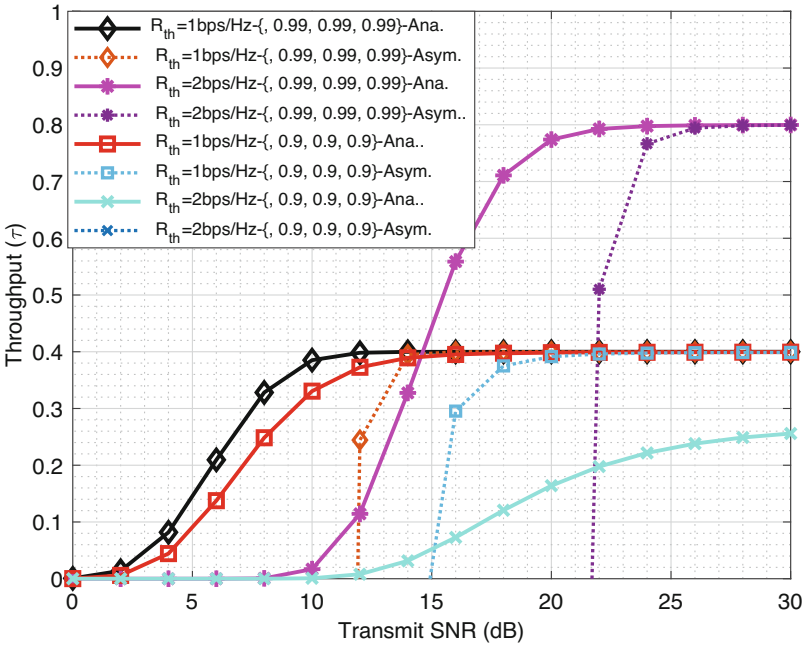


Fig. 6.5 $2 \times 2 \times 1$ MIMO AC: Throughput vs SNR in the presence of imperfect CSI conditions [13]

the delayed channel of $\mathbf{h}_{PQ}^{u}$ and which is used in antenna selection. For decoding, consider $\mathbf{h}_{PQ}^{u}$ be the estimated channel vector used. Hence, $\hat{\mathbf{h}}_{PQ}^{u}$ is conditioned over $\mathbf{h}_{PQ}^{u}$ following a Gaussian distribution. Thus, the relation between $\hat{\mathbf{h}}_{PQ}^{u}$ and $\mathbf{h}_{PQ}^{u}$ is given as [36]

$$\hat{\mathbf{h}}_{PQ}^{u} = \rho_{PQ}\mathbf{h}_{PQ}^{u} + \sqrt{1-\rho_{PQ}^{2}}\mathbf{e}, \quad 0 \leq \rho_{PQ} \leq 1 \tag{6.20}$$

where ρ_{PQ} is the correlation coefficient between $\hat{\mathbf{h}}_{PQ}^{u}$ and $\mathbf{h}_{PQ}^{u}$. The error is considered to be modeled as $\mathbf{e} \sim \mathcal{N}(0, I)$ [36].

With Doppler frequency (f_d) and delay spread (T_d) of the feedback channels, the channel correlation coefficient for Clarke's fading spectrum is modeled as $\rho_{PQ} = \mathcal{J}_o(2\pi f_d T_d)$. Further, at R energy is harvested through TS protocol from the S received signals. Energy harvesting and the IP take place as discussed in the earlier section. In the first hop, S transmits the signal with transmit power P_S to R and D, respectively. The signals received at multiple antennas at R and D MRC receiver are used, and thus the received signals at R and D are given as

$$\mathbf{y}_{SR} = \sqrt{P_S}\mathbf{h}_{SR}^{u}x + \mathbf{w}_{SR}, \tag{6.21}$$

$$\mathbf{y}_{SD} = \sqrt{P_S}\mathbf{h}_{SD}^{u} x + \mathbf{w}_{SD}, \tag{6.22}$$

where $\mathbf{w}_{SR}$ and $\mathbf{w}_{SD}$ are the AWGN vectors corresponding to the SR and SD links, respectively. In the second hop, R broadcasts the received S signal to D after amplifying with a gain by TAS through υth antenna [37]

$$G \leq \sqrt{\frac{P_R}{P_S||\mathbf{h}_{SR}^{u}||^4 + \sigma_{SR}^2||\mathbf{h}_{SR}^{u}||^2}} \approx \sqrt{\frac{K_1}{||\mathbf{h}_{SR}^{u}||^2}}, \tag{6.23}$$

and is given as

$$\mathbf{y}_{RD} = \sqrt{P_R} G(\mathbf{h}_{RD}^{v})(\mathbf{h}_{SR}^{u})^H \mathbf{y}_{SR} + \mathbf{w}_{RD}, \tag{6.24}$$

where P_R is transmit power at R and $\mathbf{w}_{RD}$ is the RD link AWGN vector. P_R is drawn from the EH phase as [11]

$$E_h = \eta\alpha T P_S||\mathbf{h}_{SR}^{u}||^2, \tag{6.25}$$

$$P_R = \frac{2E_h}{(1-\alpha)T} = \frac{2\eta\alpha P_S||\mathbf{h}_{SR}^{u}||^2}{(1-\alpha)}. \tag{6.26}$$

Signals received from both the hops at D are combined using MRC with a MMSE filter [32, 38]. Thus, the e2e SNR at D is given as

$$\Lambda_{e2e}^{(u,v)} = \Lambda_{SD}^{(u)} + \Lambda_{SRD}^{(u,v)}, \tag{6.27}$$

where,

$$\Lambda_{SRD}^{(u,v)} = \frac{\frac{P_S}{\sigma_N^2} K_1||\mathbf{h}_{RD}^{v}||^2 \frac{P_S}{\sigma_N^2}||\mathbf{h}_{SR}^{u}||^2}{\left(\frac{P_S}{\sigma_N^2} K_1||\mathbf{h}_{RD}^{v}||^2 + \frac{P_S}{\sigma_N^2}\right)} = \frac{\Lambda_{SR}^{(u)}\Lambda_{RD}^{(v)}}{1 + \Lambda_{RD}^{(v)}}, \tag{6.28}$$

where $\Lambda_{SD}^{(u)} = \overline{\Lambda}_{SD}||\mathbf{h}_{SD}^{u}||^2$, $\Lambda_{SR}^{(u)} = ||\mathbf{h}_{SR}^{u}||^2$, $\Lambda_{RD}^{(v)} = \Lambda_{K_1}||\mathbf{h}_{RD}^{v}||^2$, and $\Lambda_{K_1} = \frac{P_S}{\sigma_N^2} K_1$.

6.4.2 System Performance Metrics

6.4.2.1 Outage Probability

As the outage probability determines the reliability of the communication link, it can be analyzed through the closed-form UB expression and is given as [14]

$$\mathrm{P}_{out}^{(UB)}(\Lambda_{th}) = F_{\hat{\Lambda}_{SD}^{(U)}}(\Lambda_{th}) F_{\hat{\Lambda}_{SRD}^{(U,V)}}(\Lambda_{th}), \tag{6.29}$$

where

$$\begin{aligned}
F_{\hat{\Lambda}_{SD}^{(u)}}(\Lambda_{th}) &= 1 + \sum_{e=0}^{N_S-1} \sum_{f=0}^{e(M_{SD}-1)} \sum_{g=0}^{f} \sum_{h=0}^{M_{SD}+g-1} \binom{N_S-1}{e} \binom{f}{g} \frac{(-1)^{e+1} N_S \Gamma(M_{SD}+f)}{h! \Gamma M_{SD}} \\
&\quad \times \frac{\Phi_{n,m,M_{SD}}}{(e+1)^{M_{SD}+g-h}} \frac{\rho_{SD}^{g}(1-\rho_{SD})^{f-g} \Lambda_{th}^{h}}{\lambda_{SD}^{h}[1+m(1-\rho_{SD})]^{f+h}} e^{-\Delta_1 \Lambda_{th}}, \\
F_{\hat{\Lambda}_{SRD}^{(u,v)}}(\Lambda_{th}) &= 1 + \sum_{a=0}^{N_S-1} \sum_{b=0}^{a(M_{SR}-1)} \sum_{j=0}^{b} \sum_{i=0}^{M_{SR}+j-1} \sum_{p=0}^{N_R-1} \sum_{q=0}^{p(M_{RD}-1)} \sum_{r=0}^{q} \sum_{s=0}^{\Delta_4} \frac{(-1)^{a+p+1} 2 N_S N_R}{i!(a+1)^{M_{SR}+j-i}} \\
&\quad \times \frac{\Gamma(M_{SR}+b)}{\Gamma(M_{RD}+r)} \frac{\Gamma(M_{RD}+q)}{\Gamma M_{SR} \Gamma M_{RD}} \binom{N_S-1}{a} \frac{\rho_{SR}^{j}(1-\rho_{SR})^{b-j} \Phi_{b,a,M_{SR}}}{\lambda_{SR}^{i}[1+a(1-\rho_{SR})]^{b+j}} \binom{N_R-1}{p} \\
&\quad \times \frac{\rho_{RD}^{r}(1-\rho_{RD})^{q-r} \Phi_{q,p,M_{RD}}}{\lambda_{RD}^{M_{RD}+r}[1+p(1-\rho_{RD})]^{q+s}} \binom{b}{j} \binom{q}{r} \left(\frac{(a+1)\lambda_{RD}[1+p(1-\rho_{RD})]}{(p+1)\lambda_{SR}[1+a(1-\rho_{SR})]} \right)^{\frac{\vartheta}{2}} \\
&\quad \times \binom{\Delta_4}{s} \Lambda_{th}^{\Delta_4+1} e^{-(\Delta_5)\Lambda_{th}} K_{\vartheta}\left(2\Lambda_{th}\sqrt{\Delta_6}\right).
\end{aligned} \tag{6.30}$$

Further, $\Delta_1 = \frac{e+1}{\lambda_{SD}[1+e(1-\rho_{SD})]}$, $\Delta_2 = \frac{a+1}{\lambda_{SR}[1+a(1-\rho_{SR})]}$, $\Delta_3 = \frac{p+1}{\lambda_{RD}[1+p(1-\rho_{RD})]}$, $\Delta_5 = \Delta_2 + \Delta_3$, and $\Delta_6 = \Delta_2 \Delta_3$. Furthermore, $\Delta_4 = M_{RD}+r+i-1$, $\vartheta = s-i+1$, $M_{PQ} = m_{PQ} N_Q$, and $\lambda_{PQ} = \frac{\Lambda_{PQ}}{m_{PQ}}$.

6.4.2.2 Asymptotic Analysis

As discussed earlier, asymptotic analysis provides the system design insights such as diversity gain and coding gain. Diversity gain provides the number of independent paths between the transmitter and the receiver. Asymptotic outage probability expression is derived at $\overline{\Lambda} \to \infty$ as [14]

$$\mathrm{P}_{out}^{\infty}(\Lambda_{th}) = \begin{cases} f_{SD} f_{SR} \left(\frac{\Lambda_{th}}{\overline{\Lambda}}\right)^{d_1}, M_{SR} < M_{RD} \\ f_{SD} f_{RD} \left(\frac{\Lambda_{th}}{\overline{\Lambda}}\right)^{d_2}, M_{SR} > M_{RD} \\ f_{SD}(f_{SR} + f_{RD}) \left(\frac{\Lambda_{th}}{\overline{\Lambda}}\right)^{d_3}, \quad M \end{cases} \tag{6.31}$$

$$f_{SR} = \sum_{a=0}^{N_S-1} \sum_{b=0}^{a(M_{SR}-1)} \binom{N_S-1}{a} \frac{\Gamma(M_{SR}+b)}{M_{SR}!\Gamma M_{SR}} \frac{(-1)^a \phi_{b,a,M_{SR}} (1-\rho_{SR})^b}{[1+a(1-\rho_{SR})]^{M_{SR}+b}} \left(\frac{m_{SR}}{k_{SR}}\right)^{M_{SR}}, \tag{6.32}$$

$$f_{RD} = \sum_{p=0}^{N_R-1} \sum_{p=0}^{q(M_{RD}-1)} \binom{N_R-1}{p} \frac{\Gamma(M_{RD}+q)}{M_{RD}!\Gamma M_{RD}} \frac{(-1)^p \phi_{q,p,M_{RD}} (1-\rho_{RD})^q}{[1+p(1-\rho_{RD})]^{M_{RD}+q}} \left(\frac{m_{RD}}{k_{RD}}\right)^{M_{RD}}, \tag{6.33}$$

$$f_{SD} = \sum_{e=0}^{N_S-1} \sum_{f=0}^{e(M_{SD}-1)} \binom{N_S-1}{e} \frac{\Gamma(M_{SD}+f)}{M_{SD}!\Gamma M_{SD}} \frac{(-1)^e \phi_{f,e,M_{SD}} (1-\rho_{SD})^f}{[1+e(1-\rho_{SD})]^{M_{SD}+f}} \left(\frac{m_{RD}}{k_{SD}}\right)^{M_{SD}}, \tag{6.34}$$

where diversity order is given by d_1, d_2, and d_3. Further, $M = M_{SR} = M_{RD}$, $d_1 = M_{SD} + M_{SR}$, $d_2 = M_{SD} + M_{RD}$, and $d_3 = M_{SD} + M$. Furthermore, $k_{SD} = \frac{\overline{\Lambda}_{SD}}{\overline{\Lambda}}$, $k_{SR} = \frac{\overline{\Lambda}_{SR}}{\overline{\Lambda}}$, and $k_{RD} = \frac{\overline{\Lambda}_{RD}}{\overline{\Lambda}}$.

6.4.2.3 Throughput Analysis

We consider a delay-limited transmission mode for the analysis of throughput. Thus, in this mode throughput is evaluated for a fixed transmission rate $R_r = \log_2(1 + \Lambda_{e2e})$ bits/s/Hz from the outage probability as

$$\tau = \frac{(1 - \mathrm{P}_{out}) R_r (1-\alpha)}{2}, \tag{6.35}$$

6.4.3 Results and Discussions

In this section, we discuss the derived analytical expressions from the various system design aspects and validate them through the Monte Carlo simulations. The system modeling parameters are considered as follows: correlation parameters (CPs) as $\{\rho_{SR}, \rho_{RD}, \rho_{SD}\}$, transmission rate $R_r = 3$ bits/s/Hz, and EH efficiency factor ($\eta = 1$). In the analysis, urban macro-cell network environment is considered with path loss exponent 4 [39].

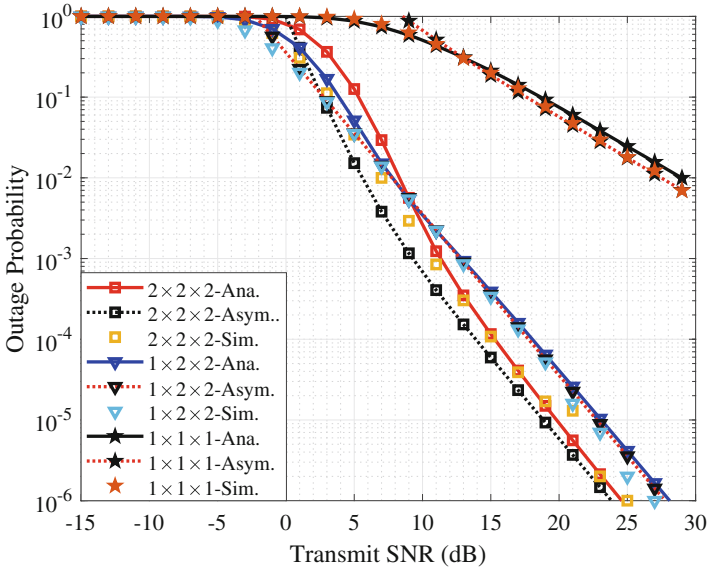


Fig. 6.6 Outage probability vs SNR for different MIMO ACs [14]

6.4.3.1 Impact of MIMO Antenna System

In Fig. 6.6, outage probability and SNR curves are presented for $2 \times 2 \times 2$ AC with FPs, $\{1, 1, 1\}$, $1 \times 2 \times 2$ AC with FPs $\{2, 1, 1\}$ and $1 \times 1 \times 1$ AC with FPs $\{1, 1, 1\}$. In the analysis, EH is performed for $\alpha = 0.2T$, and CPs are taken as $\{0.9, 0.9, 0.9\}$. The derived analytical UB results match well with the asymptotic results, and Monte Carlo results validate them. Curves illustrate that for an outage probability of 10^{-5}, $2 \times 2 \times 2$ AC performs better than $1 \times 2 \times 2$ AC with a SNR gain of ≈ 3 dB. For an outage of 10^{-2}, $1 \times 2 \times 2$ and $2 \times 2 \times 2$ ACs have a SNR gain of ≈ 20 dB over $\{1, 1, 1\}$ AC. This illustrates the improvement of system performance with MIMO antennas.

6.4.3.2 Impact of Feedback Delays

In Fig. 6.7, the impact of feedback delays is visualized with respect to (α) over the outage probability at 15 dB SNR for a $2 \times 2 \times 2$ AC with FPs $\{1, 1, 1\}$. Curves demonstrate the severity of feedback delays over the outage probability with respect to α. For $\alpha < 0.4T$, system records lower outage probability with better performance. However, with a decrease in ρ_{PQ}, outage probability increases and results in system performance degradation. Curves illustrate for $\alpha < 0.1$, system with $\{0.1, 0.1, 0.1\}$ CP records high outage probability over the other CPs values. Further, system with $\{0.9, 0.9, 0.9\}$ CP has low outage probability for $\alpha < 0.5T$. It

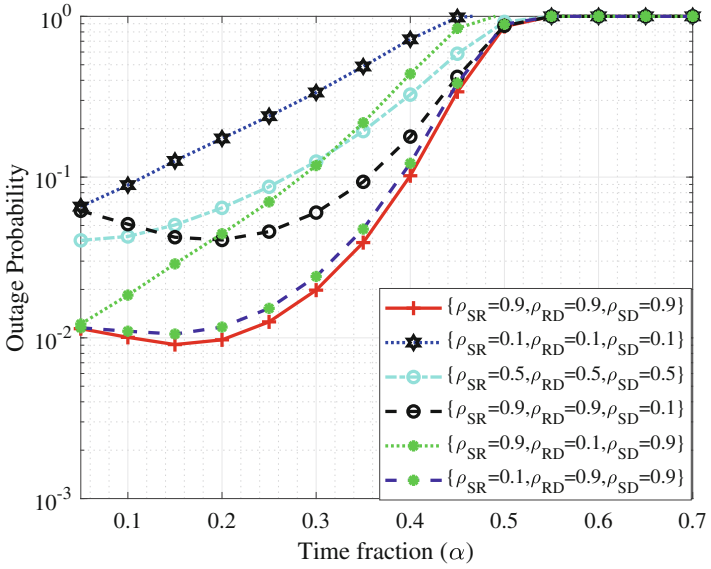


Fig. 6.7 $2 \times 2 \times 2$ MIMO AC: Outage probability vs (α) for different ρ values [14]

is noticed that system performance improves with the decrease in αT due to more IP time.

6.4.3.3 Throughput Analysis

In Fig. 6.8, throughput vs SNR curves are plotted for a $1 \times 2 \times 2$ AC with FPs $\{2, 1, 1\}$. Throughput is analyzed at two fixed transmission rates $R_r = 1$ and $R_r = 2$ bits/s/Hz and with $\alpha = 0.2T$. Curves illustrate the increase in system throughput with the increase in R_r at high SNR. System with $\{0.9, 0.9, 0.9\}$ CP with an SNR gain of $\approx$ 5 dB attains higher throughput prior to $\{0.1, 0.1, 0.1\}$ CP. For $R_r = 2$ bits/s/Hz, system with $\{0.9, 0.9, 0.9\}$ and $\{0.1, 0.1, 0.1\}$ CPs, throughput saturates to 0.8 at 38 dB and 43 dB, respectively, whereas for $R_r = 1$ bits/s/Hz, system throughput saturates to 0.4 at 10 dB and 15 dB, respectively.

6.5 Conclusion

In this chapter, concepts of energy harvesting are discussed such as energy harvesting sources, energy transmission techniques, energy harvesting models, and the impact of channel estimation errors over SWIPT devices. Among the available energy harvesting sources, radio frequency signal-based EH finds a viable solution

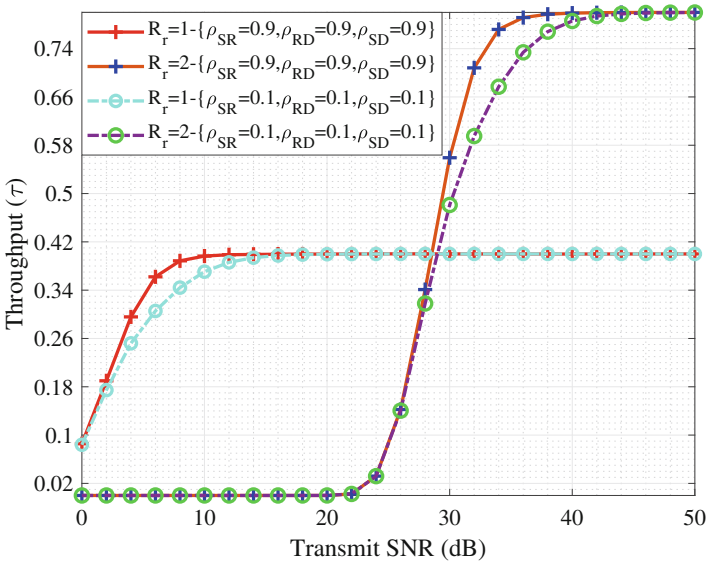


Fig. 6.8 $1 \times 2 \times 2$ MIMO AC: Throughput vs SNR [14]

for energy-constraint nodes. The state-of-the-art EH with cooperative D2D MIMO relaying in the presence of channel estimation errors including feedback delays and imperfect CSI is addressed. System performance is quantified through system metrics such as outage probability and asymptotic outage probability. Further, actual data delivery rate is analyzed through the throughput analysis for a delay-limited transmission, and useful insights are drawn. Impact of MIMO antennas, fading parameters severity, severity of the feedback error, imperfect CSI, and energy harvesting time constant over the system performance are also analyzed.

6.5.1 Summary

A brief summary is as follows:

- Impact of imperfect CSI and feedback delays over MIMO D2D system with SWIPT is analyzed.
- System performance is quantified through the outage probability and asymptotic outage probability closed-form expressions.
- Actual delivered data rate is analyzed through throughput analysis.
- Results illustrate that the system performance degrades with the increase in α values. However, for smaller values of α, system performs better in both perfect and imperfect CSI scenarios.
- Results illustrate that for $\alpha < 0.4T$, system attains low outage probability.

- At high rates, curves demonstrate that the system performance degrades under imperfect CSI conditions.
- Increase in outdated CSI deteriorates the system performance drastically.
- Diversity order of the system is lost with a slight increase in CEE, due to SNR independent CEE variance.
- Impacts of multiple antennas and fading parameters severity along with imperfect CSI are analyzed along with SWIPT.

References

1. G. Forecast, Cisco visual networking index: global mobile data traffic forecast update, 2017–2022. Update **2017**, 2022 (2019)
2. P. K. Singya, P. Shaik, N. Kumar, V. Bhatia, and M.-S. Alouini, A Survey on Higher-Order QAM Constellations: Technical Challenges, Recent Advances, and Future Trends. IEEE Open J. Commun. Soc. **2**, 617–655 (2021)
3. M. Series, IMT vision–framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU **2083**, (2015)
4. A. Gupta, R.K. Jha, A survey of 5G network: Architecture and emerging technologies. IEEE Access **3**, 1206–1232 (2015)
5. Y. Chen, *Energy Harvesting Communications: Principles and Theories* (Wiley, Hoboken, 2019)
6. V. Leonov, Thermoelectric energy harvesting of human body heat for wearable sensors. IEEE Sensors J. **13**(6), 2284–2291 (2013)
7. P. Grover, A. Sahai, Shannon meets Tesla: Wireless information and power transfer, in *IEEE International Symposium on Information Theory* (IEEE, Piscataway, 2010), pp. 2363–2367
8. H. Ju, R. Zhang, Throughput maximization in wireless powered communication networks. IEEE Trans. Wireless Commun. **13**(1), 418–428 (2013)
9. S. Bi, C.K. Ho, R. Zhang, Wireless powered communication: opportunities and challenges. IEEE Commun. Mag. **53**(4), 117–125 (2015)
10. L.R. Varshney, Transporting information and energy simultaneously, in *2008 IEEE International Symposium on Information Theory* (2008), pp. 1612–1616
11. A.A. Nasir, X. Zhou, S. Durrani, R.A. Kennedy, Relaying protocols for wireless energy harvesting and information processing. IEEE Trans. Wirel. Commun. **12**(7), 3622–3636 (2013)
12. T. Li, P. Fan, K.B. Letaief, Outage probability of energy harvesting relay-aided cooperative networks over Rayleigh fading channel. IEEE Trans. Veh. Technol. **65**(2), 972–978 (2015)
13. P. Shaik, P.K. Singya, N. Kumar, K.K. Garg, V. Bhatia, On impact of imperfect CSI over SWIPT device-to-device (D2D) MIMO relay systems, in *SPCOM* (IEEE, Piscataway, 2020), pp. 1–5
14. P. Shaik, P.K. Singya, K.K. Garg, V. Bhatia, Outage Probability analysis of SWIPT device-to-device MIMO relay systems with outdated CSI, in *IEEE Wireless Communications and Networking Conference (WCNC)* (IEEE, Piscataway, 2021), pp. 1–6
15. S. Parvez, P.K. Singya, V. Bhatia, On ASER analysis of energy efficient modulation schemes for a device-to-device MIMO relay network. IEEE Access **8**, 2499–2512 (2019)
16. P. Shaik, P.K. Singya, V. Bhatia, On impact of imperfect CSI over hexagonal QAM for TAS/MRC-MIMO cooperative relay network. IEEE Commun. Lett. **23**(10), 1721–1724 (2019)
17. T.M. Hoang, X.N. Tran, N. Thanh, et al., Performance analysis of MIMO SWIPT relay network with imperfect CSI. Mobile Netw. Appl. **24**(2), 630–642 (2019)
18. P. Shaik, K.K. Garg, V. Bhatia, On impact of imperfect channel state information on dual-hop nonline-of-sight ultraviolet communication over turbulent channel. Opt. Eng. **59**(1), 1–14 (2020)

19. Y. Alsaba, S.K.A. Rahim, C.Y. Leow, Beamforming in wireless energy harvesting communications systems: a survey. IEEE Commun. Surv. Tut. **20**(2), 1329–1360 (2018)
20. S. Bisen, P. Shaik, V. Bhatia, On performance of energy harvested cooperative NOMA under imperfect CSI and imperfect SIC. IEEE Trans. Veh. Technol. **70**(9), 8993–9005 (2021)
21. A.S. Parihar, P. Swami, V. Bhatia, Z. Ding, Performance analysis of SWIPT enabled cooperative-NOMA in heterogeneous networks using carrier sensing. IEEE Trans. Veh. Technol. **70**, 1–1 (2021)
22. D. Wang, F. Rezaei, C. Tellambura, Performance analysis and resource allocations for a WPCN with a new nonlinear energy Harvester model. IEEE Open J. Commun. Soc. **1**, 1403–1424 (2020)
23. Y. Dong, M.J. Hossain, J. Cheng, Performance of wireless powered amplify and forward relaying over Nakagami-*m* fading channels with nonlinear energy Harvester. IEEE Commun. Lett. **20**(4), 672–675 (2016)
24. Y. Chen, K.T. Sabnis, R.A. Abd-Alhameed, New formula for conversion efficiency of rf eh and its wireless applications. IEEE Trans. Veh. Technol. **65**(11), 9410–9414 (2016)
25. B. Clerckx, E. Bayguzina, Waveform design for wireless power transfer. IEEE Trans. Signal Proces. **64**(23), 6313–6328 (2016)
26. E. Boshkovska, D.W.K. Ng, N. Zlatanov, R. Schober, Practical non-linear energy harvesting model and resource allocation for SWIPT systems. IEEE Commun. Lett. **19**(12), 2082–2085 (2015)
27. M.K. Ozdemir, H. Arslan, Channel estimation for wireless ofdm systems. IEEE Commun. Surv. Tut. **9**(2), 18–48 (2007)
28. M. Seyfi, S. Muhaidat, J. Liang, Amplify-and-forward selection cooperation over Rayleigh fading channels with imperfect CSI. IEEE Trans. Wirel. Commun. **11**(1), 199–209 (2011)
29. K.K. Garg, P.K. Singya, V. Bhatia, Performance analysis of NLOS ultraviolet communications with correlated branches over turbulent channels. IEEE/OSA J. Opt. Commun. Net. **11**(11), 525–535 (2019)
30. K. Garg, P. Shaik, V. Bhatia, O. Krejcar, On the performance of relay assisted hybrid RF-NLOS UVC system with imperfect channel estimation. J. Optical Commun. Netw. (2021). [Online]. Available: https://doi.org/10.1364%2Fjocn.440819
31. Y. Liu, L. Wang, M. Elkashlan, T.Q. Duong, A. Nallanathan, Two-way relay networks with wireless power transfer: design and performance analysis. IET Commun. **10**(14), 1810–1819 (2016)
32. P. Shaik, P.K. Singya, V. Bhatia, Performance analysis of QAM schemes for non-regenerative cooperative MIMO network with transmit antenna selection. AEU-Int. J. Electron. Commun. **107**, 298–306 (2019)
33. J. Jose, P. Shaik, V. Bhatia, VFD-NOMA under imperfect SIC and residual inter-relay interference over generalized Nakagami-m fading channels. IEEE Commun. Lett. **25**, 646–650 (2020)
34. K.K. Garg, P. Shaik, V. Bhatia, Performance analysis of cooperative relaying technique for non-line-of-sight UV communication system in the presence of turbulence. Opt. Eng. **59**(5), 055101 (2020)
35. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 2014)
36. N. Yang, M. Elkashlan, P.L. Yeoh, J. Yuan, Multiuser MIMO relay networks in Nakagami-m fading channels. IEEE Trans. Commun. **60**(11), 3298–3310 (2012)
37. S. Parvez, D. Kumar, V. Bhatia, On performance of SWIPT enabled two-way relay system with non-linear power amplifier, in *NCC* (IEEE, Piscataway, 2020), pp. 1–6
38. V. Bhatia, B. Mulgrew, Non-parametric likelihood based channel estimator for Gaussian mixture noise. Signal Process. **87**(11), 2569–2586 (2007)
39. A. Goldsmith, *Wireless Communications* (Cambridge University Press, Cambridge, 2005)

Chapter 7
Vehicular Communications in the B5G Era

Leyre Azpilicueta, Cesar Vargas-Rosales, Ana Vazquez Alejos, and Francisco Falcone

7.1 Introduction

The transition from 4G to 5G did not bring disruptive changes to the physical layer or radio interface. Up to 2020, the novelties consisted of higher throughput, latency, and reliability, due to the use of OFDM and the introduction of network slicing, among other functionalities. Furthermore, base stations have incorporated antenna arrays for massive MIMO and larger bandwidths that increased notably the spectral efficiency. The successful development of the fifth generation (5G) of wireless mobile communications requires disruptive technologies to fulfill unforeseen data rates, latency, and capacity that will be necessary, among others, including network densification and virtualization, massive multiple input multiple output (MIMO), antenna beamforming, device-to-device communications, smart cells, multi-cell cooperation, and interference mitigation techniques. In this scenario, the radio interface acquires a new dimension, and its proper modeling will contribute ensuring the expected 5G quality of service and quality of experience requirements. 5G pushes first-line questions like radio spectrum, radio access, and core network that incorporate slicing and virtualization concepts. Thus, 5G represents the most striving of the technology challenges for a generation of wireless mobile communi-

L. Azpilicueta (✉) · C. Vargas-Rosales
School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL., Mexico
e-mail: leyre.azpilicueta@tec.mx; cvargas@tec.mx

A. V. Alejos
Campus Universitario, Escola de Ingeniería de Telecomunicación, Universidad de Vigo, Vigo, Spain
e-mail: analejos@uvigo.es

F. Falcone
EEC Department, Universidad Publica de Navarra, Edificio Los Tejos, Pamplona, Navarra, Spain
e-mail: francisco.falcone@unavarra.es

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_7

cations up to date. Nonetheless, at the early phase, the network deployments do not implement all the capabilities defined for 5G New Radio (5G NR) network. Network optimization and standardization is an uncompleted process, in which researchers have continued to work in parallel to identify enabling factors and promising topics beyond the present 5G situation that will unavoidably lead to the next generation of wireless communication system. This phase of research stage conforms beyond 5G (B5G), and the efforts will lay the definition of the sixth generation (6G). 6G will likely combine trends derived from 5G and B5G research phases. Several white papers [1–8] and consortiums [9–11] describe basically the same idea as presented here.

The evolution of 5G/B5G/6G accepts the following timelines or roadmap [12]:

- 5G phase 1: up to 2020 with the spectrum allocated below 6 GHz in 2015 World Radiocommunication Conference (WRC-15), [13]. It is usually referred to as simply 5G.
- 5G phase 2: likely extending to 2030 with the spectrum above 6 GHz recently allocated in 2019 World Radiocommunication Conference (WRC-19) [14], usually referred as B5G.
- Post 5G phase 2: beyond 2030, and likely laying 6G if standardization is reached.

B5G and 6G networks may then be defined as the next generation of mobile communication systems with characteristic features that will contribute to the Sustainable Development Goals (SDG), such as ultralow power consumption, ultra-security and reliability, autonomy, and scalability, jointly to optimize 5G performance in three specific service cases: enhanced mobile broadband (eMBB) for ultrafast speed and capacity (up to 10 Gbps over the air transmission rate), ultra-reliable low latency communications (URLLC) for large demanding applications such as vehicle-to-everything (with 1ms latency goal), and massive machine-type communications (mMTC) for multiple simultaneous connections. Even though most of these features are expected for 5G networks, B5G-specific network softwarization will make a difference [12], given by extended use of artificial intelligence and machine learning in order to enhance radio resource allocation, evolved core network functionalities, or data processing/handling and multi-agent integration, towards cognitive environments. The envisioned functional architecture of B5G/6G networks, where frequency, space and time are essential re-sources to be managed to make effective use of them [1, 2].

Worldwide, different spectrum bands have been released for 5G, with the main focus for network deployments in FR1 (i.e., under 6 GHz). However, higher-frequency spectrum, particularly 26 GHz [15, 16], is now available, and also the millimeter wave spectrum has recently become a 5G band. Nonetheless, despite the exploitation of higher frequencies and millimeter wave spectrum having been promoted to attain over gigabit data rates, the business model relying on expanding spectrum allocation is not valid for rural areas in terms of cost-effectiveness. Therefore, spectrum sharing of traditional and congested frequency bands as well as the inclusion of satellite communications will appear as solutions to avoid holding

back those communities of a new era of connectivity [17]. In a word: 5G should work for all.

Terrestrial 5G solutions under development will not follow an evolution track to allow connecting rural areas in a short horizon [12]. Basically, 5G can be primarily defined as an urban system. However, not few arguments will impel for improved B5G rural coverage [12]: urbanization, improve public safety, e-health, connected cars, and entertainment services. Before long, regulations with respect to rural coverage will arise and will be compulsory to fulfill [12].

The saturation of the bands below 6 GHz along with the growing demand of wireless channel capacity for new communication is pushing towards the increase on the operating frequency of the 5G/B5G/6G devices. One of the main drawbacks when the frequency increases is the higher resulting propagation losses, as well as an increase in atmospheric absorption, as well as in relation with shadowing effects related with frequency dispersive material characteristics [18].

In 2014, the third-generation partnership program (3GPP) working groups considered incorporating vehicle-to-everything (V2X, communication links related in general with the vehicular environment, enabling widespread connectivity) as an additional feature for LTE Advanced, [19–22, 22, 23]. 3GPP's Technical Specification Group on Service and System Aspects (TSG-SA) within LTE Release 15 explored the 5G-grade enhanced V2X (eV2X) use cases [19, 20, 24, 25]. The envisioned scenarios (platooning, remote driving, smart car sensory...) will require a combination of very high data rates, large reliability levels, and low latencies [19, 20, 24–26]. 3GPP introduced LTE-based vehicle-to-infrastructure communications (V2I, which enables wireless links between vehicles and road infrastructure, for enhanced and secure transportation as well as for non-critical user-oriented applications) and device-to-device D2D communication for the V2V communication [19–21]. D2D communication is considered as essential for the case where the cellular network coverage is missing or not guaranteed [19–21].

However, none of those technologies is sufficient to accomplish the stringent requirements demanded by the eV2X or LTE V2X applications. For this reason, 3GPP started developing in Release 19 [19, 20, 22], a standard that will lead to the third wave of V2X communications that will complement LTE by including the 5G New Radio (NR) technologies (NR-V2X) with special attention to the deployment in millimeter wave bands for vehicular communications [19–21, 24, 27]. 5G is intended to support higher mobility, and to improve network reliability, with a packet loss rate of around 10-5 valid to offer safety critical services [19–21].

In WRC-19 it was determined that 5G will operate, among other options, in millimeter wave (mm-wave) spectrum ranging from 24 to 300 GHz [14]. For mm-wave-based V2X communications, some interesting frequency portions have been postulated: 60 GHz, 79 GHz, and low-terahertz frequencies at 300 GHz and beyond [19, 20, 24, 28]. Standardization of NR-V2X will expand LTE V2X with massive deployments of connected vehicles [19, 20, 24].

At present, the mm-wave band is considered for vehicle-to-vehicle (V2V, which enables direct connectivity among vehicles, mainly related with critical driving context information exchange) or V2I communications, required for autonomous

Table 7.1 5G satellite network use cases and scenarios by 3GPP [31, 32]

Case	Description
1	Satellite and terrestrial network roaming capabilities, in order to provide global coverage, in the case terrestrial-based network coverage fails
2	Satellite overlay for multicast and broadcast, in order to provide delivery of global video services
3	Delivery of IoT services with global coverage. Use of low-orbit satellites for efficient radio connections
4	Satellite use in temporary way for different services (e.g., mail, messaging, or voice)
5	Optimized routing via satellite links to provide enhanced connectivity, via independent UL/DL links within the 5G terrestrial network and the satellite network
6	Multiregional/multinational satellite service provision, enabling radio access connectivity, including neutral regions
7	Capability to provide overlay services of low-orbit satellites to terrestrial links
8	Possibility of non-direct connectivity to the satellite network, aided by relay units
9	Transport connectivity (Backhaul/Midhaul) between radio stations and the core network
10	Backhaul provision for services on the move (e.g., trains, ships, etc.) to the core network from moving radio stations
11	Interoperability with other non PLMN technologies, such as WLAN
12	Enabling connectivity in remote locations and applications aided by LEO satellites, such as offshore wind farm to remote data centers

or proxy driving in future smart cars [19, 20, 22]. The V2X roadmap defined by the 5G Public Private Partnership (5G-PPP) envisioned scenarios that include awareness driving, cooperative driving, and synchronized driving with vehicles being autonomously driven [19, 20, 24, 25, 29, 30]. Some of those scenarios are integrated with satellite networks as shown in Table 7.1.

In [19], Table 6.9 shows the survey summarizing existing research reports regarding the mm-wave V2X communications.

7.2 V2X Communications

Within wireless communications and networking, several paradigms are being developed. One of those is the vehicular communications paradigm, which can be seen from a networking perspective or from a physical layer (PHY) perspective. A networking perspective presents examples such as vehicular ad hoc networks (VANETs), internet of vehicles (IoV), and vehicular sensor networks, among others, where connectivity and the network architecture are developed. The integration of these networking examples with current and future networks and technologies is of relevance since information needs to be carried by others such as 4G, 5G, beyond 5G (B5G), cognitive radio, IEEE802.11p/WAVE/C-ITS, etc. The PHY perspective provides a base on which new technologies can be developed and built, since one

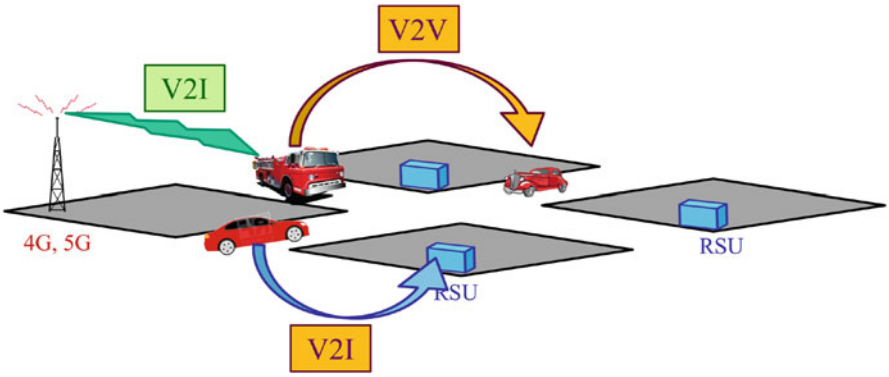


Fig. 7.1 V2X communication

can study the vehicular communications from the propagation of signals, which establishes a platform to develop analysis for coverage and expected performance of each of the links that vehicles communicate.

In vehicular communications, links that need to be established are broadly classified into vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V), although in some cases this is in general called vehicle-to-everything (V2X) communications; see Fig. 7.1 for an example.

When solutions are proposed for the V2X communications, it is important to consider some of the differences that would be involved if the solution is for V2I or V2V. In case of V2I communications, we have fixed infrastructure elements generally called roadside units (RSU) that are located along the roads and streets and work similarly to an access point in a wireless local area network. Vehicles try to communicate with the infrastructure, and some of the problems in such a link will be caused by issues such as propagation, mobility, environment, and obstacles that cause channel impairments such as losses, Doppler, multipath, fadings, and distortion. In some cases, handoffs need to be included in order to provide constant communication as the vehicle moves along the road. In contrast, for V2V communications, we have the same kind of channel impairments, in some cases in critical levels, for example, when communicating vehicles move in opposite directions. Also, depending on the installation of RSUs, there could be a direct connection between vehicle and infrastructure, whereas for the V2V channel, we could have more obstacles obstructing the line of sight (LOS) between vehicles.

For V2X (V2I and V2V), channel parameters such as coherence time and coherence bandwidth also play important roles due to the mobility. It is also important to have measurement campaigns in order to collect data that help to establish the behavior of the signals in such environments. The characterization of all possible scenarios is critical to understand the effect of the channel in the communication links.

In the automotive industry, there are two issues that are paving the road for the future. The first is electromobility that provides sustainable vehicles with low to zero emissions. The hybrid or electric vehicle will have many different systems integrated that need from sensors to monitor the environment. Intravehicular communication can be done in many different ways, and the CAN bus has been part for some years now. The second is mobility with connectivity, the provisioning of connected services to the driver and passengers of the vehicle. This needs more of the networks to connect with the outside. Information needs to be carried in some of the different technologies for that purpose, i.e., 4G, 5G, IEEE802.11p, etc. Both issues need reliable communications in order to fulfill their needs, e.g., electromobility needs from data bases opportune information for routing, for traffic assessment, and in general for transportation needs. Connected services need to provide entertainment, monitoring of vehicle, insurance needs, remote services, network connectivity for the passengers, etc.

7.3 Satellite Communications

The standardization of 5G has reached its latest milestone with the 3GPP Release 17, due out in early 2022. The 3GPP Release 17 will provide the phase 2 of architecture enhancements for 3GPP support of advanced vehicle-to-everything (V2X) services. 3GPP RAN will set the normative to support non-terrestrial access (NTN) via satellites and high-altitude platform systems (HAPS), to address the role of satellite segment in the 5G ecosystem vision.

Generally speaking, 5G satellite networks are those networks provided of the NG-RAN radio access network based on a satellite network technology [31]. In 2017, European Space Agency (ESA) launched the project "Satellite for 5G" aimed to study to introduce 5G satellite-based access components [31, 33]. In 2018, the CEPT completed the ECC Report 280 [34, 35] for determining the role of satellite component within 5G and stating that some services cannot be carried out bypassing satellites [31, 34].

Technical specifications of 3GPP [31, 32] identified several use cases and scenarios for 5G satellite networks development, herein collected and described in Table 7.1. It is noted that there is no specific use case for V2X vehicular communications. So far, there are no ongoing initiatives to deploy vehicle-to-satellite (V2S) technologies given the recent definition of both spectrum and radio access by satellite segment. It will be necessary to elucidate whether vehicular communication applications with their demanding latency requirements can be solved by direct communication between vehicle and satellite or will it be indirectly relying on the satellite network infrastructure to solve, for example, lack or deficiency of coverage.

In 2019, the WRC-19 allocated additional mm-wave frequency bands [14]: 24.25–27.5 GHz, 37–43.5 GHz, and 66–71 GHz for 5G terrestrial networks. In some countries and regions, frequency bands of 45.5–47 GHz and 47.2–48.2 GHz received complimented allocation to terrestrial segment of IMT (International

Table 7.2 Frequency bands allocated to fixed and mobile satellite services [31, 36]

Frequency range (GHz)		Bandwidth (GHz)
Uplink (UL)	Downlink (DL)	UL/DL
12.5–13.25	10.7–11.7	0.75/1.0
13.75–14.8	17.7–21.2	1.0/3.5
27.5–31.0	37.0–42.5	3.5/5.5
42.5–47.0	66.0–76.0	4.5/10.0
48.2–50.2	123.0–130.0	2.0/7.0
50.4–51.4	158.5–164.0	1.0/5.5
81.0–86.0	167.0–174.5	5.0/7.5
209.0–226.0	191.8–200.0	17.0/
252.0–275.0	232.0–240.0	23.0/8.0

Mobile Telecommunications) [14]. This decision of WRC-19 will enable to use some part of mm-wave bands on spectrum sharing basis for 5G satellite and 5G terrestrial network segments [14].

Table 7.2 shows the basic frequency bands allocated to fixed and mobile satellite services, sited within the band from 10.7 to 275 GHz, designed for satellite networks and satisfying the needs for 5G channel bandwidths [25, 31]. Typical frequency bands as Ka-band (28 GHz) and Q/V-bands (37–53 GHz) are listed, and it should be noticed that while Ka-band has a large tradition for satellite use, the V-band has not been used yet for satellite applications [31]. In this sense, WRC-19 pointed out some restrictions to apply to the different platforms [14]:

- HAPS base stations: Concern about interference in passive Earth observation operations in the band adjacent to 24 GHz was one of the most controversial topics at the WRC-19 conference. As a solution, the ITU established power limits so that 5G base stations are required to comply with the following power limits in the 24 GHz band: −33 dBW/200 MHz before and then −39 dBW/200 MHz after 2027. However, the meteorologist community that relies on passive Earth observation may not be satisfied with the compromise and considers that power limits do not adequately protect atmospheric water vapor sensors. On the other hand, lower-than-requested power limits will likely increase the number of base stations required to achieve 5G wide area coverage in the band.
- High-altitude platform systems (HAPS): Balloons, unmanned aircraft systems (UAS), and other HAPS technologies received assignments that will allow them to provide connectivity from altitudes between where the terrestrial and satellite networks operate. The ITU identified the following frequency bands for HAPS use: Global (31–31.3 GHz, 38–39.5 GHz, 47.2–47.5 GHz, and 47.9–48.2 GHz) and Region 2 (America) (21.4–22 GHz and 24.25–27.5 GHz).
- Satellites in geostationary satellite orbit (GSO): Earth stations in motion (ESIM) can now communicate with GSO satellites in the Ka-band (17.7–19.7 GHz and 27.5–29.5 GHz) and satellites in non-geostationary satellite orbit (NGSO).

For the deployment of the 5G satellite segment, the 3GPP [37] presented three options for orbits [31]:

- Geostationary satellites (GEO): Provision of full Earth coverage between 70°N and 70°S by a constellation composed of one to three satellites: propagation delay ranging from 120 to 140 ms.
- Medium Earth orbit (MEO): Providing full Earth coverage by a constellation composed from 10 up to 12 satellites: propagation delay ranging from 27 to 43 ms.
- Low Earth orbits (LEO): Continuous coverage by satellite network constellation composed of 50 up to 100 satellites: propagation delay ranging from 3 to 15 ms.

The 5G satellite options will allow 5G service delivery in areas with impeded provision by the 5G terrestrial segment and also for the services supported by satellite systems [31]. As reported in [35], the 5G satellite segment will complement 5G services, which are delivered especially on road, rail, and waterways as well as in rural regions, where access to 5G terrestrial segment is unavailable [31]. It should be noticed that 5G services provided via 5G satellite segment are not only data and voice communications but also connection provision for IoT/M2M devices, access to broadcasting services, as well as other delay-tolerant services [31].

As noted above and stated in the ITU Network 2030 vision [38], the space-terrestrial integrated network is considered as one of the key enablers in the B5G ecosystem. In this context, the proliferation of satellite networks faces many challenges in terms of integration with other networks or the adoption of new network management schemes according to this dynamic topology [37, 39]. To serve a fully mobile and connected society, 5G/B5G networks must experience remarkable growth in connectivity, density, and volume of data traffic, besides requiring multilevel ultra-densification. Hence, the satellite segment arises as a necessary part of the B5G ecosystem development [37].

B5G/6G systems should be characterized by advanced flexibility and reprogrammability, seamless global mobility in heterogeneous environments, and support for multiple radio access technologies. In fact, 3GPP clearly states that satellite communications can provide significant benefits in terms of coverage extension, increased network availability, and mobility management [40]. Thus, space networks can complement and extend terrestrial networks, and its integration can bring many benefits especially in terms of continuity, scalability, service management, reliability, and efficient resource utilization for data delivery [37].

Disruptive architecture concepts must be analyzed for application to high-capacity and high-frequency satellite communications, including links between satellites [37]. Enabling technologies for massive LEO architectures will be proposed, as well as the integration of space and terrestrial systems within the framework of 5G networks, considering the characteristics of geostationary systems and constellations of medium and low orbit within the NTN systems defined by the 3GPP [37, 41]. The aim is the design of novel mobility management solutions able to deal with the integrated space-terrestrial system and new radio access networks to support roaming between 5G satellite network and 5G terrestrial network [37].

On the other hand, the implementation of new low-orbit satellite thousand constellation communications systems (massive LEO) requires analyzing a new set of enabling radio technology. There are new system aspects that must be addressed such as interference management. Due to the dynamic nature of the orbits in LEO constellations, connection disruptions occur with both fixed access on earth and across the constellation between neighboring satellites on different planes [mip]. Obtained results [42, 43] demonstrate the high dependency of the mobility management algorithms in the channel model.

As aforementioned, the mm-wave spectrum will be a pioneer band to develop B5G/6G wireless mobile communications, for terrestrial and space applications. The incipient development of mm-wave B5G/6G communications will require disruptive technologies to successfully fulfill the expected features and, finally, the study of new case uses and applications. In this scenario, radio propagation behavior acquires a new dimension, and its proper modeling will contribute to ensuring the expected B5G/6G capacity [40].

The vehicle-to-satellite use case (V2S) is yet to be defined. The characterization of the mobile radio channels envisioned in B5G/6G systems for V2X use case, terrestrial or space, at millimeter frequency bands with high bandwidths, requires considering their highly changing temporal and spatial variability [42, 43]. Now the assumed condition of stationarity and uncorrelation is not fulfilled for most of the applications due to double mobility at the ends of the radio link, terrestrial or space, and forces us to rethink the existing radio channel models in order to account for the specific features.

Far from considering that the demand for radio channel characterization is complete, the results obtained also indicate that there are still many open research questions that are pending response [19]. One of those questions is the demand for developing more complex channel models that try to reflect as many possible propagation mechanisms and impairments taking place in vehicular communication [19], under the new scenario of the satellite segment.

7.4 Vehicular Wireless Communication Link Case Study

In order to gain insight in relation to performance metrics in relation with wireless channel behavior, a vehicular scenario communication link case study operating at 28 GHz frequency has been performed by means of an in-house three-dimensional ray-launching (3D-RL) code. By means of deterministic techniques such as the 3D-RL software, large- and small-scale propagation characterization can be performed. The 3D-RL algorithm is based on geometrical optics (GO) and the uniform theory of diffraction (UTD). For interested readers, the complete description of the algorithm can be found in [44], and its application in vehicular communication scenarios can be seen in [45].

In this case study, a typical urban scenario within a medium-sized city has been considered, where two different roads with parked cars are considered. A schematic

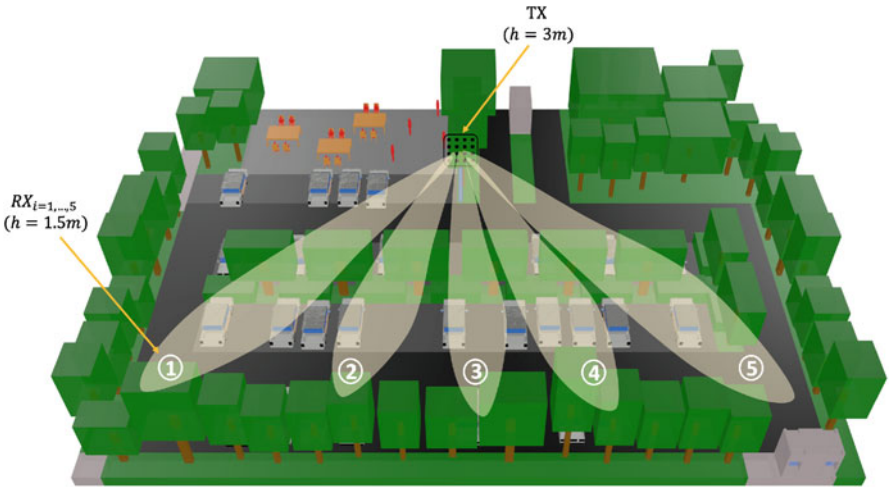


Fig. 7.2 Vehicular communication link scenario: infrastructure-to-vehicle (I2V) study case

Table 7.3 Simulation parameters

Parameters	Values
TX power	25 dBm
Carrier frequency	28 GHz
Bit rate	4.62 Gbps
Antenna type/beamwidth	Directional/14°
3D ray-launching: angular resolution/rebounds	0.2°/6
Scenario size/unitary volume analysis	(54 × 42 × 15) m/1m^3 (1 × 1 × 1) m

view of the modeled scenario can be seen in Fig. 7.2, where different types of trees and shrubbery are contemplated. The transmitter has been placed at 3 meters height on one side of the road, and five different receiver points have been analyzed, named as $RX_{i=1,...,5}$, at a typical vehicular receiver point (1.5 m height), emulating a receiver antenna on the vehicle ceiling. Simulation parameters can be seen in Table 7.3, based on previous extensive convergence analysis studies performed in relation to angular resolution, adaptive meshing, number of employed rays until extinction, and the consideration of diffraction phenomena, among others [44, 45].

As it can be seen in Fig. 7.2, five different beams have been considered for the transmitter antenna, in order to analyze the vehicular link communication of a receiver vehicle moving in the opposite road of the scenario. It is observed that the transmitter beams are affected by a tree barrier placed in the middle of the road, at a distance from the transmitter of approximately 11 m, which can affect signal propagation and lead to non-line-of-sight (NLOS) links in some cases. The path loss for each beam has been analyzed, and the fit to the close-in path loss model has been performed by means of the following equation:

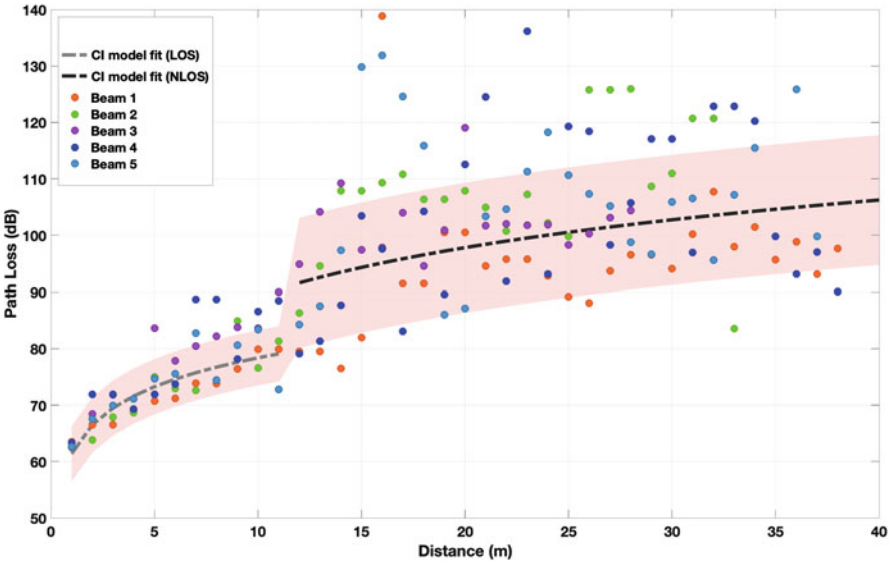


Fig. 7.3 Path loss for the five different beams considered in the vehicular scenario and close-in path loss model fit

$$PL_{CI}(f,d)[\text{dB}] = FSLP(f,d(1\,\text{m}))[\text{dB}] + 10n\log_{10}(d) + X_\sigma, \tag{7.1}$$

where $FSLP\,(f,d\,(1\,\text{m}))\,[\text{dB}] = 20\log_{10}\frac{4\pi f}{c}$ denotes the free space path loss in dB at a transmitter-receiver (TR) separation distance of 1 m at the carrier frequency f; c is the speed of light; n is the path loss exponent (PLE); d is the 3D separation distance between the transmitter and the receiver; and X_σ is the random variable for the shadow fading with standard deviation σ for large-scale signal fluctuations. Figure 7.3 shows the path loss for each of the beams and the close-in path loss model fit.

From Fig. 7.2, it is observed that path loss attenuation is under line-of-sight (LOS) conditions up to 11 m, where the different trees in the middle of the road are located. The obtained mean path loss exponent under LOS conditions is between 1.7 and 1.8. The highly reflective environment of the parked vehicles reports better conditions than free space losses. From the trees and vegetation barrier distance, NLOS conditions are encountered, with an increase in the path loss dispersion for the different beams. The obtained mean path loss exponent under NLOS conditions has been 2.79 with a mean standard deviation around 11.5 dB. As it can be seen, the tree barrier in the middle of the road highly affects signal propagation for this specific environment, typical for urban vehicular scenarios.

In order to have insight into the received power at different points in the scenario for the different beams, Fig. 7.4 presents a bidimensional plane of received power

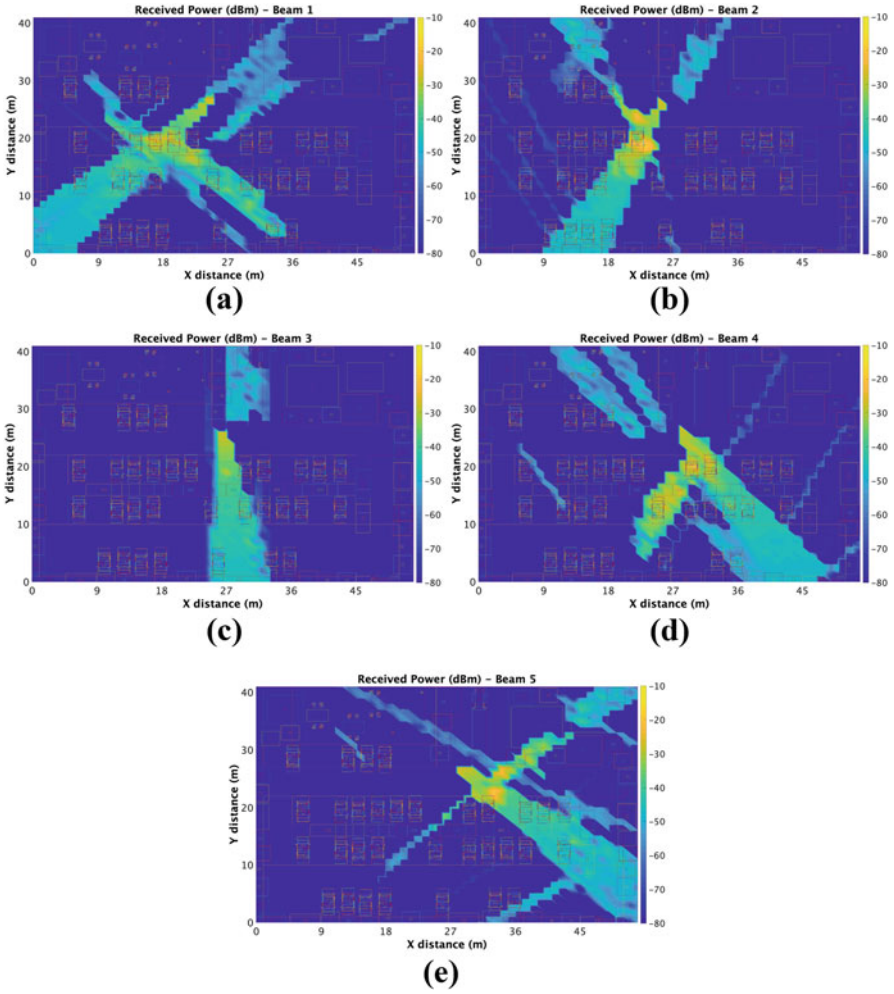


Fig. 7.4 Bidimensional planes of received power at 1.5 m height in the considered vehicular scenario for each beam: (**a**) beam 1, (**b**) beam 2, (**c**) beam 3, (**d**) beam 4, (**e**) beam 5

at 1.5 m height (typical receiver point in a vehicular scenario), for each one of the considered beams. As it can be seen from the figures, the morphology and topology of the scenario have a high impact in the signal attenuation. In this specific case study, the tree barrier and vegetation in the middle of the road have considerable effects, due to the NLOS condition impacts that provoke the different beams.

7.5 Challenges

Challenges in vehicular communication networks can be seen in terms of three principal perspectives. The first is the development and integration with other technologies, the second is about the main problems being addressed, and the third is about the paradigm on which the service needs to operate. Figure 7.5 shows the main issues in each of these three perspectives for the future challenges. In this manuscript, we discuss briefly some elements of these perspectives.

The evolution to address all the three perspectives needs to be developed within the smart communications (SC) paradigm. Smart communications (SC) are adaptive solutions integrating signal processing techniques with networking and communications and artificial intelligence (AI) algorithms. SC generates a platform on which paradigms such as VANETs and cognitive radio networks (CRN) can evolve integrating informed decisions and resource-aware adaptation. Adaptation, reconfigurability, and learning are components to become strong future wireless communications technologies. SC will change how we deal with complex environments with massive numbers of heterogeneous devices such as in smart cities, autonomous vehicles, unmanned aerial vehicles (UAV), intelligent transportation systems (ITS), internet of vehicles (IoV), 5G, and beyond 5G (B5G). In order to provide high-quality services for these massive numbers of devices, large amounts of information need to be stored, secured, processed, and transmitted.

With respect to technologies, the evolution towards 6G is key for a definition of the future. However, several recent initiatives have become relevant. The evolution of Internet of Things (IoT) has changed technologies, and now we have internet of

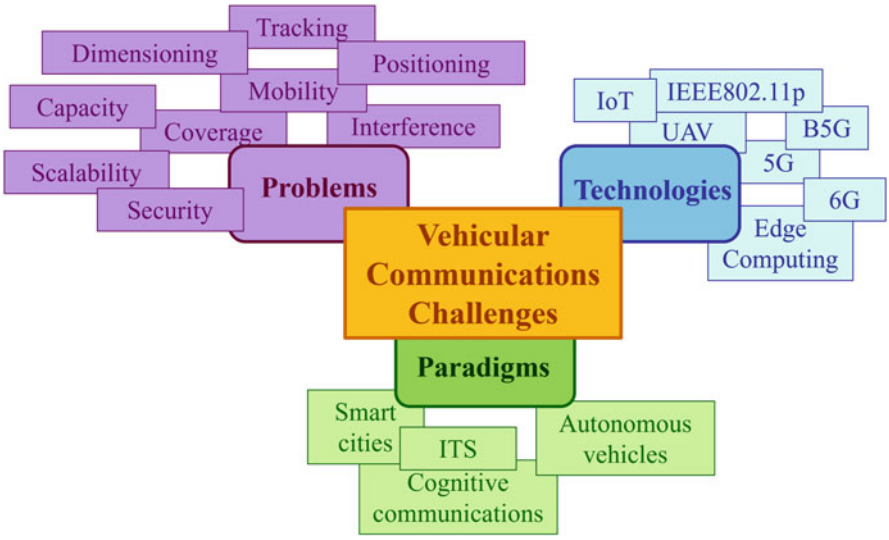


Fig. 7.5 Challenges seen from three different perspectives

vehicles (IoV) as a concept that connects anything, anyone, anytime, and anyplace [46]. IoV will serve to reduce problems such as collisions while at the same time increasing connectivity through low-latency links with reliability and high bandwidth. Resource sharing to achieve connectivity will be essential, and some solutions have been based on games and cooperative algorithms. The service will need infrastructure composed of distributed antenna systems and wireless devices or sensors that work with an open and distributed network architecture to guarantee service provisioning.

The integration of vehicles with UAV or drones is essential. UAV could help in safety and security tasks and will help to increase coverage in certain conditions of high volume of vehicular traffic. UAVs could be relays and also could serve as access points for certain activities, e.g., emergency scenarios with critical information delivery. The applications of UAVs are expanding from delivery of objects to supporting air traffic control tasks and infrastructure inspection; see [47]. For some of those activities, low latencies and moderate data rates are essential, depending on the need of a real-time connectivity. UAVs can even provide relay functions for remote infrastructure or even satellite communications integration with 5G networks or vehicular networks. Vehicles will have VANETs where a network can be established without the need for a support infrastructure, and the routing elements of the network are the same for vehicles. In comparison, UAVs can also be generating their own network, usually known as flying ad hoc network (FANET). In fact, FANETs are more flexible than VANETs, since the topology defining the connectivity of the nodes can vary in a way with structures such as mesh, star, linear, etc. The only restriction from the infrastructure point of view is regulation since UAVs must not interfere with other services and infrastructure. For VANETs this will not be possible due to the obvious restrictions encountered in the physical environment surrounding the vehicles.

As it was mentioned, UAVs could extend the coverage of cellular-based communications, especially in those remote areas where coverage is reduced, and UAVs could also integrate different communications solutions and establish links with different technologies and standards. UAVs could help extend line-of-sight (LOS) links or even create some where there are obstructions between transmitter and receiver. They could also help with cellular traffic management such as quality-of-service (QoS) delivery and handoff management. They could also help to create a 3D cellular network if equipped with appropriate technology.

In all wireless communications technologies, there are channel impairments that need to be measured, characterized, and modeled so that they help in the design of solutions to improve performance. A typical wireless multipath scenario that vehicles, mobile devices, and UAVs can experience is shown in Fig. 7.6.

Modeling of channel propagation phenomena is extremely important in order to establish a network plan for deployment, dimensioning, extension, and coverage. PHY layer and air interface solutions are based on these channel characterizations, and even future technologies consider them, especially with the changing radio spectrum where more bands become available for public communications services.

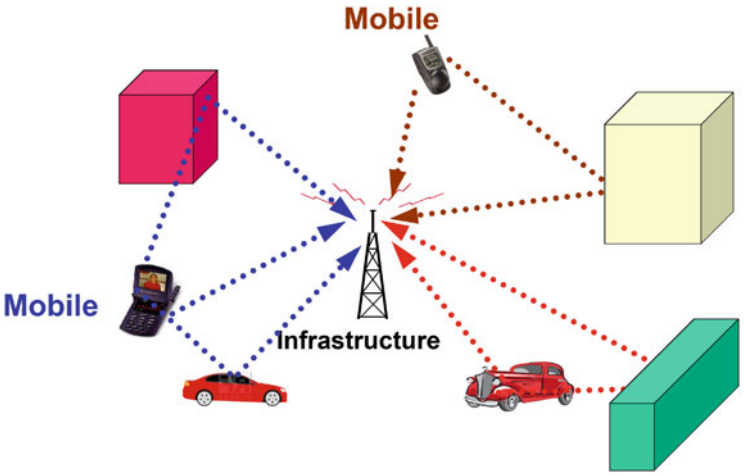


Fig. 7.6 Typical multipath channel for wireless communications

For UAVs, channel propagation measurements are essential since the scenarios involved are more complex than the classical planar point of view for cellular networks. Multipath environment is different and needs to be assessed in many different situations in urban and suburban environments. The use of multiple-input multiple-output antennas in UAVs is another challenge since the traditional radiation patterns can now be modified with beamforming techniques, but with a different advantage, UAVs can "accommodate" the beams not only by using signal processing algorithms for beamforming but also with physically changing positions that could improve communications [47].

There have been network architectures proposed for 5G [46]; however, architectures that consider V2X communications have not been widely developed. In [48], the authors propose an architecture for V2X communications in 5G that not only tries to deliver connectivity with reliability but also tries to solve security and privacy issues. It considers a network edge that works with the 5G infrastructure to deliver services and applications. Such network edge is connected to V2X application servers via a core network that is software based for flexibility. Even though vehicles can form VANETs, communication through the architecture is carried out from/to the 5G infrastructure. The V2X concept comprehends the classic V2I and V2V, and two more types are added, the vehicle to pedestrian (V2P) used for collision avoidance and driver alerts and the vehicle to network (V2N) where applications are in dedicated servers to which vehicles can connect for services such as traffic and navigation of weather. It is also important to include cooperation to achieve efficiency in the network.

From the perspective of problems needed to be addressed, it is important to have models validated with measurements and simulations in order to propose solutions for problems such as interference. As the number of vehicles grows, and the need

for information for autonomous driving and smart operation, interference becomes relevant to having reliable communications and network links. Densification, power control, new air interface standards, and coexistence strategies are key to solve the problem of interference [49]. Algorithms from statistical signal processing to mitigate or cancel interference need to evolve towards the smart paradigm where new machine learning and deep learning techniques are developed to be suitable to a scenario where mobility can change the environment in a very short time. Spatial stochastic modeling is important to establish behavior of coverage, capacity positioning, and tracking of vehicles and pedestrians. In contrast, security and privacy are needed for some of these operations in order to avoid cyberattacks and identity theft. Scalability for urban environments in vehicular networks will increase interference and reduce coverage unless vehicles and devices work together and align with coexistence, cooperation, and consensus to achieve high levels of quality in the network. Resource allocation and sharing will be essential, but the organization of the resources will also be fundamental to achieve an efficient communication network that delivers quality and reliability under widely varying conditions such as those in vehicular scenarios [50]. Signal processing for channel estimation needs to evolve by generating solutions in those widely varying conditions. The imperfect knowledge of the channel does not help to achieve high levels of quality and to mitigate channel impairments to a degree needed under these highly dense and highly mobile scenarios.

The road towards 6G is full of software-defined solutions to add flexibility and automation of network operations and functions. The main idea is to have a network as an autonomous system (AS) that can adapt itself according to network conditions such as traffic, outage, signal-to-noise ratio, cyberattacks, etc. Intelligence is key for this paradigm where in many operations human intervention will be eliminated. The network elements working cooperatively need to be able to also protect the infrastructure, the information, the use of the cloud, and the privacy of the users. Even service providers need to work cooperatively in order to achieve such adaptability, intelligence, and automation. Edge computing will assist in the resource sharing tasks and the communication and delivery of the information necessary for control of the network. New network paradigms where IoV, VANETs, and B5G work together are being proposed; see [51].

The challenges are many, and IoV is evolving and needs help to propose solutions to those challenges. In [51], some of the issues that we have mentioned are addressed, such as scalability, but from the perspective of a reliable delivery of information through an optimized way of sharing resources for vehicles. A discussion is also included about modern multiple access techniques for vehicles using beacons, connectivity of the vehicular network is analyzed, and convenient deployment of RSU within a network planning methodology is presented.

Challenges also evolve as technology does, and new smart techniques need to be developed in order to cope with new problems that arise with new communications services and technologies. Smart communications will help create a logical infrastructure where solutions can be developed that will permeate technologies,

paradigms, and problems finding optimized ways of communicating information across digital platforms such as those on ground mobile, aerial, or fixed networks.

7.6 Conclusions

Vehicular communications are one of the key elements in order to provide frameworks for the adoption of intelligent transportation systems, particularly for the case of connected and autonomous vehicles. The capabilities foreseen in current 5G systems and future B5G/6G systems enable wider range of V2X communication links, providing higher reliability in terms of reduced latency and higher achievable transmission bit rates, as well as increased coverage and security levels. Challenges in terms of network coordination between different network segments (e.g., ground-aerial-satellite), mobility, and interference control are currently being addressed in order to fulfill the foreseen requirements in future vehicular communication scenarios.

References

1. *Beyond 5G/6G Whitepaper*, National Institute of Information and Communications Technology (NICT) (2021)
2. *Beyond 5G/6G Whitepaper*, KDDI (2021). (Available Online) https://www.kddi-research.jp/tech/whitepaper_b5g_6g/
3. A. Pärssinen, in *White Paper on RF Enabling 6G – Opportunities and Challenges from Technology to Spectrum*, 6G Research Visions, No. 13 (2021)
4. *Beyond 5G Promotion Roadmap -Roadmap to 6G*, Ministry of Internal Affairs and Communications, Japan (2020). (Available Online) https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/presentation/pdf/Beyond_5G_Promotion_Strategy-Roadmap_towards_6G-.pdf
5. *IOWN initiative*, NTT. (Available Online) https://www.rd.ntt/iown/
6. *DoCoMo 6G White Paper*, DoCoMo (2021). (Available Online) https://www.nttdocomo.co.jp/corporate/technology/whitepaper_6g/
7. *Beyond 5G Vision White Paper*, NEC (2021). (Available Online) https://jpn.nec.com/nsp/5g/beyond5g/pdf/NEC_B5G_WhitePaper_1.0.pdf
8. *The Next Hyper – Connected Experience for All*, Samsung (2021). (Available Online) https://cdn.codeground.org/nsr/downloads/researchareas/20201201_6G_Vision_web.pdf
9. *Beyond 5G Promotion Consortium* (2021). (Available Online) https://b5g.jp
10. *NEXT G ALLIANCE* (2021). (Available Online) https://nextgalliance.org/
11. *6G Wireless Summit* (2021). (Available Online) https://www.6gsummit.com/
12. R.L. Aguiar, in *White Paper for Research Beyond 5G*, Networld (2015)
13. ITU *WRC-15: Provisional Final Acts*, World Radiocommunication Conference, Geneva (Italy) (2015). (Available Online) https://www.itu.int/dms_pub/itu-r/opb/act/R-ACT-WRC.12-2015-PDF-E.pdf
14. ITU *WRC-19: Provisional Final Acts*, World Radiocommunication Conference, Sharm El-Sheikh (Egypt) (2019). (Available Online) https://www.itu.int/en/ITU-R/conferences/wrc/2019/Documents/PFA-WRC19-E.pdf

15. *26 GHz and 28 GHz are Both Needed for 5G* (2021). (Available Online) https://www.gsma.com/spectrum/resources/26-ghz-28-ghz/
16. *GSMA Spectrum Position* (2020). (Available Online) https://www.gsma.com/spectrum/wp-content/uploads/2020/03/5G-Spectrum-Positions.pdf
17. *Satellite Technologic Evolution to the 5G ecosystem* (2019). (Available Online) https://www.itu.int/en/ITU-R/space/workshops/2019-SatSymp/Presentations/103%20%205G%20Satellites%20Eutelsat.pdf
18. L. Wei, R. Hu, Y. Qian, G. Wu, Key elements to enable millimeter wave communications for 5G wireless systems. IEEE Wirel. Commun. **21**(6), 136–143 (2014)
19. L. Azpilicueta, C. Vargas-Rosales, F. Falcone, A. Alejos, *Radio Wave Propagation in Vehicular Environments* (Editorial SciTech Publishing, 2020). ISBN 978-1-78561-823-9
20. F.A. Rodríguez-Corbo, L. Azpilicueta, M. Celaya-Echarri, A. Alejos, F. Falcone, Propagation models in vehicular communications. IEEE Access **9**, 15902–15913 (2021)
21. C.M. Silva, B.M. Masini, G. Ferrari, I. Thibault, A survey on infrastructure-based vehicular networks. Mob. Inf. Syst. **2017**, 6123868 (2017)
22. Study on LTE-based V2X services, TR36.885, 3GPP, (2016)
23. A. Bazzi, B.M. Masini, A. Zanella, Performance analysis of V2v beacon-ing using LTE in direct mode with full duplex radios. IEEE Wirel. Commun. Lett. **40**(6), 685–688 (2015)
24. V. Petrov, G. Fodor, J. Kokkoniemi, D. Moltchanov, J. Lehtomäki, On unified vehicular communications and radar sensing in millimeter-wave and low terahertz bands. IEEE Wirel. Commun. Lett. **26**(3), 146–153 (2019)
25. A.G. Volkswagen et al., 5G automotive vision, 5G Infrastructure Public Private Partnership (5G PPP), White paper (2015)
26. V. Petrov et al., Achieving end-to-end reliability of mission-critical traffic in softwarized 5G networks. IEEE J. Sel. Areas Commun. **36**(3), 485–501 (2018)
27. S. Chen et al., Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G. IEEE Commun. Stand. Mag. **1**(2), 70–76 (2017)
28. J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C.R. Bhat, R.W. Heath, Millimeter-wave vehicular communication to support massive automotive sensing. IEEE Commun. Mag. **54**(12), 160–167 (2016)
29. *5G Automotive Vision*, 5G-PPP Project (2014). (Available Online) https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf
30. B.M. Masini, A. Bazzi, A. Zanella, A survey on the roadmap to mandate on board connectivity and enable V2V-based vehicular sensor networks. Sensors, **18**(7), 2207 (2018)
31. V. Tikhvinskiy, V. Koval, Prospects of 5G Satellite Networks Development. Moving Broadband Mobile Communications Forward - Intelligent Technologies for 5G and Beyond, 1–17 (2021). https://doi.org/10.5772/intechopen.90943
32. *3rd Generation Partnership Project; Technical Specification Group Ser-vices and System Aspects; Study on Using Satellite Access in 5G; Stage 1 (Release 16)*, 3GPP TR 22.822 (2018)
33. *Space for 5G & 6G*, European Space Agency. (Available Online) https://artes.esa.int/esa-5g6g-hub
34. *Satellite Solutions for 5G*, CEPT ECC Report 280 (2018). (Available Online) https://docdb.cept.org
35. *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Scenarios and Requirements for Next Genera-tion Access Technologies; (Release 15)*, 3GPP TR 38.913 (2016)
36. F. Khan, *Mobile Internet from the Heavens* (Samsung Electronics, Richardson, 2015)
37. *Integrated Terrestrial and Space-based Millimeter Wave Radio System for Future Communication Ecosystem beyond 5G (InTerSpaCE-5G+)*, Project website . (Available Online) https://blogs.upm.es/interspace
38. *A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond*, ITU, Focus Group Network (2030). (Available Online) https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf

39. A.N. Uwaechia, N.M. Mahyuddin, A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: feasibility and challenges. IEEE Access **8**, 62367–62414 (2020)
40. *Service requirements for the 5G system*. 3GPP TS 22.261 v17.2.0 (2020)
41. K. Lioli, Use cases and scenarios of 5G integrated satellite-terrestrial networks for enhanced mobile broadband: The SaT5G approach. Int. J. Satellite Commun. Netw. **37**(2), 91–112 (2019). https://doi.org/10.1002/sat.1245
42. E.L. Cid, M. Portelas, M. García Sánchez, A. z Alejos, Microcellular radio channel characterization at 60 GHz for 5G communications. IEEE Antennas Wirel. Propag. Lett. **16**, 1476–1479 (2016)
43. E.L. Cid, M. Garcia Sanchez, A. Alejos, Wideband analysis of the satellite communication channel at Ku and X bands. IEEE Trans. Vehic. Technol. **65**(4), 2787–2790 (2016)
44. L. Azpilicueta, M. Rawat, K. Rawat, F. Ghannouchi, F. Falcone, A ray launching-neural network approach for radio wave propagation analysis in complex indoor environments. IEEE Trans. Antenn. Propag. **62**(5), 2777–2786 (2014)
45. L. Azpilicueta, C. Vargas-Rosales, F. Falcone, A.V. Alejos, *Radio Wave Propagation in Vehicular Environments*. An Imprint of The Institution of Engineering and Technology (SciTech Publishing, London, 2020)
46. M. Agiwal, A. Roy, N. Saxena, Next generation 5G wireless networks: a comprehensive survey. IEEE Commun. Surveys Tutor. **18**(3), 1617–1655 (2016)
47. Y. Zeng, Q. Wu, R. Zhang, Accessing from the sky: a tutorial on UAV communications for 5G and beyond. Proc. IEEE **107**(12), 2327–2375 (2019)
48. R. Lu, L. Zhang, J. Ni, Y. Fang, 5G vehicle-to-everything services: gearing up for security and privacy. Proc. IEEE **108**(2), 373–389 (2020)
49. W. Wu, R. Liu, Q. Yang, T.Q.S. Quek, Robust resource allocation for vehicular communications with imperfect CSI. IEEE Trans. Wirel. Commun. **20**(9), 5883–5897 (2021) https://doi.org/10.1109/TWC.2021.3070894
50. W. Zhuang, Q. Ye, F. Lyu, N. Cheng, J. Ren, SDN/NFV-empowered future IoV with enhanced communication, computing, and caching. Proc. IEEE **108**(2), 274–291 (2020) https://doi.org/10.1109/JPROC.2019.2951169
51. Y. Ni et al., Toward reliable and scalable internet of vehicles: performance analysis and resource management. Proc. IEEE **108**(2), 324–340 (2020). https://doi.org/10.1109/JPROC.2019.2950349
52. *5G White Paper. Satellite communications services: an integral part of the 5G ecosystem*, ESOA (2018). (Available Online) https://esoa.net/cms-data/positions/1693%20ESOA%205G%2016pp%20Booklet%2
53. Study on LTE support for Vehicle to Everything (V2X) services. TR22.885, 3GPP, (2016)
54. M. Celaya-Echarri, L. Azpilicueta, A. Vazquez Alejos, M. Garcia Sanchez, F. Falcone, P. Lopez-Iturri, Deterministic 3D ray-launching millimeter wave channel characterization for vehicular communications in Urban environments. Sensors, **20**(18), 5284 (2020)
55. M. García Sánchez, E. Lemos Cid, A. Vazquez Alejos, Empirical modeling of radiowave angular power distributions in different propagation environments at 60 GHz for 5G. Electronics **7**(2), 365 (2018)
56. V.V. Chetlur, H.S. Dhillon, Coverage and rate analysis of downlink cellular vehicle-to-everything (C-V2X) communication. IEEE Trans. Wirel. Commun. **19**(3), 1738–1753 (2020). https://doi.org/10.1109/TWC.2019.2957222

Chapter 8
Implementation of Context-Aware Environments with Massive IoT Systems

Imanol Picallo, Mikel Celaya-Echarri, Peio Lopez-Iturri, Leyre Azpilicueta, and Francisco Falcone

8.1 Introduction

The interconnection of devices and users within all application areas and sectors is the main driver for the implementation of Internet of Things (IoT). Since the initial conceptualization of IoT over 20 years ago, the evolution in microelectronic development and integration, wireless communication systems, database, and information handling capabilities has led to the adoption and extension of the IoT framework to different specialized sectors, such as healthcare (Internet of Medical Things, IoMT), Industry 4.0 (Industrial IoT), or vehicular communications (Internet of Connected Vehicles, IoCV), among others. Current forecasts indicate an increase to approximately 30 billion connected devices by 2025, with an increase to 50 billion devices by 2030 [1–4].

In this context, wireless communication systems play a key role in order to provide adequate coverage/capacity solutions while adequately handling interference levels in scenarios with operating devices exhibiting strong constraints in terms of form factor, cost, energy availability, and limited computational capabilities. Specific wireless communication system solutions have been proposed or are currently under development, such as LPWAN, 802.11ah, or PLMN-based, such as NB-IoT, LTE Cat. M, and the corresponding MTC/D2D connections within 5G NR (mainly within frequency range 1) [5–7]. This pathway, from 4G to 5G, is already enabling the number of connections to increase up to approximately one million

I. Picallo · P. Lopez-Iturri · F. Falcone (✉)
Electric, Electronic and Communication Engineering Department, Public University of Navarra, Pamplona, Navarra, Spain
e-mail: imanol.picallo@unavarra.es; peio.lopez@unavarra.es; francisco.falcone@unavarra.es

M. Celaya-Echarri · L. Azpilicueta
School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL, Mexico
e-mail: mikelcelaya@tec.mx; leyre.azpilicueta@tec.mx

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_8

devices/km^2. Future evolution in B5G systems foresees an even higher increase in device density, with inherently more advanced interference handling and energy optimization and reduction schemes. This will open the path for service provision in different application domains, such as holographic-type communications, extended reality, tactile internet, multisense experience, digital twins, pervasive intelligence, intelligent transport and logistics, and enhanced onboard and ubiquitous communications, among others [7]. The expected requirements are the following:

- Peak transmission rate in the order of 1 Tbps
- User experienced data rate in the order of 1 Gbps
- Latency: minimum values in the 10–100μs range and consideration of end-to-end delay metrics
- Mobility: maximum vehicular speeds of up to 1000 km/h
- Device connectivity density: up to 10^7 links/km^2, mainly focused on massive machine-type communications
- Energy efficiency: reduction in the 10 to 100 times ranges in terms energy efficiency per bit as compared with current 5G metrics
- Peak spectral efficiency: up to 90 bps/Hz in the downlink and 45 bps/Hz in the uplink

Several functionalities within B5G and 6G communication systems are foreseen in order to enable the required capabilities in order to adequately deliver service provision:

- Extended spectral use: new frequency bands are under consideration in the mm-wave range (up to 300 GHz, with main focus in 60 GHz bands for indoor use), THz range (up to THz, mainly focused on indoor communications, wireless backhaul, stringent QoS applications, Internet of Everything, and ultra-massive machine-type communications), visible light communications (VLC, from 400 to 800 THz, relying on line-of-sight communications, amenable to high-density hot spots, and massive IoT deployments), and optical wireless communications (OWC, spanning from IR range, visible and up to the lower UV wavelengths). The use of this extended frequency range can be further improved by using dynamic spectrum management, supported by cognitive radio and listen-before-talk contention schemes. The use of higher-frequency bands benefits from theoretical higher values of bandwidth and in some cases reduced interference levels as compared with sub 6 GHz. However, given the propagation characteristics, especially in upper mm-wave bands, operation is going to be mainly constrained to line-of-sight links with moderate link distances. Therefore, in the case where coverage extension is required, sub 6 GHz frequency bands will be employed in combination with future extended bands in order to comply with coverage/capacity requirements.
- Advanced network implementation and operation: more flexible network design and deployment are foreseen using network function virtualization and software-defined networks, aided by machine learning and artificial intelligence techniques in order to carry out complex orchestration tasks. Flexibility in terms of service

provision is achieved by taking advantage of network slicing and implementation time, and cost can be accelerated by means of initiatives such as Open RAN (O-RAN), providing an open source/open hardware architecture. In addition, new security considerations and mechanisms are being studied, in order to consider end-to-end security, quantum communications, or distributed authentication, among others. In relation to the network architecture, a more extensive configuration is foreseen, including satellite constellations, high-altitude platforms, and unmanned air vehicles, among others.

- Advanced air interface: further evolutions are being proposed in order to fully enhance the physical layer characteristics in B5G/6G systems. In this sense, different approaches are followed, such as the use of massive MIMO, intelligent reflecting surfaces, coordinated multipoint (such as cooperative MIMO), and new modulation schemes (such as evolutions in OFDMA or the use of power domain non-orthogonal multiple access, PD-NOMA).

Extended spectrum use is one of the main characteristics within B5G/6G communication systems, opening the path to enhanced communication capabilities for applications such as unmanned aerial vehicles in mm-wave range, providing integrated access and backhaul communication capabilities [8]. This extended spectral usage requires characterization of wireless channel characteristics in relation to gain insight of parameters such as path loss in order to perform coverage/capacity estimations. In this sense, propagation losses in extended frequency ranges starting at the mm-wave bands are considerable, which in general will require line-of-sight operating conditions. Wireless channel characterization is thus required in order to provide adequate insight in relation to frequency/power characteristics, such as received power level distributions as well as with time-domain-related characteristics, such as Doppler spectrum and power delay profiles. Multiple approaches can be employed, depending on the characteristics of the environment (i.e., outdoor/indoor, vehicular, industrial, human body presence), system functionalities under analysis, or required accuracy, going from empirical/statistical model to full-wave electromagnetic simulation or geometric/stochastic techniques based on ray tracing/ray launching models [9–20]. Wireless channel characterization can also provide additional information related to interference sources, models, and distribution, which have a direct impact on the coverage/capacity thresholds. Models, such as BG5CM, have been recently proposed in order to provide high-level wireless channel characterization, generalizable from mm-wave to THz frequency bands [21], as well as more specific models for application in scenarios related to IoT applications [22].

If the focus is placed in IoT communications, integrated sensing, computation, and communication capabilities are provided in order to enable massive IoT connectivity. This can be achieved by following multiple joint approaches, such as the use of non-orthogonal wireless communications at the sensing level, novel computational schemes to reduce latency, and advanced interference reduction mechanisms [23]. Enhancements have also been proposed in relation to network slicing operation, by means of federated learning approaches in order to minimize

raw data exchange while preserving network slice service-level agreement quality of service conditions [24]. Energy consumption reduction and control is another relevant aspect in the design and implementation of future B5G/6G networks, especially when considering ultra-dense network deployments. Approaches have been proposed in order to reduce energy consumption as well as overall interference levels by combining different energy reduction strategies, such as dynamic coverage selection and individual user power control [25]. Within the foreseen evolution in the adoption of future B5G/6G systems, the integration of IoT applications is being analyzed in terms of new architecture paradigms, focused on seamless integration and high-end device management [26].

Connectivity of devices within the context of IoT applications implies the consideration of additional factors inherent to their intended use:

- Reduced energy consumption: usually IoT applications must consider anytime, anywhere, and anything connectivity options. This leads to operating scenarios in which IoT nodes can be located in locations with non-guaranteed/non-stable energy resources. Moreover, the potentially massive deployment of IoT nodes makes it economically nonviable to perform periodic battery replacements. This requires the use of ambient energy sources, by means of different energy harvesting schemes, such as compact photovoltaic cells, thermoelectric effects (e.g., Peltier cells), piezoelectric materials, or electromagnetic energy harvesting, based on the use of rectennas. The use of ambient energy is usually combined with a rechargeable battery or supercapacitor, with an energy supervision and control mechanism, in order to further guarantee energy availability either when ambient energy availability is low or when specific operation duty cycles require more intensive energy usage. Constrained energy consumption can take advantage of energy reduction mechanisms based on energy-aware routing strategies and optimized device states in order to further increase operational time span of the devices within the network.
- Flexible wireless connectivity: devices can be connected in a wide variety of locations, in which line-of-sight as well as partial line-of-sight or non-line-of-sight conditions occur. This is strongly dependent on the application as well as on the scenario under analysis. Telemetering applications of distributed wireless sensors in the case of lampposts for street lights, for example, will generally exhibit line-of-sight or partial line-of-sight connectivity, owing to the inherent layout of the lampposts in urban settings (i.e., series connection to one electric control box). If we are connecting electric smart meters, there will usually be non-line-of-sight conditions (especially in the case of blocks of apartment houses, where the meters will be located in an indoor facility within the building) until connection is established with the corresponding aggregator/gateway device. In this way, different network topologies can be supported, such as star configurations (e.g., LPWAN LoRa/LoRaWAN, with the possibility of simultaneous connection to more than one gateway), mesh configurations (e.g., ZigBee mesh, IEEE 802.11s, or Bluetooth Low Energy), or cooperative communication schemes (such as device-to-device connectivity

within 5G). Multiple frequency bands can also be considered, although the vast majority of solutions take advantage of lower path losses and lower cost for sub 6 GHz frequency bands.

- Compact platforms: IoT applications inherently focus on large node densities, which require low-cost and small form factors. Additional considerations deal with operational conditions of the IoT nodes, which can be deployed in outdoor scenarios or industrial shop floors or logistical warehouses, among others. This leads to additional requirements in terms of ruggedization, with specific casings, connectors, or the use of resins in order to protect electronic circuits. Moreover, compact platforms are mandatory in the case of biophysical monitoring applications, with additional ergonomic requirements in the case of wearables or biocompatibility in the case of implantable devices.

In the following section, deterministic wireless channel analysis will be performed in order to gain insight in relation to potential deployments of IoT solutions. They will focus in the use of sub 6 GHz frequency bands, taking advantage of deployments using LPWAN, WLAN, and 5G connectivity, owing to increased coverage range, reduced device cost, and ease of integration and implementation with current transceiver technology as well as with incumbent wireless connectivity technologies such as MIoTy or rotating polarized wave (RPW) communication systems. These results can be coupled with B5G/6G solutions for fronthaul connectivity (i.e., massive data aggregation) as well as with other future localized hot spot solutions, such as VLC communications with line-of-sight connectivity requirements.

8.2 Wireless Connectivity Analysis

The need to provide massive connectivity in terms of node density, number of links per unit surface, variable coverage area, and inherent low-cost requirements recommends the use of sub 6 GHz spectrum for large transceiver deployments in relation to IoT applications. In order to gain insight in relation to IoT wireless channel connectivity performance as a function of node location/density and network topology, different outdoor as well as indoor scenarios have been analyzed. Wireless channel analysis has been performed with the aid of a deterministic geometric-stochastic approach, based on an in-house implemented 3D ray-launching algorithm, which considers the full complexity of the scenario under analysis in terms of the properties of all the scatterers within the scenario [27–31]. Three different scenarios will be described in this section, taking into account an outdoor urban setting, a hybrid indoor/outdoor setting, and an indoor location.

Measurement-based techniques prove to be accurate approaches to channel modeling given a specific environment. Nevertheless, these techniques can result costly both in economic and time terms [32]. In order to reduce these costs, geometry-based modeling approaches can be used achieving a good trade-off

between accuracy and time. In this work, an in-house three-dimensional ray-launching technique has been used to characterize the radio channel of the proposed environments. The algorithm is based on geometrical optics (GO) and the uniform theory of diffraction (UTD). It is worth noting that owing to the geometric-stochastic nature of the proposed algorithm, it consider any potential type of link for the whole 3D volume of the scenario under analysis (i.e., LOS, NLOS, or partial NLOS).

The transmitted and reflected waves are modeled by Fresnel equations. For a given interface of two mediums 1 and 2, where equivalent ray interaction occurs, the reflection coefficient $R^{\parallel\perp}$ and the transmission coefficient $T^{\parallel\perp}$ for the parallel and perpendicular polarization are obtained by the following expressions:

$$R^{\parallel} = \frac{E_r^{\parallel}}{E_i^{\parallel}} = \frac{\eta_1 \cos(\psi_i) - \eta_2 \cos(\psi_t)}{\eta_1 \cos(\psi_i) + \eta_2 \cos(\psi_t)} \tag{8.1}$$

$$T^{\parallel} = \frac{E_t^{\parallel}}{E_i^{\parallel}} = \frac{2\eta_2 \cos(\psi_i)}{\eta_1 \cos(\psi_i) + \eta_2 \cos(\psi_t)} \tag{8.2}$$

$$R^{\perp} = \frac{E_r^{\perp}}{E_i^{\perp}} = \frac{\eta_2 \cos(\psi_i) - \eta_1 \cos(\psi_t)}{\eta_2 \cos(\psi_i) + \eta_1 \cos(\psi_t)} \tag{8.3}$$

$$T^{\perp} = \frac{E_t^{\perp}}{E_i^{\perp}} = \frac{2\eta_2 \cos(\psi_i))}{\eta_2 \cos(\psi_i) + \eta_1 \cos(\psi_t)} \tag{8.4}$$

where $\eta_1 = 120\pi/\sqrt{\varepsilon_{r1}}$, $\eta_2 = 120\pi/\sqrt{\varepsilon_{r2}}$, ε_{r1} and ε_{r2} are the relative permittivity of the mediums 1 and 2 and ψ_i, ψ_r, and ψ_t are the incident, reflected, and transmitted angles, respectively. In the same way, diffraction phenomenon is taken into account by considering wedge diffraction, where new diffracted rays are generated. In the case of finite conductivity bidimensional wedge conditions, the calculation of diffracted components is characterized by detecting the corresponding object edges and applying the following diffraction coefficients [33]:

$$D^{\parallel\perp} = \frac{-e^{(-j\pi/4)}}{2n\sqrt{2\pi k}} \left\{ \begin{array}{l} \cot\left(\frac{\pi+(\phi_2-\phi_1)}{2n}\right) F(kLa^{+}(\phi_2-\phi_1)) \\ +\cot\left(\frac{\pi-(\phi_2-\phi_1)}{2n}\right) F(kLa^{-}(\phi_2-\phi_1)) \\ +R_0^{\parallel\perp} \cot\left(\frac{\pi-(\phi_2+\phi_1)}{2n}\right) F(kLa^{-}(\phi_2+\phi_1)) \\ +R_n^{\parallel\perp} \cot\left(\frac{\pi+(\phi_2+\phi_1)}{2n}\right) F(kLa^{+}(\phi_2+\phi_1)) \end{array} \right\} \tag{8.5}$$

where $n\pi$ is the wedge angle, F the line source model, L the distance parameter (variable as a function of the propagating wave), a_{-}^{+} the diffraction angle-dependent coefficient, $R_{0,n}$ the reflection coefficients as a function of the corresponding material layer interfaces, and ϕ_2 and ϕ_1 the angles to the observation points before and after the diffraction wedge [34]. The resulting total E-field components are given by:

$$E_{GO}^{\perp\parallel} = \sqrt{\frac{P_{rad} D_t(\theta_t, \emptyset_t)\eta_0}{2\Pi}} \frac{e^{-j\beta_0 r}}{r} X^{\perp\parallel} L^{\perp\parallel} \tag{8.6}$$

$$E_{UTD}^{\perp\parallel} = e_0 \frac{e^{-jks_1}}{s_1} D^{\perp\parallel} \sqrt{\frac{s_1}{s_2(s_1+s_2)}} e^{-jks_2} \tag{8.7}$$

where $\beta_0 = 2\pi f_c \sqrt{\varepsilon_0 \mu_0}$, $\varepsilon_0 = 8.854 \cdot 10^{-12} F/m$, $\mu_0 = 4\pi \cdot 10^{-7} H/m$, and $\eta_0 = 120\pi\,\Omega$. P_{rad} is the radiated power of the transmitter antenna, whereas $D_t(\theta_t, \emptyset_t)$ is the directivity considering a spherical coordinate system at an elevation angle θ_t and an azimuth angle ϕ_t. Parameter e_0 is the free space field strength, k is the propagation constant, and $s1$ and $s2$ are the distances from the source to the edge and from the edge to the receiver point, respectively. This 3D RL tool can provide radio channel results such as RF power distribution, path loss, and time-domain results (e.g., power delay profiles). Table 8.1 presents the simulation parameters employed in order to perform the different simulation runs for the three different environments under analysis in this chapter. The parameters have been chosen based on criteria from previous convergence and meshing analysis studies [31].

8.2.1 University Campus Scenario

The first scenario under analysis is the campus of the Universidad Pública de Navarra (UPNA), located in the city of Pamplona, in Spain. The campus has an approximate surface of 200.000 m^2, with 16 buildings on premises and with a community of 8000 students and over 1500 research, teaching, and administrative staff. Within the campus, IoT connectivity solutions have been planned for multiple applications, such as botanical garden monitoring or energy smart metering options. A schematic representation of the simulation model implemented to perform wireless channel analysis is depicted in Fig. 8.1. Different transceivers have been located within the scenario, in order to model the behavior of potential IoT nodes performing infrastructure monitoring tasks, in locations such as lampposts. Wireless channel estimations have been obtained for the complete scenario under test, considering any potential receiver location within the simulation volume, for two different frequencies of operation (868 MHz, for the case of LPWAN; 2.4 GHz, for the case of WLAN/BT/BLE; and 3.5 GHz, for the case of 5G NR FR1 frequency bands).

Estimations of received power levels are obtained for the complete volume of the scenario under analysis, including the inner portions of buildings, which have also been modeled, considering their topology as well as the frequency dispersive material characteristics (i.e., conductivity and dielectric constant). In the scenario, multiple transceivers with variable locations can be considered in order to analyze frequency/power as well as time-dependent characteristics. Estimations for different frequencies of operation are depicted in Fig. 8.2, considering a given transceiver location, and for the different frequencies of operation under analysis

Table 8.1 3D ray-launching simulation parameters

Parameter	Scenario 1 (UPNA)	Scenario 2 (city center)	Scenario 3 (warehouse)
Output power	0 dBm	0 dBm	0 dBm
@freq	868 MHz, 2.4 GHz, 3.5 GHz	868 MHz,2.4 GHz, 3.5 GHz	868 MHz, 2.4 GHz, 3.5 GHz
Antenna type	Monopole	Monopole	Monopole
Max #reflections	6	6	6
Cuboid resolution	5 m × 5 m × 5 m	6 m × 6 m × 2 m	0.5 m × 0.5 m × 0.5 m
Ray resolution	1°	1°	1°
Diffraction effect	Activated	Activated	Activated

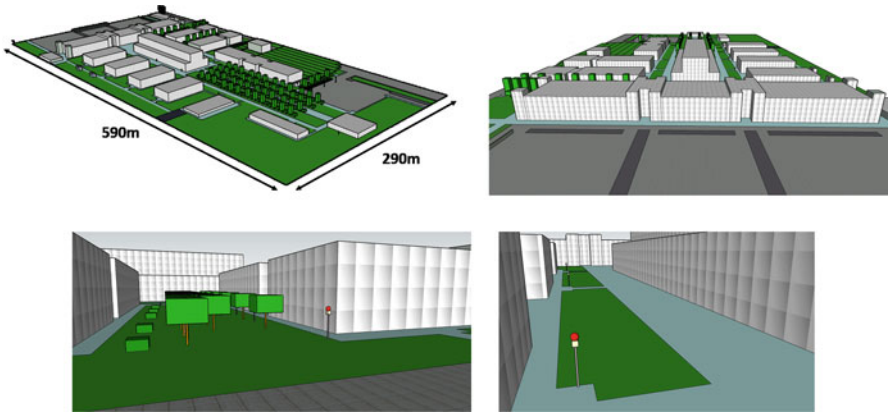


Fig. 8.1 Representation of the campus of the Universidad Pública de Navarra, Pamplona, Spain. The schematics represent the simulation model implemented, which considers the complete building structures of the campus. Details for two of the transceiver locations within the scenario are depicted in the bottom two images

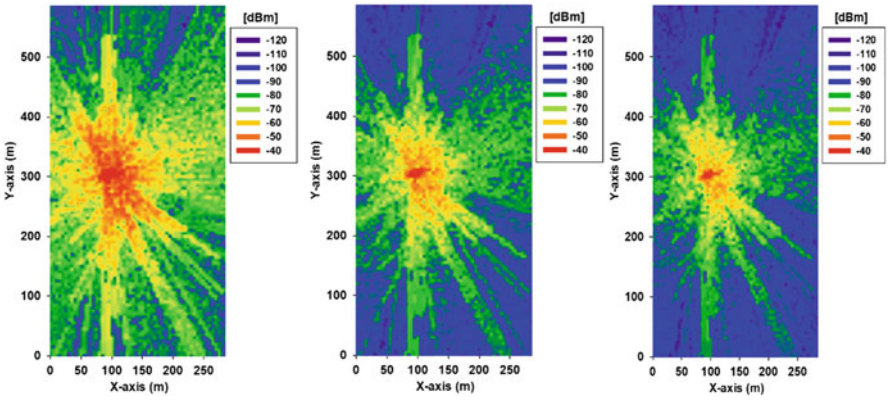


Fig. 8.2 Estimation of distribution of received power levels for the UPNA campus scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for constant height horizontal planes at a cut-plane height of 3.2 m

(i.e., 868 MHz, 2.4 GHz, 3.5 GHz), for horizontal bidimensional planes of constant height. It can be seen that by increasing the frequency of operation, received power levels decrease, as a direct consequence in higher path losses as well as in the increase in shadowing losses for refracted waves within the scenario. Similar results have also been obtained and are depicted in the case of vertical planes, for the three frequencies of operation under analysis in Fig. 8.3.

More specific details in relation to wireless signal variation as a function of transceiver location and surrounding environment can be seen by considering linear radial distributions of received power level in TX-RX links, shown in Fig. 8.4.

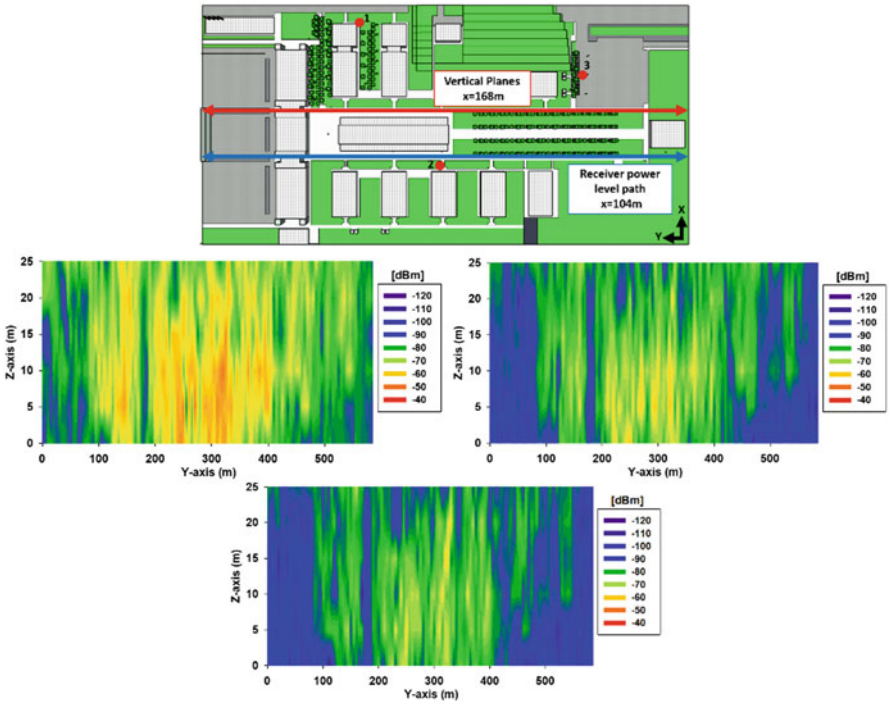


Fig. 8.3 Estimation of distribution of received power levels for the UPNA campus scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for constant vertical planes at cut planes. The location for the cut planes is depicted schematically in the top image

Different distributions are depicted in figure for all three frequencies of operation and considering two different link heights (i.e., $h = 3.2$ m and $h = 8.7$ m). The effects of fast fading can be clearly seen, with power dispersion in excess of 40 dB as a function of frequency and location, with a significant increase as the frequency of operation increases from 868 MHz to 3.5 GHz.

It can be seen from the results obtained that an increase in the frequency of operation leads as expected to increased path losses and, hence, reduced coverage levels. Moreover, the presence of infrastructure elements of different types strongly influences coverage levels, owing to shadowing effects, penetration losses by propagating wave refraction, multipath propagation, diffraction effects on edges (e.g., buildings), and diffuse scattering effects (i.e., leaves in tree canopies or equivalent rough surfaces by non-uniform wall facades). This can be clearly seen in the case of multipath propagation, given by the large density of scattering elements within the scenario under analysis, leading to rich multipath components in time domain. In this sense, estimations of power delay profiles have also been obtained for the scenario and are depicted in Fig. 8.5.

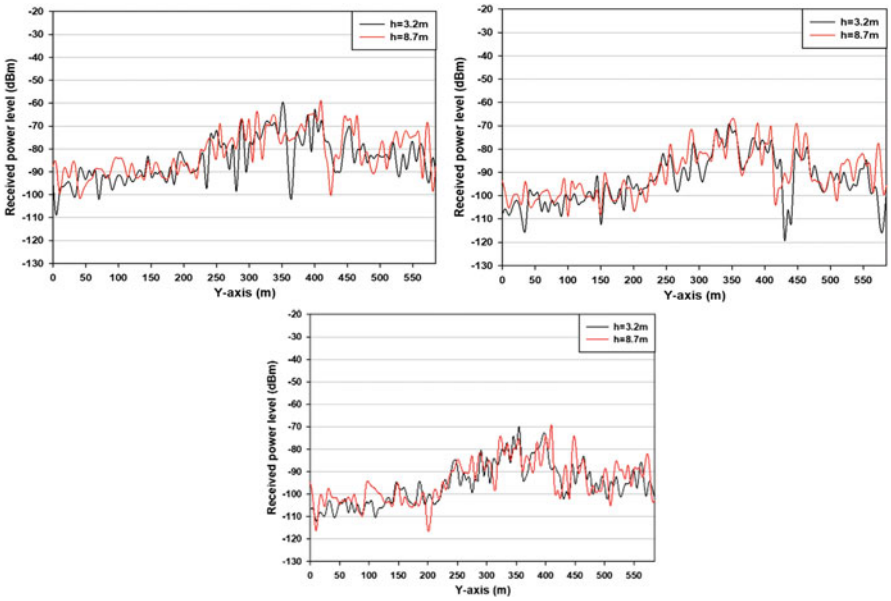


Fig. 8.4 Estimation of distribution of received power levels for the UPNA campus scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for a given linear radial TX-RX distribution, considering heights of 3.2 m and 8.7 m, respectively

The results obtained for the power delay profile distributions show relevant differences in terms of the number of multipath components detected, as well as their time distribution, with delay spread values in the range of 2000 ns to over 5000 ns, depending on the transceiver location. This behavior is directly related to the location of the transceivers as well as by the configuration of the scattering elements in the scenario under test. This enables to retrieve additional parameters, such as delay dispersion distributions for the complete volume of operation, which in turn provides information related to coherence time or Doppler shift. In this way, spatial locations for the potential network node deployments can be optimized in order to enhance their operation.

8.2.2 Urban City Center Scenario

A second scenario has been analyzed, in this case an urban setting within the center of the city of Pamplona in Spain. This scenario can be considered a typical dense urban scenario, in which a relevant amount of vehicular traffic as well as domestic households will be present. A schematic of the scenario that has been implemented in order to perform the deterministic 3D RL analysis is presented in Fig. 8.6. As in the case of the previous scenario, it enables the analysis of indoor/outdoor hybrid

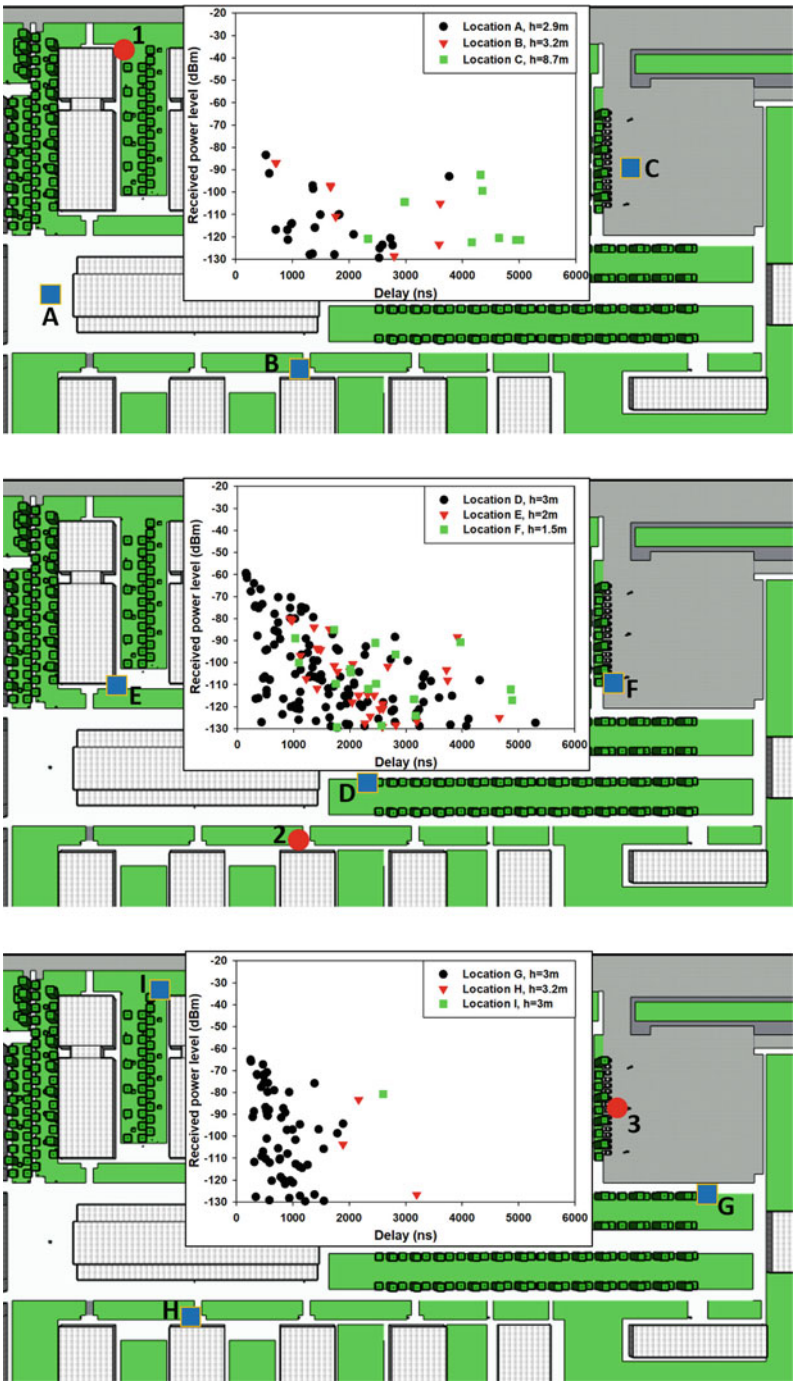


Fig. 8.5 Estimation of power delay profiles for three different transceiver locations, considering frequency of operation of f@868 MHz. The location for each one of the transceiver locations is given by a red dot within the scenario

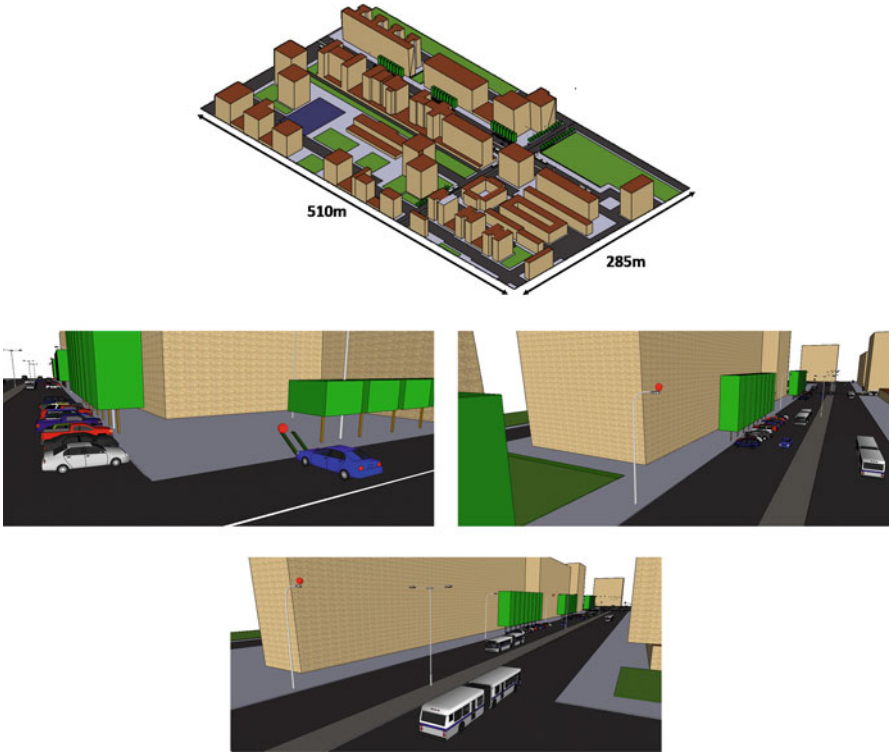


Fig. 8.6 Representation of a portion of the city center of Pamplona, Spain. The schematics represent the simulation model implemented, which considers the complete building structures within the city. Details for three of the transceiver locations within the scenario are depicted

links, although the main focus in this case is potential connectivity of elements with higher channel dynamics, such as different vehicle types.

Received power levels have been obtained within the scenario under analysis, and horizontal as well as vertical cut planes have been obtained and are depicted, for different frequencies of operation, in Figs. 8.7 and 8.8.

As in the previous case of the UPNA campus scenario, received power level distributions vary as a function of the frequency of operation, exhibiting increased path losses as the frequency is increased. In this case, it is worth noting a more uniform received power level distribution in the case of the central vertical portion of the scenario which is coincident with the presence of one of the main streets within the scenario, enabling equivalent guiding effect from the surrounding buildings. Received power levels are in general lower, which is given by the fact that the receiver height considered in the results depicted in Fig. 8.7 has been 1.5 m, consistent with opportunistic sensors located with users or vehicles. This leads to higher percentage of partial line-of-sight or non-line-of-sight conditions. Thus, transceiver location changes imply modifications in line-of-sight conditions,

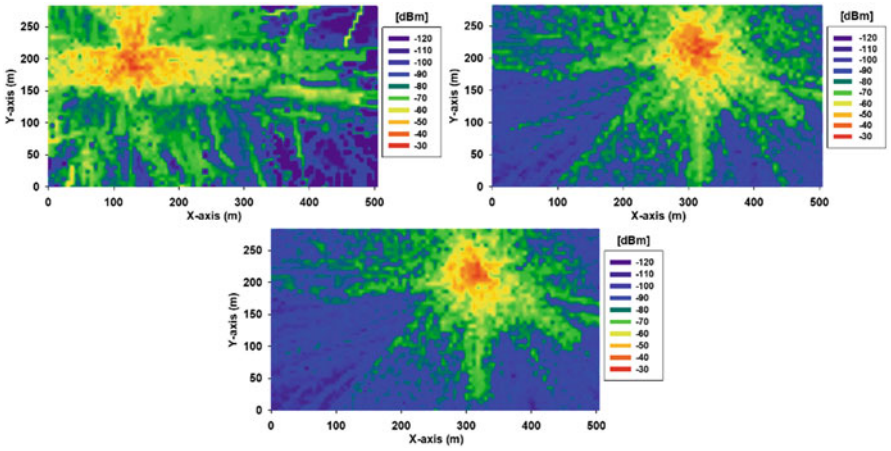


Fig. 8.7 Estimation of distribution of received power levels for the Pamplona city center scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for constant height horizontal planes at a cut-plane height of 3.2 m

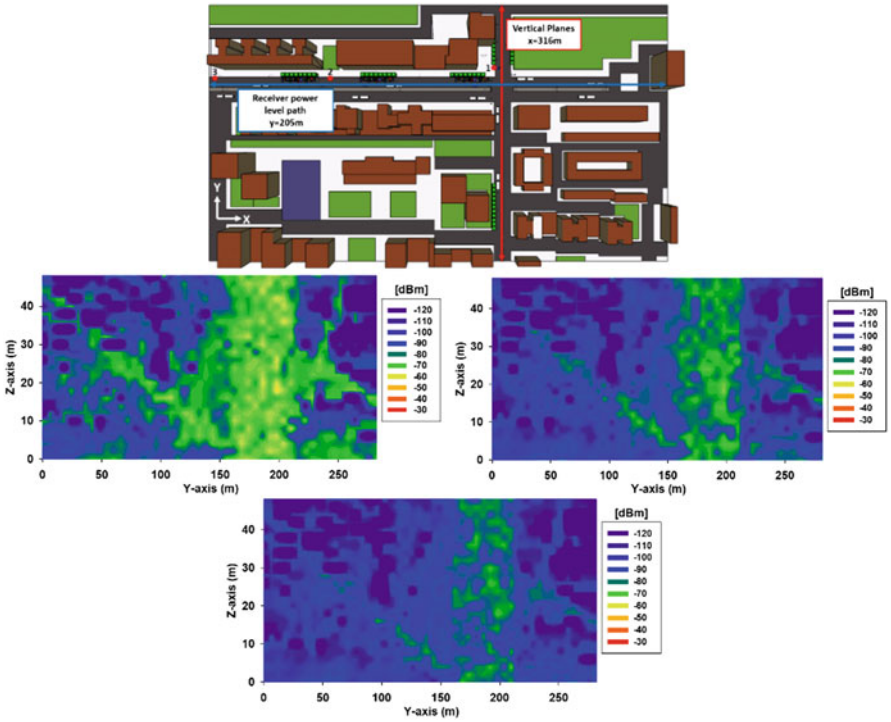


Fig. 8.8 Estimation of distribution of received power levels for the Pamplona city center scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for constant vertical planes at cut planes. The location for the cut planes is depicted schematically in the top image

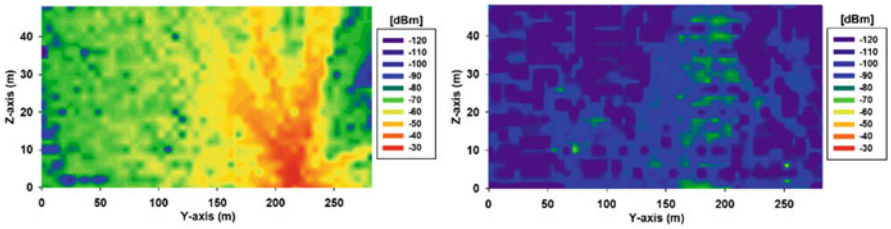


Fig. 8.9 Estimation of distribution of received power levels for the Pamplona city center scenario, considering frequency of operation of f@2.4 GHz, considering central (left image) and lateral (right image) transceiver location

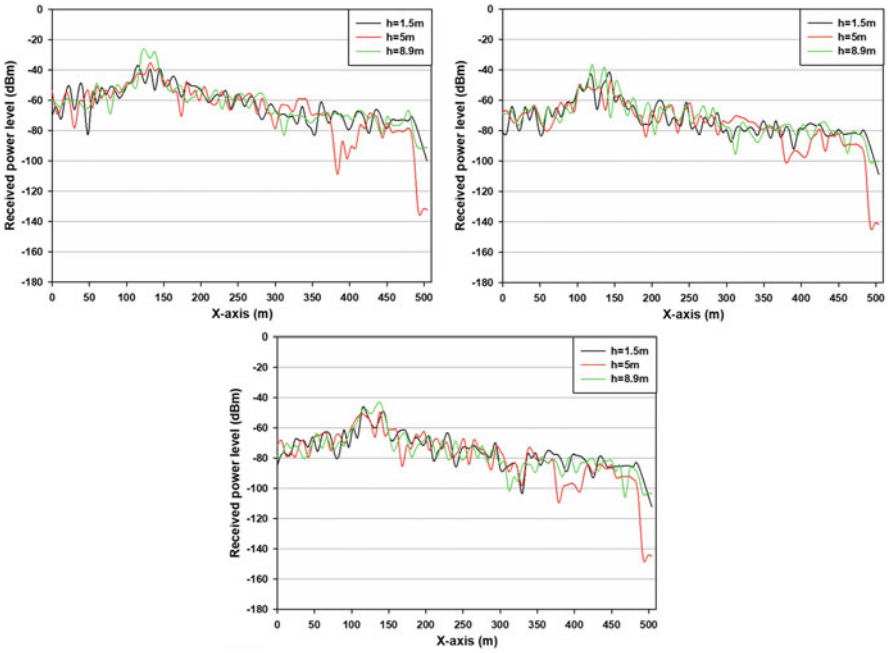


Fig. 8.10 Estimation of distribution of received power levels for the Pamplona city center scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for a given linear radial TX-RX distribution, considering heights of 1.5 m, 5 m, and 8.9 m, respectively

increasing the percentage of line-of-sight or partial line-of-sight links, which can be observed in the results depicted in Fig. 8.9, considering a central antenna position and a lateral antenna position, for a frequency of operation of 2.4 GHz. Linear radial TX-RX distributions have also been obtained and have been depicted in Fig. 8.10, for all frequencies of operation and considering link heights of 1.5 m, 5 m, and 8.9 m, corresponding to potential location of user as well as infrastructure nodes within the scenario, showing clear path loss dependence with frequency, with increasing losses as the frequency of operation spans from 868 MHz to 3.5 GHz.

Time-domain characterization for the city center portion of Pamplona scenario has also been obtained by means of power delay profile distributions for three different antenna locations, considering a frequency of operation in this case of 3.5 GHz, shown in Fig. 8.11. As in the previous case, clear differences can be seen in the multipath component distribution as a function of the transceiver location and the distribution of infrastructure elements within the scenario under test, which in turn impacts reflection and scattering phenomena of the propagating wireless signals. Large variations can be observed in terms of delay spread, with values ranging from 1000 ns to over 6000 ns as a function of transceiver location.

8.2.3 Indoor Warehouse Scenario

A final test case has been considered, corresponding to a large indoor warehouse location. This specific type of scenario is of particular interest in the case of IoT applications, as wireless nodes can provide multiple functionalities in relation to location and tracking of assets, as well as for building automation and control purposes among others. The schematic representation of the scenario under analysis, which corresponds to a real logistic warehouse location in the vicinity of the city of Pamplona, is depicted in Fig. 8.12.

In this case, the wireless channel characteristics will be strongly dependent on the density and type of objects within the scenario, as in principle, a higher percentage of communication links will be operating in partial line-of-sight or non-line-of-sight conditions. Therefore, in order to consider the impact of this condition, the scenario has been implemented with a variable load setting in the pallets that are located in the vertical rack structures. To illustrate the impact of loading assets within the scenario (and hence, modifying the line-of-sight/partial line-of-sight/non-line-of-sight conditions), two different results have been obtained: for an empty warehouse (i.e., no loading on the pallets) and for a fully loaded warehouse (i.e., all of the pallets have been introduced inside the simulation scenario). The results for received power level distributions, for different frequencies of operation under consideration and the two loading level conditions, are shown in Figs. 8.13 and 8.14. The presence of assets within the scenario leads to higher path loss and, hence, to lower received power levels, which is given by a strong decrease in line-of-sight and partial line-of-sight propagating conditions in the case of loaded pallets. The results shown in Fig. 8.14 have been obtained for different transmitter locations, which also indicate the strong dependence of the wireless channel characteristics with network topology, which is more relevant in the case of indoor scenarios. This is once again given by the lower clearance conditions of wireless links within the indoor scenario, usually limited by the presence of walls and higher density of clutter. Different effects such as signal guiding within the central corridor or predominant shadowing effects within lateral locations of the scenario can be seen. Similar results have also been obtained for vertical cut planes within the indoor warehouse scenario for unloaded and loaded pallets, which are shown in Figs. 8.15 and 8.16, respectively.

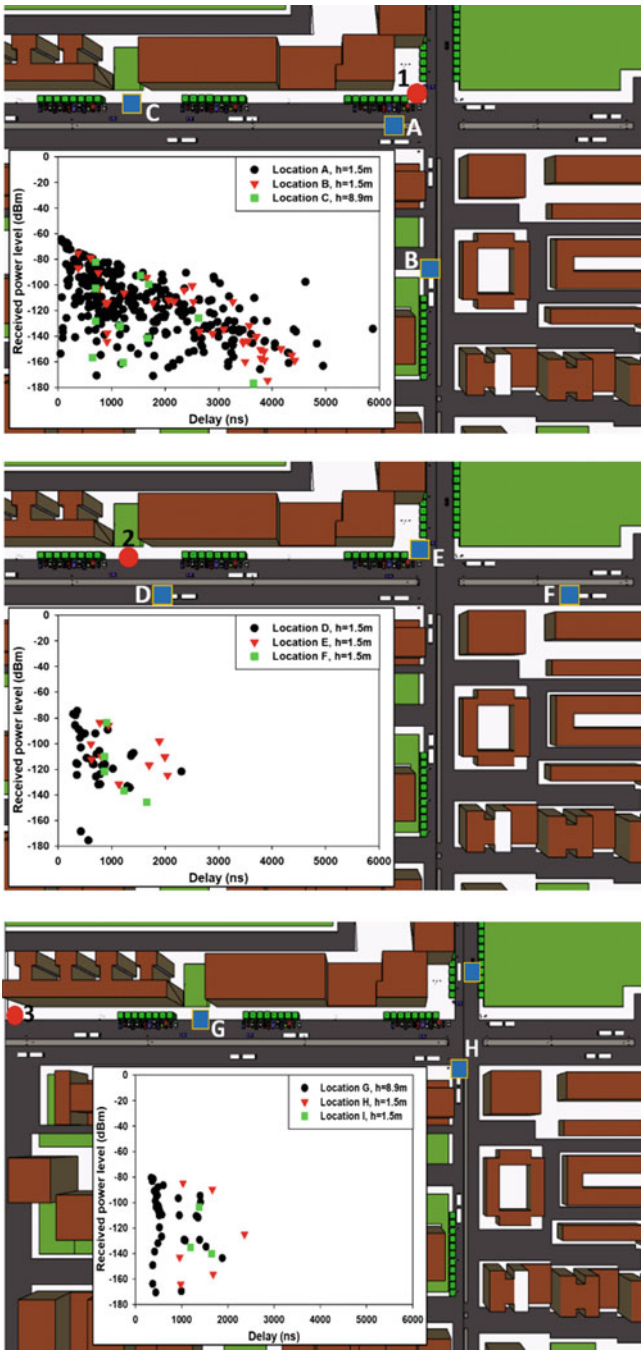


Fig. 8.11 Estimation of power delay profiles for three different transceiver locations, considering frequency of operation of f@868 MHz. The location for each one of the transceiver locations is given by a red dot within the scenario

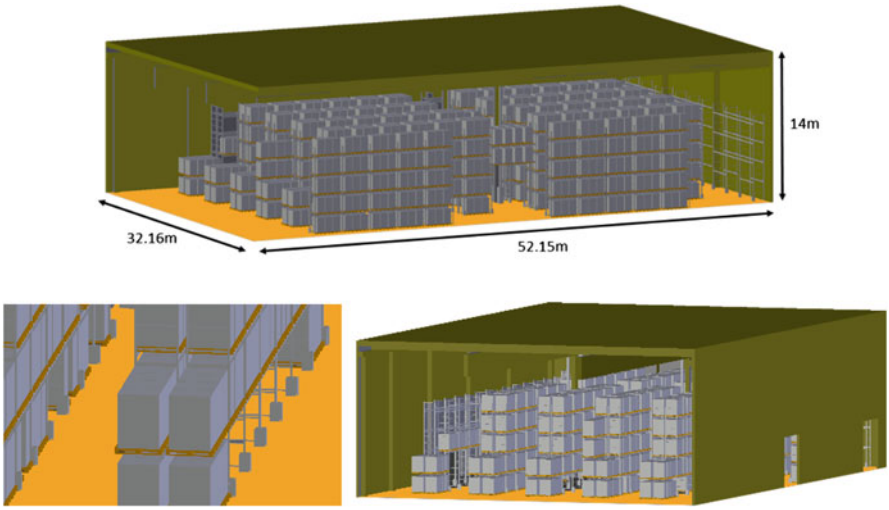


Fig. 8.12 Representation of the warehouse scenario implemented in order to perform wireless channel analysis based in 3D RL approach. The scenario is a logistic operator warehouse, in which a varying density of pallets with variable loads of assets can be located at different height levels within the scenario

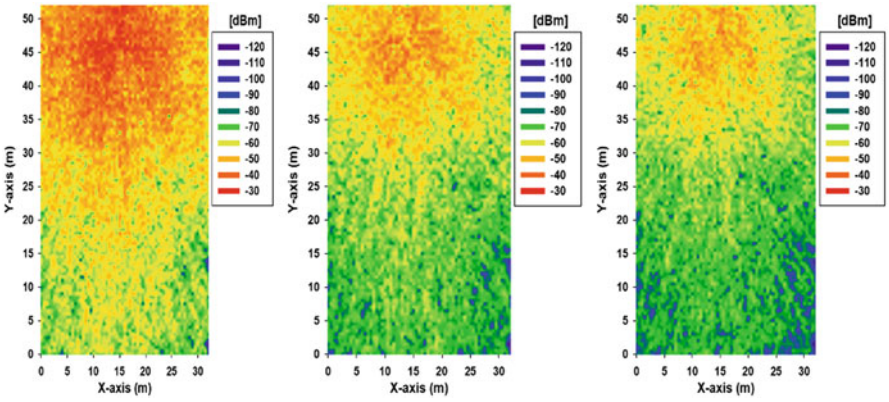


Fig. 8.13 Estimation of distribution of received power levels for the indoor warehouse scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for constant height horizontal planes at a cut-plane height of 1.5 m in the case of unloaded pallets within the scenario

As in the previous cases, linear TX-RX radial distributions of received power levels have been obtained for the scenario, considering both unloaded and loaded pallets within the warehouse, depicted in Fig. 8.17. Three different link heights have been considered (i.e., $h = 2\,\text{m}$, $h = 3\,\text{m}$ and $h = 4\,\text{m}$), corresponding to potentially different distribution of transceivers (either for static infrastructure as well as coupled to the pallets). Significant fast fading components can be observed

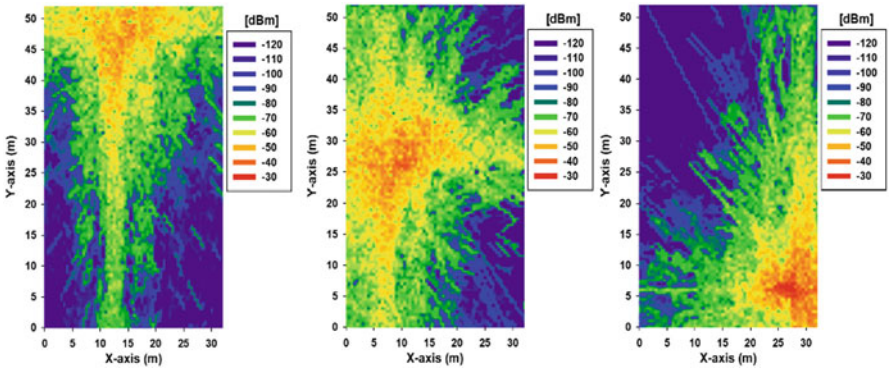


Fig. 8.14 Estimation of distribution of received power levels for the indoor warehouse scenario, considering frequency of operation of f@3.5 GHz, for constant height horizontal planes at a cut-plane height of 1.5 m in the case of loaded pallets within the scenario. Three different transmitter antenna locations have been depicted, to illustrate varying channel characteristics owing to topological variations within the scenario

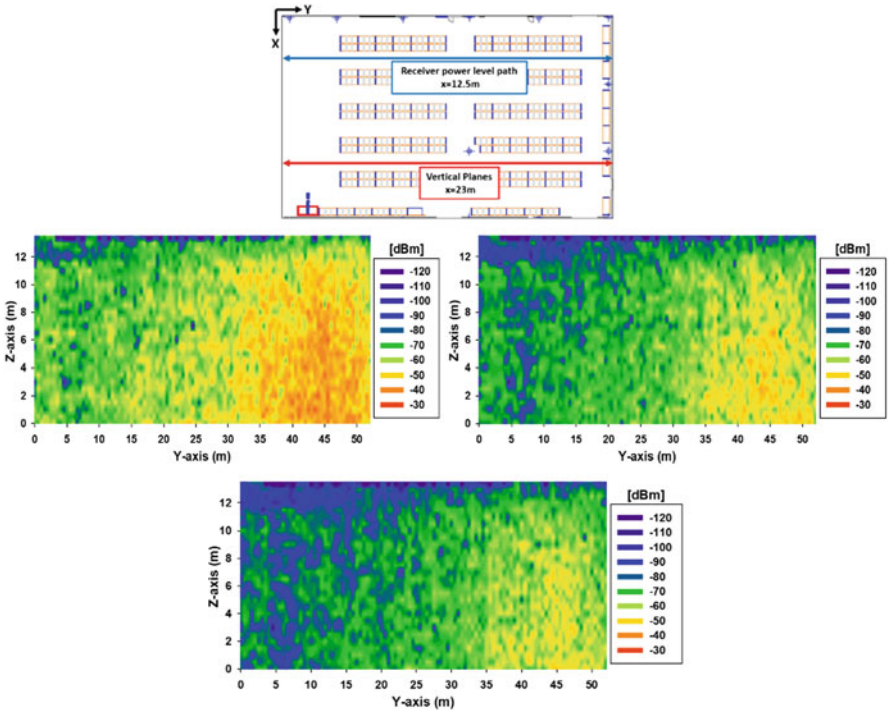


Fig. 8.15 Estimation of distribution of received power levels for the indoor warehouse scenario, considering frequencies of operation f@868 MHz, f@2.4 GHz, and f@3.5 GHz, for vertical planes in the case of unloaded pallets within the scenario. The location of the vertical planes is shown in the upper image of this panel

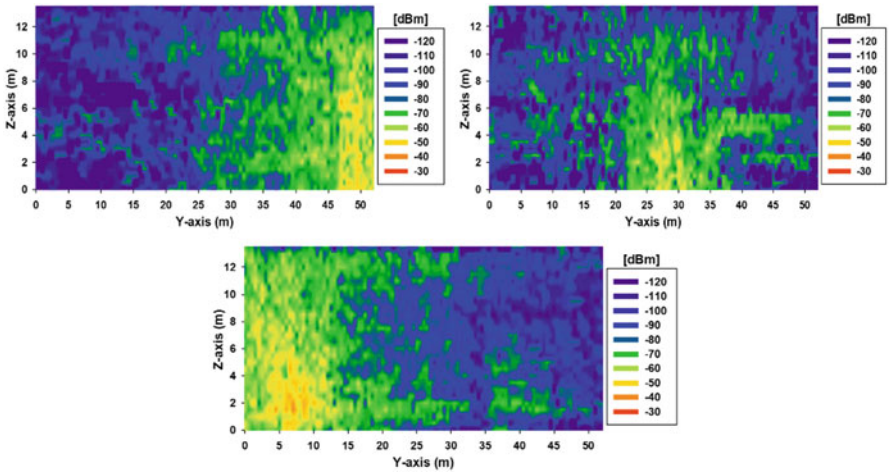


Fig. 8.16 Estimation of distribution of received power levels for the indoor warehouse scenario, considering frequency of operation of f@3.5 GHz, for vertical in the case of loaded pallets within the scenario. Three different transmitter antenna locations have been depicted, to illustrate varying channel characteristics owing to topological variations within the scenario

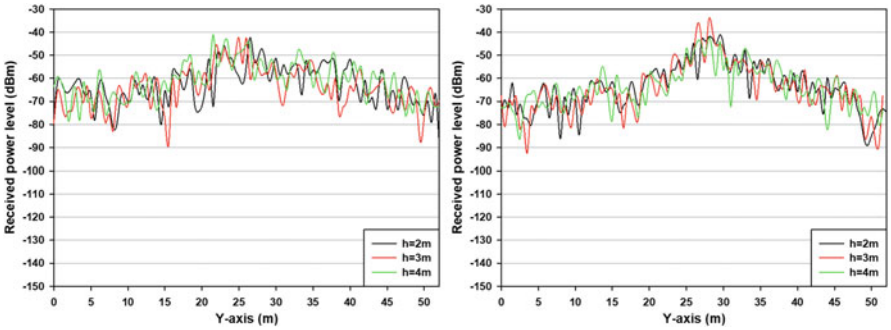


Fig. 8.17 Estimation of distribution of received power levels for the indoor warehouse scenario, considering frequency of operation of f@3.5 GHz, for linear TX-RX radials in the case of unloaded (left) and the case of loaded pallets (right) within the scenario, considering three different heights of operation ($h = 2$ m, $h = 3$ m, and $h = 4$ m)

in the results, with fading dips in excess of 30 dB, as well as overall increase in path loss in the case of loaded pallets, consistent with the fact that a higher percentage of links will be operating in partial line-of-sight or non-line-of-sight conditions in the latter case. There is an increase in the received maximum power level in the case of loaded pallets, which can be given by constructive interference in the line-of-sight components in the vicinity of the transceiver location for the specific scenario configuration.

The analysis of the results obtained for the different scenarios presented shows that wireless channel connectivity is strongly influenced by the type of environment

(indoor/outdoor, dense urban, dense indoor), leading to variations in path loss values as well as in delay dispersion in time domain. In summary:

- The use of higher carrier frequencies from 868 MHz to 3.5 GHz leads to lower values of received power levels, owing to increase in path loss as well as to shadowing losses, given higher-frequency dispersive material losses, with average differences up to 40 dB.
- The topomorphological characteristics as well as the network topology have a significant impact in terms of multipath components, with variations in terms of delay spread in the range of 1000 ns to over 6000 ns in the scenarios under test described. This can be extended also to variable link types given in different conditions in outdoor, indoor, and outdoor/indoor links, varying from line-of-sight, partial line-of-sight, to non-line-of-sight conditions.

The requirements of IoT massive deployments in terms of node density, cost, and area of operation favor the use of sub 6 GHz frequency bands, which can be provided by combining LPWAN, WLAN, and 5G NR FR1 wireless communication systems. Future evolutions in B5G/6G will allow to further enhance operation, by adding additional access networks for IoT, such as VLC and different options of fronthaul links mainly for aggregation purposes related to ultra-high node density environments, such as massive machine-type communications.

8.3 Conclusions

The current needs and future evolution of IoT applications require wireless communication capabilities in order to enable connectivity of increasing density of devices. In this sense, compliance of coverage/capacity requirements foresees the use of sub 6 GHz frequency spectrum in order to enable large coverage areas, mainly focused in low to moderate transmission rates for telemetry and telecontrol applications. Current solutions based on LPWAN, WLAN, and 4G/5G machine-type connections will be coupled with future B5G/6G systems, for high-density line-of-sight scenarios (such as VLC), applications with very stringent QoS requirements in terms of delay and transmission bit rate, and new node/device aggregation schemes, which can benefit for fronthaul/backhaul solutions in the extended mm-wave/THz range. The combined use of these technologies, following a cooperative heterogeneous network approach, will expand overall coverage levels as well as application domains of IoT solutions.

References

1. A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, Internet of things for smart cities. IEEE Int. Things J. **1**, 22–32 (2014). https://doi.org/10.1109/JIOT.2014.2306328

2. R. Morello, S.C. Mukhopadhyay, Z. Liu, D. Slomovitz, S.R. Samantaray, Advances on sensing technologies for smart cities and power grids: a review. IEEE Sensors J. **17**, 7596–7610 (2017). https://doi.org/10.1109/JSEN.2017.2735539
3. M.A.A. Mamun, M.R. Yuce, Sensors and systems for wearable environmental monitoring toward IoT-enabled applications: a review. IEEE Sensors J. **19**, 7771–7788 (2019). https://doi.org/10.1109/JSEN.2019.2919352
4. C. Paniagua, J. Delsing, Industrial frameworks for internet of things: a survey. IEEE Syst. J. **15**, 1149–1159 (2021). https://doi.org/10.1109/JSYST.2020.2993323
5. W. Ayoub, A.E. Samhat, F. Nouvel, M. Mroue, J. Prévotet, Internet of mobile things: overview of LoRaWAN, DASH7, and NB-IoT in LPWANs standards and supported mobility. IEEE Commun. Surveys Tutor. **21**, 1561–1581 (2019). https://doi.org/10.1109/COMST.2018.2877382
6. U. Raza, P. Kulkarni, M. Sooriyabandara, Low power wide area networks: an overview. IEEE Commun. Surveys Tutor. **19**, 855–873 (2017). https://doi.org/10.1109/COMST.2017.2652320
7. W. Jiang, B. Han, M.A. Habibi, H.D. Schotten, The road towards 6G: a comprehensive survey. IEEE Open J. Commun. Soc. **2**, 334–366 (2021). https://doi.org/10.1109/OJCOMS.2021.3057679
8. N. Tafintsev et al., Aerial access and Backhaul in mmWave B5G systems: performance dynamics and optimization. IEEE Commun. Mag. **58**, 93–99 (2020). https://doi.org/10.1109/MCOM.001.190031
9. S. Vitturi, C. Zunino, T. Sauter, Industrial communication systems and their future challenges: next-generation ethernet, IIoT, and 5G. Proc. IEEE **107**, 944–961 (2019). https://doi.org/10.1109/JPROC.2019.2913443
10. D. Baumann, F. Mager, U. Wetzker, L. Thiele, M. Zimmerling, S. Trimpe, Wireless control for smart manufacturing: recent approaches and open challenges. Proc. IEEE **109**, 441–467 (2021). https://doi.org/10.1109/JPROC.2020.3032633
11. J. Zhu, Y. Zou, B. Zheng, Physical-layer security and reliability challenges for industrial wireless sensor networks. IEEE Access **5**, 5313–5320 (2017). https://doi.org/10.1109/ACCESS.2017.2691003
12. W. Mao, Z. Zhao, Z. Chang, G. Min, W. Gao, Energy efficient industrial internet of things: overview and open issues. IEEE Trans. Ind. Inf. **17**, 7225–7237 (2021). https://doi.org/10.1109/TII.2021.3067026
13. K.C. Chen, S.C. Lin, J.H. Hsiao, C.H. Liu, A.F. Molisch, G.P. Fettweis, Wireless networked multirobot systems in smart factories. Proc. IEEE **109**, 468–494 (2021). https://doi.org/10.1109/JPROC.2020.3033753
14. Y. Ding et al., Experimental investigation of the packet loss rate of wireless industrial networks in real industrial environments. IEEE Int. Conf. Inf. Autom. 1048–1053 (2015). https://doi.org/10.1109/ICInfA.2015.7279441
15. R. Dionísio, T. Lolić, P. Torres, Electromagnetic interference analysis of industrial IoT networks: from legacy systems to 5G, in *2020 IEEE Microwave Theory and Techniques in Wireless Communications (MTTW)* (2020), pp. 41–46. https://doi.org/10.1109/MTTW51045.2020.9245057
16. X. Yin, X. Cheng, *Propagation Channel Characterization, Parameter Estimation and Modeling for Wireless Communications* (Wiley, Singapore, 2016)
17. Z. Yun, M.F. Iskander, Ray tracing for radio propagation modeling: principles and applications. IEEE Access **3**, 1089–1100 (2015). https://doi.org/10.1109/ACCESS.2015.2453991
18. P. Jörke, S. Böcker, F. Liedmann, C. Wietfeld, Urban channel models for smart city IoT-networks based on empirical measurements of LoRa-links at 433 and 868 MHz, in *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (2017), pp. 1–6. https://doi.org/10.1109/PIMRC.2017.8292708
19. M. Mohamed, M. Cheffena, F. Perez Fontan, A. Moldsvor, A dynamic channel model for indoor wireless signals: working around interference caused by moving human bodies. IEEE Antennas Propag. Mag. **60**, 82–91 (2018). https://doi.org/10.1109/MAP.2018.2796022

20. E.I. Adegoke, R.M. Edwards, W.G. Whittow, A. Bindel, Characterizing the indoor industrial channel at 3.5GHz for 5G, in *2019 Wireless Days (WD)* (2019), pp. 1–4. https://doi.org/10.1109/WD.2019.8734160
21. J. Bian, C.X. Wang , X. Gao, X. You, M. Zhang, A general 3D non-stationary wireless channel model for 5G and beyond. IEEE Trans. Wirel. Commun. **20**, 3211–3224 (2021). https://doi.org/10.1109/TWC.2020.3047973
22. T. Jiang et al., 3GPP standardized 5G channel model for IIoT scenarios: a survey. IEEE Int. Things J. **8**, 8799–8815 (2021). https://doi.org/10.1109/JIOT.2020.3048992
23. Q. Qi, X. Chen, C. Zhong, Z. Zhang, Integrated sensing, computation and communication in B5G cellular internet of things. IEEE Trans. Wirel. Commun. **20**, 332–344 (2021). https://doi.org/10.1109/TWC.2020.3024787
24. H. Zhuang, J. Chen, R. Gilimyanov, Hierarchical energy optimization with more realistic power consumption and interference models for ultra-dense networks. IEEE Trans. Wirel. Commun. **19**, 4507–4518 (2020). https://doi.org/10.1109/TWC.2020.2984504
25. H. Chergui, L. Blanco, C. Verikoukis, Statistical federated learning for beyond 5G SLA-constrained RAN slicing. IEEE Trans. Wirel. Commun. (2021). https://doi.org/10.1109/TWC.2021.3109377
26. D. Mishra, N.R. Zema, E. Natalizio, A high-end IoT devices framework to foster beyond-connectivity capabilities in 5G/B5G architecture. IEEE Commun. Mag. **59**, 55–61 (2021). https://doi.org/10.1109/MCOM.001.2000504
27. L. Azpilicueta, M. Rawat, K. Rawat, F.M. Ghannouchi, F. Falcone, A ray launching-neural network approach for radio wave propagation analysis in complex indoor environments. IEEE Trans. Antenn Propag. **62**, 2777–2786 (2014). https://doi.org/10.1109/TAP.2014.2308518
28. L. Azpilicueta, F. Falcone, R. Janaswamy, A hybrid ray launching-diffusion equation approach for propagation prediction in complex indoor environments. IEEE Antenn. Wirel. Propag. Lett. **16**, 214–217 (2017). https://doi.org/10.1109/LAWP.2016.2570126
29. L. Azpilicueta, F. Falcone, R. Janaswamy, Hybrid computational techniques: electromagnetic propagation analysis in complex indoor environments. IEEE Antenn. Propag. Mag. **61**, 20–30 (2019). https://doi.org/10.1109/MAP.2019.2943297
30. F. Casino, L. Azpilicueta, P. Lopez-Iturri, E. Aguirre, F. Falcone, A. Solanas, Optimized wireless channel characterization in large complex environments by hybrid ray launching collaborative filtering approach. IEEE Antenn. Wirel. Propag. Lett. **16**, 780–783 (2017). https://doi.org/10.1109/LAWP.2016.2604021
31. L. Azpilicueta, M. Rawat, K. Rawat, F. Ghannouchi, F. Falcone, Convergence analysis in deterministic 3D ray launching radio channel estimation in complex environments. ACES J. **29**, 256–271 (2014)
32. J. Karedal, S. Wyne, P. Almers, F. Tufvesson, A.F. Molisch, A measurement-based statistical model for industrial ultra-wideband channels. IEEE Trans. Wirel. Commun. **6**, 3028–3037 (2007)
33. R.J. Luebbers, A Heuristic UTD slope diffraction coefficient for rough lossy wedges. IEEE Trans. Antenn. Propag. **37**, 206–211 (1989). https://doi.org/10.1109/8.18707
34. R.J. Luebbers, Comparison of lossy wedge diffraction coefficients with application to mixed path propagation loss prediction. IEEE Trans. Antenn. Propag. **36**, 1031–1034 (1998). https://doi.org/10.1109/8.7210

Chapter 9
Feasibility of LDM to Serve User-IoT Pairs in the Future Wireless Network

Md Shantanu Islam, Raouf Abozariba, A. Taufiq Asyhari, Mohammad Patwary, and Mohammad Abdul Matin

9.1 Introduction

The future wireless system's development is focused on achieving higher capacity and flexibility than the existing systems. The future network should provide seamless services to new types of services such as the internet of things (IoT), enhanced mobile broadband (eMBB) communications, and tactile internet. These services are setting the standards for the long-term evolution of fifth generation (5G) and beyond 5G (B5G) [1]. Multiple technologies have the potential to achieve this seamless connectivity with increased user capacity. Non-orthogonal multiple access (NOMA) is one such potential technology [2, 3].

In this chapter we are exploring the application and feasibility of LDM in future wireless communication to achieve higher network capacity while delivering the required standard of user and IoT communications. We have developed a reliable downlink transmission framework and usage model that can take advantage of the different characteristics of the LDM layers. LDM is a layer division or power-based NOMA which can be configured with diverse power levels (layers) to provide different services using its upper and lower layer [4]. LDM can serve two different services using a single traffic channel (OFDM channel as will be presented later) as both layers can use the available time and frequency slot of

M. S. Islam (✉) · R. Abozariba · A. T. Asyhari
School of Computing and Digital Technology, Birmingham City University, Birmingham, UK
e-mail: mdshantanu.islam@mail.bcu.ac.uk; raouf.abozariba@bcu.ac.uk; taufiq.asyhari@bcu.ac.uk

M. Patwary
Department of Computer Science, University of Wolverhampton, Wolverhampton, UK
e-mail: patwary@wlv.ac.uk

M. A. Matin
Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_9

the traffic channel simultaneously. This characteristic makes LDM a more spectral efficient system as it can use the available channel capacity for both layers with intelligent layer configuration. One such simple structure is used in the integration of LDM into advanced television systems committee (ATSC 3.0) PHY layer baseline technology [5]. In this case, a robust configuration is implemented in the upper (core) layer (UL), oriented to portable and mobile receivers. On the other hand, in the lower (enhance) layer (LL), a high-capacity configuration is chosen to deliver high data rate services, such as ultrahigh-definition television (UHDTV) or multiple high-definition television (HDTV) services, to fixed receivers. Motivated from the successful integration of LDM in ATSC 3.0, we have investigated the possibilities of convergence between the user and IoT devices based on their difference in required data rate and network condition. In our model, the UL is used to serve the IoT communications, while LL is used to serve the users (e.g., mobile handsets). As the model is developed for future wireless needs, it has to perform with user mobility. We have aimed to implement LDM UL and LL for the mobile user, which is a challenge as in ATSC 3.0, and static users are considered LL receivers.

We have developed an OFDM physical layer framework to adopt LDM into our transmission model. In principle, heterogeneity for the 5G network is dictated by QoE/QoS of different applications—in terms of latency, privacy, data rate, accuracy, and robustness requirements. LDM upper and lower layers can address these heterogeneous requirements as UL offers higher accuracy and robustness with lower latency; in contrast, LL can ensure better privacy and a higher data rate. In this chapter, we are looking into a downlink transmission scenario for IoT and users. Moreover, the impact of mobility on the formation and termination of IoT pairs will be explained. A way to calculate LDM link sustainability for combining different mobility models of both receivers is developed and explained. This chapter aims to give readers an insight into the LDM's applicability in future wireless communication.

The remainder of the chapter is structured as follows: key works in LDM for broadcasting and 5G wireless communication are discussed in Sect. 9.2. The use case scenario of this transmission system is described in Sect. 9.3. The system model is described for the downlink communication scenario in Sect. 9.4. An analytical framework for LDM performance evaluation is presented in Sect. 9.5. The device mobility model and its impact on the LDM pair is discussed in Sect. 9.6. Section 9.7 describes our findings and relevant analysis. Moreover, the outcome is summarized in the final Sect. 9.8.

9.2 Literature Review

LDM is a power-based NOMA technology that was introduced in cloud transmission (a flexible multilayer system that uses spectrum overlay technology to deliver multiple program streams simultaneously) [6] in 2012. Later, LDM was accepted for use in the physical layer design of advanced television systems committee

(ATSC) 3.0 due to its higher degree of flexibility and performance advantages over existing orthogonal multiple access (OMA) techniques [7]. LDM can combine various services in a single radio frequency (RF) channel to support multiple user applications using the same traffic channel. Since adapting LDM into ATSC 3.0 physical layer design, much work has been done on finding the LDM's performance for the downlink broadcast transmission. Reference [8], for example, investigated this performance trade-off between UL and LL where capacity and coverage performance of LL at the cost of that of UL is shown for ATSC 3.0. Their work shows a better performance of LDM compared with time-division multiplexing (TDM) and frequency-division multiplexing (FDM). Reference [9] proposed a multiple physical layer pipe (M-PLP) configuration based on multilayer LDM and performed a capacity analysis of this configuration to determine the lower capacity bound (approximately 1 Mbps) of UL. In [10] the authors showed that LDM UL provides higher channel capacity than TDM/FDM at a low signal-to-noise ratio (SNR), and LL can do the same at high SNR conditions. Due to increased capacity gain over TDM and FDM, LDM is gaining interest in 5G deployment. Many have investigated the possibility of using LDM in point to multipoint (P2MP) and broadcasting transmission in 5G wireless communication. Both [11, 12] investigated the capacity improvement of LDM over OMA techniques for providing multimedia services in 5G and found that LDM offers higher channel capacity. On the other hand, in [13] higher network throughput is achieved by using LDM for a unicast-broadcast convergence. These works show that LDM can serve heterogeneous devices and increase network capacity. Most of the works on LDM are focused on exploiting its advantages for broadcast multimedia transmission; these works motivated us to look for other use cases where LDM can be successful. In our previous work [14], we proposed a system model that uses LDM to take advantage of the heterogeneity requirements of user and IoT devices. Our analysis showed that LDM increased capacity is achieved only at higher SNR for both UL and LL. Due to this higher SNR requirement, the impact of the rapid changes in the channel condition arising from the mobility needs to be determined. Therefore, mobility management for LDM pairs is crucial for the successful broader implementation of LDM in the future wireless network.

The following section describes the usage scenario where such user IoT pairing would be needed. Both urban and rural scenarios showing the usage model are presented. The focus is mainly on the IoT usages as these are the expected new inclusion in future wireless communication.

9.3 Use Case Scenario of LDM in Future Wireless Network

Our system model focuses on heavy downlink communication for IoT and users [14]. It is widely assumed within the predecessor communication technologies that IoT communication does not require much downlink data. However, the future usage models are expected to demand more uplink and downlink data transmission

for both users and IoT devices. This section aims to describe a few such future transmission scenarios where our proposed system model can be of use to support such transmission. We have considered both urban and rural scenarios for such possible heavy downlink transmission dependency for mMTC communication. These possible usages scenarios justify our downlink heavy transmission model for IoT communication.

9.3.1 Urban Use Case Scenario for IoT Downlink Communication

The urban areas are expected to be high device density areas. Figure 9.1 depicts a transmission scenario for the downlink communication requirements for autonomous vehicle systems and cloud-controlled drones in the future city context. Small cell or picocells are expected to adopt in urban cellular infrastructure. Hence, a transmitter is expected to take advantage of existing urban structures such as radioheads installed on the lamppost, as shown in Fig. 9.1 [15]. Alongside the massive number of IoT devices attached to users, such as smartwatches and other devices, cloud control drones and autonomous vehicle systems will add to the total number of IoT users in future cities. The drones in Fig. 9.1 are controlled from a central cloud location, requiring constant control information from the cloud depicting high downlink communication dependency. Autonomous vehicles also have a heavy downlink dependency as they also require continuous traffic and control information from the cloud to navigate the urban streets[16]. Most of

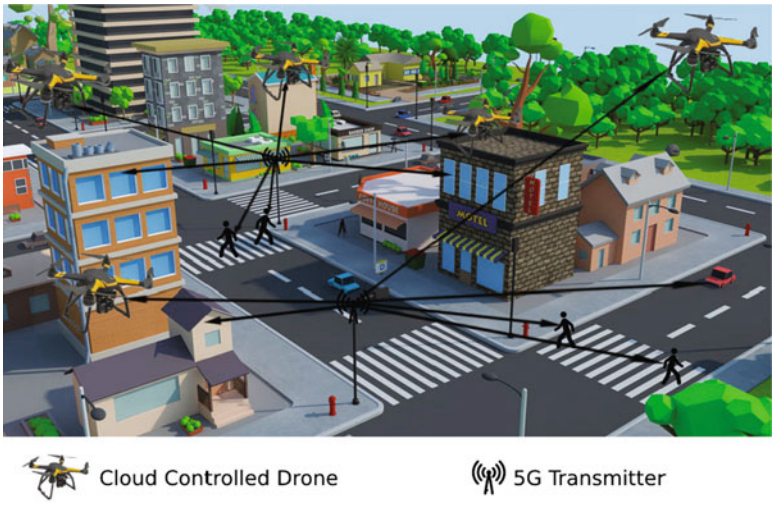


Fig. 9.1 Downlink transmission model for IoT devices in urban area

these communications are expected to be real-time communication; hence, they require ultrareliable low-latency communication[17]. As the number of users and IoT devices are expected to multiply significantly, future wireless networks need to achieve the capability to manage such a high number of connections. Due to the similarities of payload information in most of the associated machine-type devices, any IoT device can be paired with a suitable user device. In this configuration, the UL layer is suitable for all IoT devices due to its robustness and lower latency.

9.3.2 Rural Use Case Scenario for IoT Downlink Communication

The application of IoT in the rural scenario is different from that of the urban scenario depicted in the earlier section. Many future farming and agriculture systems will be controlled and monitored by IoT sensors, and cloud-controlled drones can be useful for surveillance, as shown in Fig. 9.2. The drones need real-time control information as described in urban scenarios; sensors also have significant use for future rural applications, for example, to monitor water flow and assess stored food conditions and different agriculture-related monitoring. These devices can serve multiple services based on the requirement and are expected to be self-reconfigurable upon receiving control information from the clouds. Subsequently,

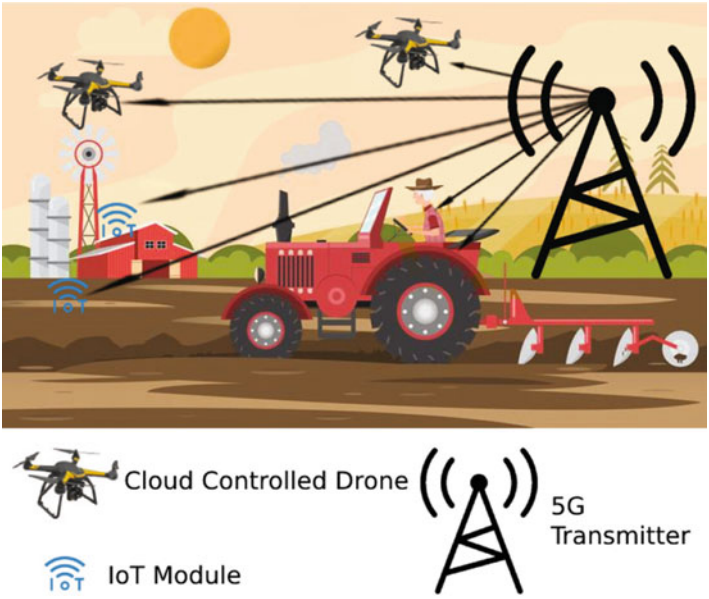


Fig. 9.2 Downlink transmission model for IoT devices in rural area

these devices require a heavy downlink transmission. One major change in the network configuration of rural areas is the use of cells with large coverage areas. This large cellular coverage area will make the user IoT pair more sustainable even when both receiving devices are mobile. However, the design of LL will be more critical to the extensive coverage area and hence will be more susceptible to channel conditions. However, the slow nature of changes in rural areas makes for better channel prediction. Hence, the UL and LL can serve these IoT and user combinations in such scenarios.

A physical layer transceiver framework that adopts LDM within a standard OFDM model is described next. The model is for a two-layer LDM transmission where UL serves the IoT devices, and the LL is used to serve the user.

9.4 Transceiver Framework Adopting LDM

We have proposed the integration of LDM within an OFDM system as shown in Fig. 9.3. We are focusing on benchmarking the performance of LDM layers within an OFDM framework and, for that purpose, have not included any error-correcting code in our proposed model. Like any communication model, the system is described as per transmitter, channel, and receiver.

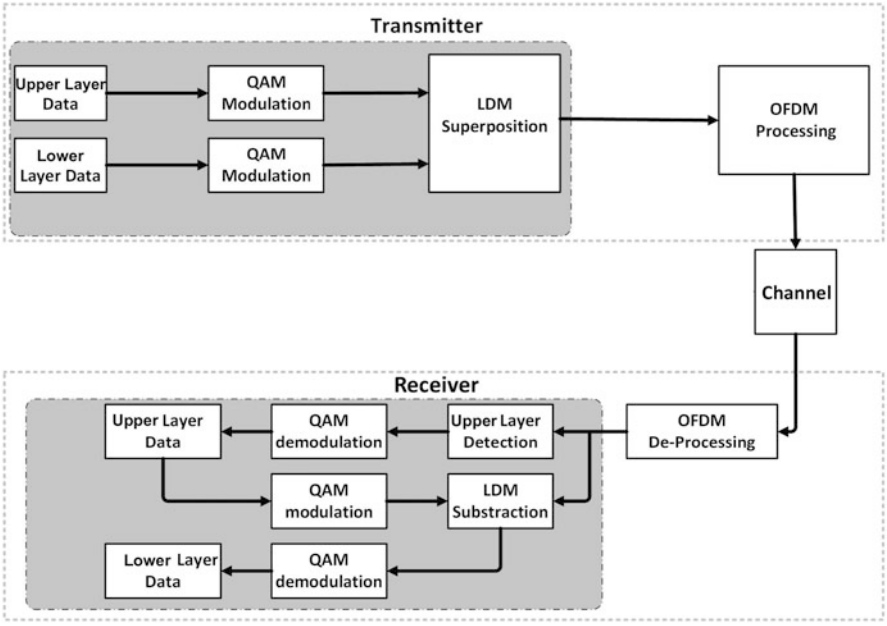


Fig. 9.3 OFDM transmission framework with LDM adaptation

9.4.1 Transmitter Framework with LDM Superposition

The processing of both UL and LL data is done in parallel at the transmitter end, as shown in Fig. 9.3. Both layers can have different transmission bit rate with different quadrature amplitude modulation (QAM) modulation scale M. The UL will have a lower bit rate hence a lower M, and the opposite is assumed for the LL. However, M is chosen in a way that will result in an equal number of UL and LL QAM symbols. LDM superposition is done in the next step, resulting in the same number of LDM symbols. The power of the LL symbol is reduced during the superposition process to have a smaller power portion of the total transmit power. The total transmission power is the same as any single-core transmission. Equation (9.1) represents the LDM superposition where $\mathbf{X}(k)$ represents the LDM symbols of kth subcarrier, g is the power ratio between layers, and $\mathbf{X_{UL}}(k)$ and $\mathbf{X_{LL}}(k)$ represent UL and LL symbols of kth subcarrier, respectively. As a result, each LDM symbol contains a UL layer symbol and a LL symbol where the total number of LDM symbols is the same as that of an OFDM system in a subcarrier.

$$\mathbf{X}(k) = \mathbf{X_{UL}}(k) + g\,\mathbf{X_{LL}}(k). \tag{9.1}$$

9.4.2 Channel Model

A single traffic channel serves user-IoT paired devices with LL and UL layer data. In this analysis, we have considered the AWGN channel model. The IoT device and user come under various mobility models, and hence the channel condition will vary for both of them. However, the IoT devices are assumed to be in poorer channel conditions due to their versatile locations and power restrictions, making them ideal recipients of UL data. The user device is more sensitive to channel conditions, but it can enjoy a higher data rate due to the properties of LL. We assume perfect channel estimation and perfect reception of control signaling at the receivers, and receivers have all the necessary information needed to detect the signal.

9.4.3 Receiver Framework for LDM Detection

The LDM signal at the receiver can be expressed by

$$\mathbf{Y}(k) = \mathbf{X_{UL}}(k).\,\mathbf{H}(k) + g\,\mathbf{X_{LL}}(k).\,\mathbf{H}(k) + \mathbf{N}(k), \tag{9.2}$$

where $\mathbf{Y}(k)$ is the received signal of kth sub-channel, $\mathbf{H}(k)$ is the channel matrix, and $\mathbf{N}(k)$ is the added noise. The dimensions of all the parameters used in (9.2) are the same as the number of OFDM symbols of kth subcarrier, which we assumed

to be 64 in our simulation. The UL detection is done in a simple OFDM detection process where the LL signal from (9.2) is treated as added interference. Equation (9.3) shows the total noise and interference for the UL detection. UL detection does not require complex computing; hence, this is suitable for low-power IoT devices. The signal at the receiver will have the originally transmitted signal with added noise as well as channel effects and can be expressed as

$$\mathbf{N_{UL}}(k) = g\,\mathbf{X_{LL}}(k)\,.\,\mathbf{H}(k) + \mathbf{N}(k). \tag{9.3}$$

The detection of LL data is done in the next phase, where the detected UL data is processed the same way as it is done at the transmitter. The reconstructed UL data was then subtracted from the original received signal. Then the remaining signal is amplified and detected in that order. Equation (9.4) shows the subtracted signal for the LL detection.

$$\mathbf{Y_{LL}}(k) = \mathbf{X_{UL}}(k).\,\mathbf{H}(k) + g\,\mathbf{X_{LL}}(k)\,.\,\mathbf{H}(k) + \mathbf{N}(k) - \mathbf{X_{UL\ re}}(k). \tag{9.4}$$

Assuming a perfect UL detection with perfect channel estimation, we can derive LL signal as

$$\mathbf{Y_{LL}}(k) = g\,\mathbf{X_{LL}}(k)\,.\,\mathbf{H}(k) + \mathbf{N}(k). \tag{9.5}$$

The signal from (9.5) is amplified with a factor of $1/g$ before the detection of LL data. As we cannot separate noise during this phase of detection, the noise is also amplified by the same ratio, and it makes the LL data detection more prone to channel noise power level. Successful UL detection is essential in this transmission framework for successful LL detection, so the UL needs to be reliable and robust. Moreover, the LL data detection process is more computationally intensive, so user devices require higher computation capability.

The analytical model to evaluate the UL bit error rate (BER) is developed empirically and described in the following section. The channel capacity distribution and the maximum UL channel capacity are also derived.

9.5 Theoretical Evaluation

We analyzed the system models described in the last section to evaluate the performance of LDM in our proposed usage models. To justify the results of this analysis and have a reference point for future work, we have developed an empirical equation to examine the performance bound of the LDM UL. Due to the incorporation of LL data, the maximum UL channel capacity is bounded by the g. This capacity bound is also defined in this section.

9.5.1 Bit Error Rate of LDM Upper Layer

We considered the performance of LDM in an uncoded OFDM system in the AWGN channel. We have obtained the equation for symbol error rate (SER) of an uncoded QAM OFDM system from [18] as

$$SER_k^{AWGN} = 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3\,\rho_k}{M-1}}\right) - 4\left(1 - \frac{1}{\sqrt{M}}\right)^2 Q\left(\sqrt{\frac{3\,\rho_k}{M-1}}\right)^2, \tag{9.6}$$

where

$$Q(x) = \frac{1}{pi}\int_0^{\frac{pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) dx \tag{9.7}$$

and ρ_k is the SNR of the kth symbol.

By assuming Gray coding for QAM constellation, which gives a single bit error for each symbol error, we get the relation between BER and SER as BER = SER$/M$. An analytical model for the BER calculation of the UL is developed from the equations above. For evaluation, no intersymbol interference is assumed. LL has lower power than the UL; therefore, the value of g is always negative in dB. We apply this setup for all the different QAMs used. We identified that the UL data rate and g follow the relation in an uncoded OFDM system as

$$g = -4\,M_c \tag{9.8}$$

where M_c is the QAM order for the UL. Now, the UL SNR is calculated from channel SNR as

$$\rho_{kcl} = \frac{2\,\rho_k}{-g}. \tag{9.9}$$

Using the values of UL SNR from (9.9) in (9.6), we calculate the SER of the UL data as

$$SER_k^{AWGN} = 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{6\,\rho_k}{g(1-M)}}\right) - 4\left(1 - \frac{1}{\sqrt{M}}\right)^2 Q\left(\sqrt{\frac{6\,\rho_k}{g(1-M)}}\right)^2. \tag{9.10}$$

9.5.2 Channel Capacity Distribution

The channel capacity of an AWGN channel can be written as

$$C = \log_2\left(1 + \frac{P_s}{P_n}\right), \tag{9.11}$$

where P_s is the signal power and P_n is the noise power. This shows the dependency between channel SNR and capacity and can be used to calculate the channel capacity for LDM layers as [10]

$$C_{UL} = \log_2\left(1 + \frac{P_{UL}}{P_{LL} + P_n}\right) \tag{9.12}$$

and

$$C_{LL} = \log_2\left(1 + \frac{P_{LL}}{P_n}\right). \tag{9.13}$$

As the power ratio g is known, we can evaluate the UL's maximum system capacity with a fixed value of g from

$$C_{UL} = \log_2\left(1 + \frac{1}{g}\right). \tag{9.14}$$

The following section considers a model where both user and IoT receivers are mobile. A detailed analysis of LDM pair sustainability with mobility is presented in the next section.

9.6 Mobility Model

We are considering both user and IoT mobility in this analysis. In ATSC 3.0, a 4k ultrahigh-definition TV or multiple enhanced HDTV services are transmitted using LL to a fixed receiver with advanced antennas[19]. The proposed model is developed to test the UL and LL layers' performance with various mobility models for users and IoT devices. Moreover, the future wireless network is expected to use advanced transmission technology such as massive MIMO to transmit configurable signals with custom power and beamwidth management. These technologies will allow the system to manage interference and frequency reuse more efficiently. However, these techniques will lower the coverage area with narrower beamwidth and lower transmission power. This smaller coverage area will be more challenging to sustain the mobile LDM pair. Ideally, we want our LDM pair to be within the coverage area for the duration of the entire data transmission to minimize the

effort needed to form new IoT-user pair. Therefore, there is a trade-off between the transmission beam width and power and the optimum coverage area for a different receiver mobility group.

9.6.1 IoT and User Mobility

As shown in Fig. 9.4, the IoT devices are assumed to be distributed randomly over the entire transmission, and the initial beam is directed at the user location; hence, it is at a zero degree angle with the transmission beam. The IoT devices can start with any random angle θ_{2i} with transmission direction. Figure 9.4 shows the mobility of IoT and user devices from an initial position to a final position. The base station (BS) is assumed to be located at position (x_b, y_b). The initial IoT and user location is assumed to be at (x_1, y_1). Then the initial distance L_{initial} can be calculated using

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{9.15}$$

The transmission is assumed to be directed at the user's initial position. Therefore, the user is initially located on the x-axis, as shown in Fig. 9.4. The distance between the user's initial position and the base station is the same as the x values of the user's initial coordinates. The initial angle between IoT device and BS θ_{2i} can be calculated as

$$\theta_{2i} = \cos^{-1}\left(\frac{l_{\text{initial}}^2 + x_1^2 - y_1^2}{2 l_{\text{initial}}\, x_1}\right). \tag{9.16}$$

If the device moves at a random speed v for a time t to reach its final position, then the distance between IoT's initial and final position d can be written as

$$d = v\,t. \tag{9.17}$$

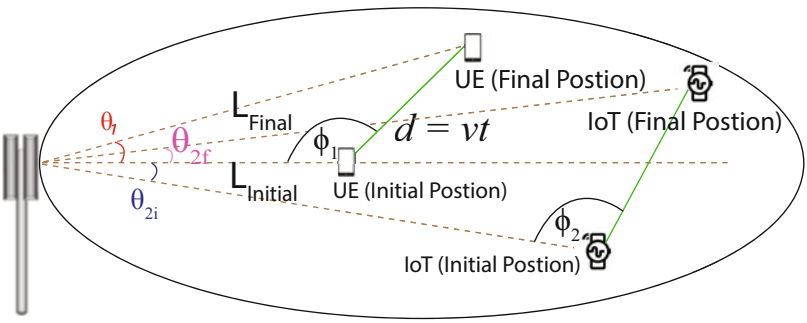


Fig. 9.4 IoT and user mobility in LDM pairing

The IoT device is assumed to move at a random angle ϕ to the (x_2, t_2) after time t. Then the final position can be calculated using

$$x_2 = x_1 + d\cos(\phi), \tag{9.18}$$

and

$$y_2 = y_1 + d\sin(\phi). \tag{9.19}$$

The coordinates of BS and IoT are now known. The distance between the initial and final location of IoT and with BS can be measured using (9.15). At the final position, the angle between BS and IoT is taken as θ_2 as shown in Fig. 9.4 which can be calculated as

$$\theta_{2f} = \cos^{-1}\left(\frac{l_{\text{final}}^2 + x_2^2 - y^2}{2l_{\text{final}}\, x_2}\right). \tag{9.20}$$

Similarly for user device, the final angle θ_f can be calculated using the following values of user's initial and final position.

$$\theta_f = \cos^{-1}\left(\frac{l_{\text{initial}}^2 + l_{\text{final}}^2 - d^2}{2l_{\text{initial}}\, l_{\text{final}}}\right). \tag{9.21}$$

Assume t_p as the power transmitted from the BS with n_p being the white noise power and h being the channel fading. Path loss L_p can be calculated as [20]

$$L_p = 32.4 + 20\log_{10}(f_c) + 20\log_{10}(r), \tag{9.22}$$

where f_c is the carrier frequency in GHz and r is the distance between transmitter and receiver in meters. Then finally, we can calculate the SNR at IoT devices from

$$\rho = \frac{t_p\, h\cos\theta}{n_p\, L_p}, \tag{9.23}$$

where h is the time selective fading coefficient. For L number of transmission paths, h depends on the career frequency f_c and Doppler frequency f_d. The following equation provides the relation between h and these above parameters as a function of time.

$$h(t) = \sum_{i=0}^{L-1} a_i\, e^{-j\,2\pi\, f_c\, \tau_i}\, e^{j\,2\pi\, f_d t}. \tag{9.24}$$

From this relation, we can calculate h from the carrier frequency f_c and the speed of the devices, which is the cause of the Doppler frequency f_d. We assume the SNR

of IoT at the initial position as ρ_1. IoT data is transmitted using the UL, and LDM configuration is developed based on ρ_1 and user network conditions. As the IoT and user device moves from the initial position, the SNR condition will keep changing. The LDM pair needs to break when the minimum SNR required for the IoT devices or the user is greater than the channel SNR of either devices.

9.6.2 The Range of IoT Device's Initial Position

When forming LDM pair, both user and IoT devices need to be within the coverage area. As the transmission is directed toward the user, we need to sort out the boundary condition for the IoT device that will keep it within the transmission angle. To simplify, we are assuming the base station location as (0,0); we get the initial distance between BS and IoT as

$$\begin{aligned} L_{initial} &= \sqrt{(x_b - x_1)^2 + (y_b - y_1)^2} \\ &= \sqrt{x_1^2 + y_1^2}. \end{aligned} \tag{9.25}$$

Using (9.16) and (9.25), we can derive the following relation

$$\begin{aligned} \cos(\theta_1) &= \frac{L_{initial}^2 + x_1^2 - y_1^2}{2\, L_{initial}\, x_1} \\ &= \frac{x_1^2 + y_1^2 + x_1^2 - y_1^2}{2\, L_{initial}\, x_1} \\ &= \frac{2\, x_1^2}{2\, L_{initial}\, x_1} \\ &= \frac{x_1}{L_{initial}}. \end{aligned} \tag{9.26}$$

For the IoT device's initial position to be within the transmission area, the maximum value of θ_1 can be θ_{max} = Beamwidth/2. For this θ_{max} angle, we can calculate the maximum value of the initial y position from a given x position using (9.26) and derive the following relation.

$$\begin{aligned} \frac{\cos(\theta_1)}{x_1} &= \frac{1}{L_{initial}} \\ L_{initial} &= \frac{x_1}{\cos(\theta_1)} \\ x_1^2 + y_1^2 &= \frac{x_1^2}{\cos^2(\theta_1)} \end{aligned}$$

$$y_1^2 = \frac{x_1^2}{\cos^2(\theta_1)} - x_1^2$$

$$y_1 = \sqrt{\frac{x_1^2}{\cos^2(\theta_1)} - x_1^2}. \tag{9.27}$$

Equation (9.27) gives the maximum limit on the initial y position for IoT device for any given x value and transmission beamwidth.

9.6.3 Mobility Model

In the proposed scenario, we have considered IoT devices attached to human usages, such as smartwatches, car sensors, and similar devices. So the mobility model of these devices is similar to the human mobility model. We have considered the following three mobility models for IoT movement.

9.6.3.1 Random Waypoint Mobility Model

This mobility model represents static to downward movement in urban and rural areas. This movement model applies to nonmotorized movements such as walking, running, and cycling. The speed of the users in this model is assumed to be within the range of 0–10 kph, and the angle is between 0 and 360° The movement pattern is random both in terms of speed and direction as both can be changed randomly at any time [21]. For simplicity, we have assumed that the devices can alter their speed after every t time interval and direction after every 100 m.

9.6.3.2 Manhattan Mobility Model

This model refers to the urban street movement for motorized vehicles [21]. We assume a grid-like design for all urban roads resulting in a discrete perpendicular or parallel change in direction. We assume the block lengths to be 200 m, which means the user can change direction every 200 m with a speed range between 10 and 40 kph.

9.6.3.3 Freeway Mobility Model

This mobility model is for unidirectional mobility. In this model, the device keeps moving in a constant direction for the whole scenario with variable speed. We have taken this model to represent vehicle movement on the motorway. This model

assumed the vehicle speed to be between 40 and 100 kph. Moreover, as the vehicle does not take turns frequently and moves freely via a single road at a higher speed, we can assume their movement to be unidirectional for a more petite time frame. This model is a simple linear model where the receiver's position can be predicted.

9.7 Results and Analysis

We are presenting two sets of analyses to evaluate our transmission framework. In the first set, we test the performance of LDM transmission without any additional channel coding. The results show different channel SNR requirements for UL and LL data detection. Moreover, our analytical model is also verified in this analysis. In the second set, we test the link sustainability of LDM pair with mobility. All our previously discussed mobility models are considered in this simulation, and the results show the feasibility of LDM for mobile receivers.

9.7.1 The Performance of LDM in an OFDM Framework

In this simulation, two independent data sets are transmitted using LDM within an OFDM framework using an AWGN channel model. The SNR values presented in the results refer to the overall channel SNR for the OFDM signal. Table 9.1 presents the parameter used in these simulations. The analytical model in (9.10) is also evaluated alongside the simulated values. The distribution of channel capacity between UL and LL is also given, which provides a clear indication of the applicability of both layers for IoT devices and users, respectively.

The performance bound of the UL layer presented in the earlier section is evaluated in Fig. 9.5. The BER performance follows a waterfall curve as the BER decreases with channel SNR. The SNR does not calculate the additional interference to UL data from LL data. We have tested the UL performance for different data rates, which is varied by QAM modulation index M (16, 32, and 64), and the power ratio g is selected as per (9.8). From the figure, we see overlapping BER performance for analytical and simulation results, which shows the correctness of our developed performance bound. This equation can be used to evaluate future works on LDM

Table 9.1 OFDM parameters used for LDM simulation

Parameter	Values
No. of carriers	64
Single frame size	96 bits
Total no of frame	1000
No. of pilot bits	4
Cyclic extension	16 bits

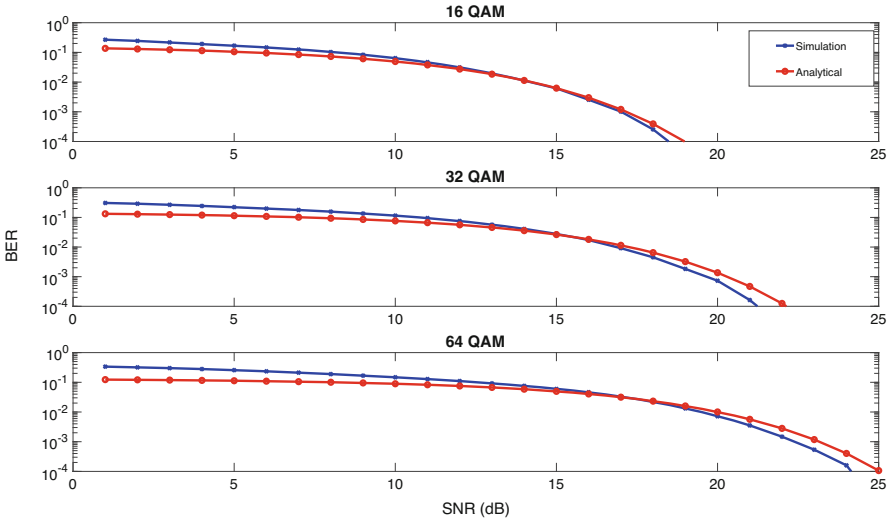


Fig. 9.5 Performance comparison of LDM UL between simulation and analytical results for 16, 32, and 64 QAM modulation

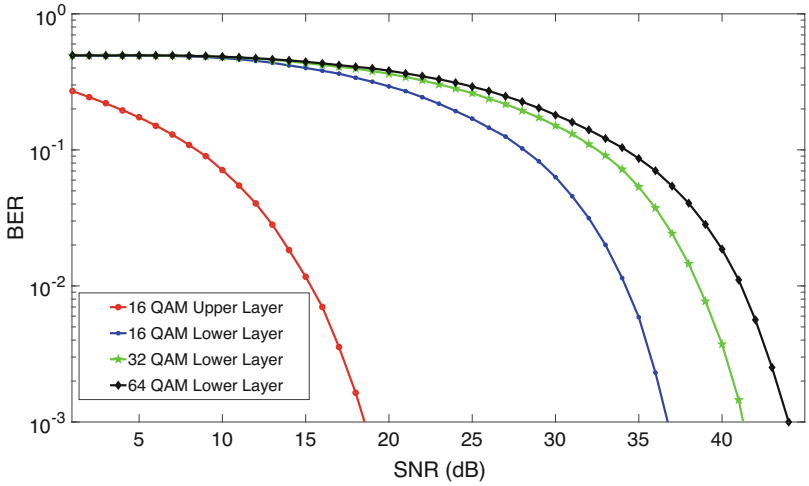


Fig. 9.6 BER performance of LDM LL transmission for 16, 32, and 64 QAM constellation with a fixed UL data rate

and OFDM. Another noticeable characteristic from the results is the increment of SNR values for similar BER performance as the data rate increases.

Figure 9.6 compares the performance of UL and LL and evaluates their performance with our proposed IoT-user LDM usage models. The power ratio g is set fixed at -10 dB in this analysis. We simulated the system with different data rate

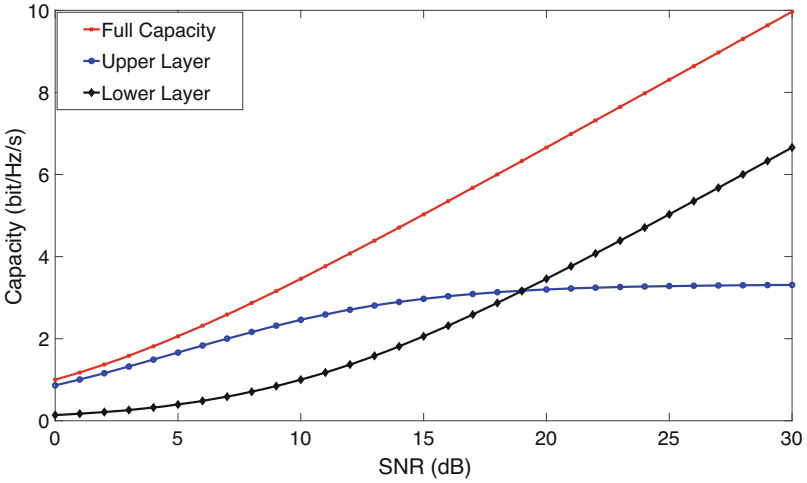


Fig. 9.7 Distribution of channel capacity between LDM UL and LL for fixed power ratio g

combinations for the UL and LL. The UL data are fixed at 192 bits (16 QAM) per frame, while the LL data ranges from 192 bits (16 QAM), 240 bits (32 QAM), to 288 bits (64 QAM) per frame. The UL performance is unchanged with the different data rates of the LL. Moreover, we obtain a bit error rate of 10^{-4} at around 22 dB. In [22], the authors investigated the UL performance for QPSK with a strong low-density parity-check (LDPC) coding of 4/15. They achieved the same 10^{-4} bit error rate at 7 dB. They also attained a similar LL performance for 64 QAM at 20 dB. In another work, [23] used BPSK for the UL with 1/8 Turbo coding for error correction and QPSK for LL with 1/2 turbo coding and achieved a similar performance at 1 and 15 dB, respectively. We get a similar performance of LL around 40 dB channel SNR for uncoded OFDM system. Our model requires higher SNR due to lack of error correction ability as we were focused on finding the performance of LDM itself. However, we also have around 15 dB higher SNR requirement for LL than UL, which is similar to the results found in [23]. Similar to the earlier results, higher SNR is needed to achieve similar BER performance with a higher data rate. Figure 9.6 shows the channel condition required for different receivers. The receiver of the UL can be in the worse transmission area with a poor SNR, whereas the receiver of the LL needs to be in a good coverage area for successful detection. The BER performance of UL at lower SNR works well for small IoT devices as they will be distributed among different places with varying channel conditions. On the other hand, LLs need higher SNR values, as seen in Fig. 9.6, which is more suitable for users due to their better channel condition and data requirement.

Channel capacity distribution between the LDM data layers is explained using (9.11), (9.12), and (9.13). Figure 9.7 shows the total channel capacity distribution between the LDM layers. The figure shows that the capacity distribution is lossless and nonlinear. At lower SNR, the UL capacity is higher than that of the LL.

However, it gets saturated with an increase in SNR, which can be calculated using (9.14) and shows the maximum UL capacity in any given LDM configuration. This distribution works well for IoT devices in our usage model as the devices are assumed to be in different locations, which can cause bad channel conditions for some devices. Due to the robustness of the UL against the channel conditions, it can be used to serve all IoT devices. Moreover, the UL capacity is lower, which also fits with the IoT devices data requirements we assumed in our usage model. On the other hand, the LL has a low capacity at lower SNR. The capacity of LL increases significantly as the SNR values improve. LL has most of the available channel capacity in good channel conditions.

9.7.2 Receiver Mobility

We have simulated the user and IoT mobility based on the earlier mobility models. The LDM pair performance is evaluated for link sustainability for a different mobility model and transmission beamwidth combination. In the first case, the user is assumed to be static while the IoT device is mobile. The x position of the IoT devices is between 0 and 500 m and is randomly chosen in each iteration. The range of y position is calculated using the known transmission beamwidth, x and (9.27). In the second case, similar simulation is done for the user movement, assuming that the transmission angle is set based on the user's initial position. Moreover, we simultaneously moved both IoT and the user to determine the sustainable link time in the final setup. We have combined the random waypoint and Manhattan model to test the link sustainability of such mobility combination. These comparisons will help the system form LDM pairs more efficiently based on the mobility group of the receivers. Both models represent urban areas, whereas the freeway mobility model is for separate geographical areas.

In both Figs. 9.8 and 9.9, the IoT devices are moving according to the random waypoint mobility model while we change the mobility model for users. In Fig. 9.8, both IoT and user devices are in the same mobility group. User mobility has less impact on pair sustainability as the transmission is directed toward the user. The IoT devices go out of the transmission area faster as there are cases when the IoT devices initial position can be at the network edge, which will break the LDM pair more quickly. The combined results follow a similar trend as IoT mobility, and as expected, the combined mobility offers a shorter sustainable link time. In Fig. 9.9, the user is on a higher mobility model, hence performs worse than IoT. We see a similar pattern in this case as well, where the combined mobility serves less than the user mobility. From these results, this is clear that the most efficient pairing would be IoT devices and users in the same mobility group; otherwise, the system will need to find a new device for pairing as one of the devices goes out of range.

In both Figs. 9.10 and 9.11, the Manhattan mobility model is used for IoT mobility while we used different mobility models for user mobility. In Fig. 9.10, again, just as in the case of the random waypoint mobility model, the user stays

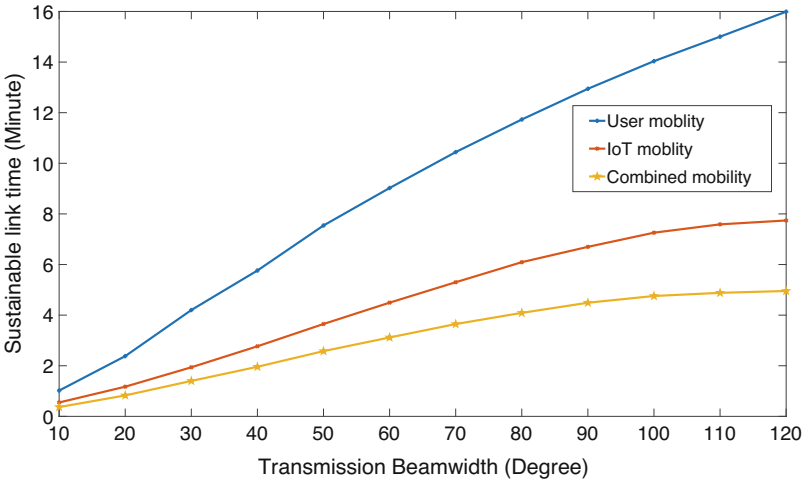


Fig. 9.8 Impact of the LDM pair mobility on the sustainable link time for random waypoint mobility model

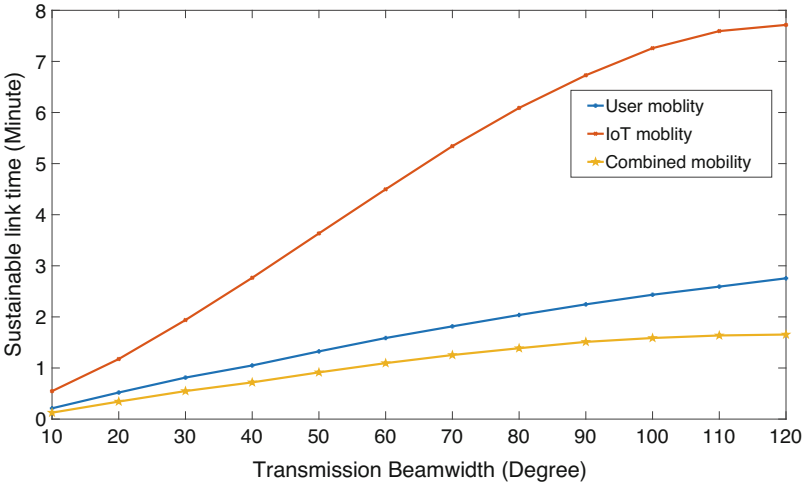


Fig. 9.9 Impact of the LDM pair mobility on the sustainable link time for random waypoint model for IoT and Manhattan model for user

within the transmission area for a significantly longer period. At a 120-degree transmission angle, user mobility offers almost double link sustainability time than IoT mobility. Figure 9.11, on the other hand, shows similar results with one exception; unlike the other results here, the combined mobility is almost the same as IoT mobility. This behavior is due to the gap in the link sustainability time of user and IoT mobility. As user mobility offers a much higher link sustainable time, the

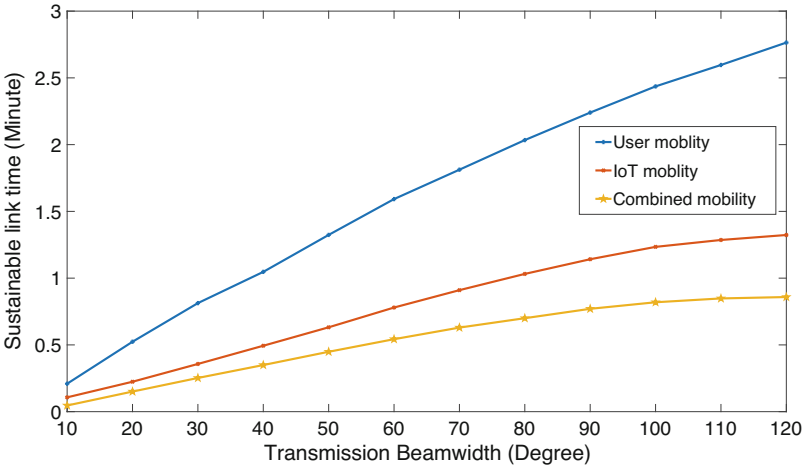


Fig. 9.10 Impact of the LDM pair mobility on the sustainable link time for Manhattan mobility model

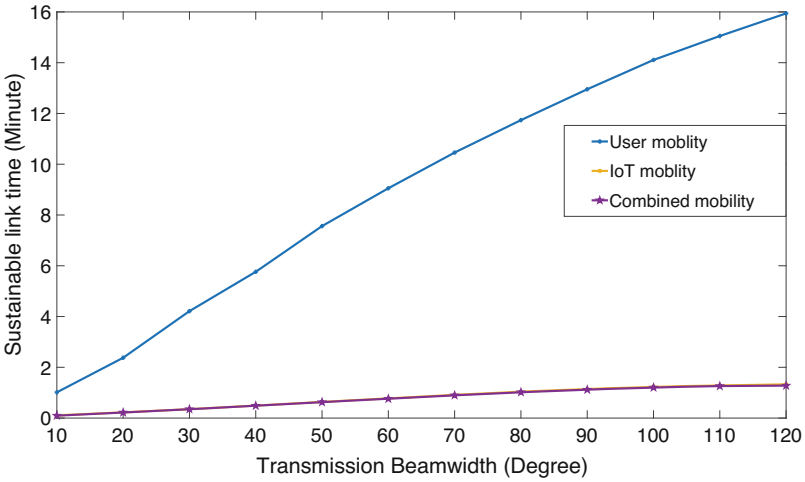


Fig. 9.11 Impact of the LDM pair mobility on the sustainable link time for Manhattan model for IoT and random waypoint model for user

pair needs to be broken almost every time due to the IoT device's network condition. This pattern can help the BS monitor one of the device's conditions with higher frequency based on their mobility model, lowering the computation load at the BS with higher efficiency.

In Fig. 9.12, both IoT device and user are moving as per the freeway mobility model. Both devices are moving at a higher speed, and hence the sustainable time is much lower than the urban scenario. However, the gap in performance between IoT

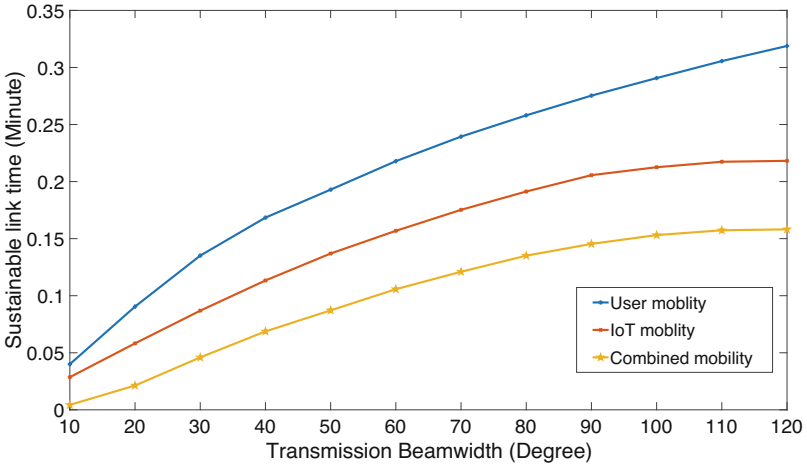


Fig. 9.12 Impact of the LDM pair mobility on the sustainable link time for freeway mobility model

Table 9.2 Combined performance comparison for 60-degree transmission angle

Mobility model	IoT	User	Both	Ratio IoT	Ratio user
Random waypoint	4.49	9.02	3.1	0.69	0.34
Random waypoint Manhattan	4.49	1.58	1.09	0.24	0.68
Manhattan	0.78	1.59	0.54	0.7	0.34
Manhattan random waypoint	0.77	9.05	0.76	0.98	0.08
Freeway	0.15	0.21	0.1	0.64	0.46

and user devices is lower in this scenario. In future work, other technologies such as beam following can be applied to improve the performance in this scenario as the movement is unidirectional and predictable. Table 9.2 shows the comparison of link sustainability at 60-degree transmission angle for all five transmission scenarios. When both devices are in the same mobility group, IoT devices are more likely to go out of range before the user. As we can see, when the gap between user and IoT sustainability is too significant, the gap between IoT and combined mobility performance becomes smaller. The BS can use this knowledge of known patterns during LDM pairing and make a more efficient IoT-user LDM pair that can improve the LDM performance in future wireless networks.

9.8 Conclusion

The analysis presented in this chapter focuses on the performance of LDM and its sustainability when both receivers have various degrees of mobility. LDM can

take advantage of the diversity of channel conditions and user requirements. An adaptive power-sharing ratio with extensive experimental measurement to use LDM UL for mMTC is developed in this work. The modeled power-sharing ratio has been explored to derive an analytical model that defines the LDM UL performance bound. The proposed analytical model has been shown to be robust for any order of modulation size, which justifies the LDM UL's feasibility for downlink communications for the future wireless network in certain downlink heavy use cases. In this chapter, we have considered three mobility models to represent the movement of both IoT devices and users. In each case, the transmission direction is set up based on the initial user location; hence, the user has a more maneuverable distance in all directions before it goes out of the coverage area. In contrast, the IoT devices are positioned randomly on the total coverage area, creating a scenario where IoT mobility plays a more dominant role in the sustainability of the LDM pair. The simulation results show the possibility of LDM in our proposed scenario. Future work on this area may consider machine learning to improve the accuracy of LDM pairing based on the mobility model and position of the user and IoT devices.

References

1. G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag et al., 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications. IEEE Commun. Mag. **52**(2), 97–105 (2014)
2. L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, Z. Wang, Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. **53**(9), 74–81 (2015)
3. J.P. Lemayian, J.M. Hamamreh, Recurrent neural network-based channel prediction in mMIMO for enhanced performance in future wireless communication, in *2020 International Conference on UK-China Emerging Technologies (UCET)* (IEEE, Piscataway, 2020), 1–4
4. L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H.M. Kim, J.-Y. Lee, P. Angueira, J. Montalban, Layered-division-multiplexing: Theory and practice. IEEE Trans. Broadcasting **62**(1), 216–232 (2016)
5. S.I. Park, J.-Y. Lee, S. Myoung, L. Zhang, Y. Wu, J. Montalbán, S. Kwon, B.-M. Lim, P. Angueira, H.M. Kim et al., Low complexity layered division multiplexing for ATSC 3.0. IEEE Trans. Broadcasting **62**(1), 233–243 (2016)
6. Y. Wu, B. Rong, K. Salehian, G. Gagnon, Cloud transmission: a new spectrum-reuse friendly digital terrestrial broadcasting transmission system. IEEE Trans. Broadcasting **58**(3), 329–337 (2012)
7. M. Earnshaw, K. Shelby, H. Lee, Y. Oh, M. Simon, Physical layer framing for ATSC 3.0. IEEE Trans. Broadcasting **62**(1), 263–270 (2016)
8. D. Gómez-Barquero, O. Simeone, LDM versus FDM/TDM for unequal error protection in terrestrial broadcasting systems: an information-theoretic view. IEEE Trans. Broadcasting **61**(4), 571–579 (2015)
9. H. Kim, J. Kim, S.-I. Park, J.-y. Lee, S. Kwon, N. Hur, Capacity analysis and improvement of ldm-based Multiple-PLP configurations in ATSC 3.0. IEEE Trans. Broadcasting **67**, 450–462 (2021)
10. L. Zhang, Y. Wu, W. Li, H.M. Kim, S.-I. Park, P. Angueira, J. Montalban, M. Velez, Channel capacity distribution of layer-division-multiplexing system for next generation digital broadcasting transmission, in *2014 IEEE International Symposium on Broadband Multimedia*

Systems and Broadcasting (IEEE, Piscataway, 2014), pp. 1–6
11. L. Zhang, Y. Wu, W. Li, K. Salehian, S. Lafleche, X. Wang, S.I. Park, H.M. Kim, J.-y. Lee, N. Hur et al., Layered-division multiplexing: an enabling technology for multicast/broadcast service delivery in 5G. IEEE Commun. Mag. **56**(3), 82–90 (2018)
12. L. Zhang, W. Li, Y. Wu, Y. Xue, E. Sousa, S.-I. Park, J.-Y. Lee, N. Hur, H.-M. Kim, Using non-orthogonal multiplexing in 5G-MBMS to achieve broadband-broadcast convergence with high spectral efficiency. IEEE Trans. Broadcasting **66**, 490–502 (2020)
13. E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu, W. Li, Using NOMA for enabling broadcast/unicast convergence in 5G networks. IEEE Trans. Broadcasting **66**, 503–5014 (2020)
14. M.S. Islam, M. Patwary, R. Tait, E. Peytchev, Layer division multiplexing for 5G DL transmission within ultra-dense heterogeneous networks, in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)* (IEEE, Piscataway, 2020), pp. 1–7
15. M. Patwary, S.K. Sharma, S. Chatzinotas, Y. Chen, M. Abdel-Maguid, R. Abd-Alhameed, J. Noras, B. Ottersten, Universal intelligent small cell (UnISCell) for next generation cellular networks. Digital Commun. Netw. **2**(4), 167–174 (2016)
16. S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, W. Shi, Edge computing for autonomous driving: Opportunities and challenges. Proc. IEEE **107**(8), 1697–1716 (2019)
17. H. Wang, B. Kim, J. Xie, Z. Han, E-auto: A communication scheme for connected vehicles with edge-assisted autonomous driving, in *ICC 2019–2019 IEEE International Conference on Communications (ICC)* (IEEE, Piscataway, 2019), pp. 1–6
18. M. Roy, H. Jamadagni, Performance analysis of MQAM-OFDM based WLAN in presence of zigbee interference in AWGN and rayleigh fading channel, in *2009 Sixth International Conference on Information Technology: New Generations* (IEEE, Piscataway, 2009), pp. 1178–1183
19. L. Zhang, W. Li, Y. Wu, K. Salehian, S. Laflèche, Z. Hong, S.-I. Park, H.M. Kim, J.-Y. Lee, N. Hur et al., Using layered-division-multiplexing to deliver multi-layer mobile services in ATSC 3.0. IEEE Trans. Broadcasting **65**(1), 40–52 (2018)
20. F. Khan, Z. Pi, S. Rajagopal, Millimeter-wave mobile broadband with large scale spatial processing for 5G mobile communication, in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, Piscataway, 2012), pp. 1517–1523
21. R. Kumar, M. Dave, Mobility models and their affect on data aggregation and dissemination in vehicular networks. Wirel. Personal Commun. **79**(3), 2237–2269 (2014)
22. E. Garro, J.J. Giménez, D. Gómez-Barquero, S.-I. Park, Performance evaluation of layer division multiplexing (LDM) combined with time frequency slicing (TFS), in *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (IEEE, Piscataway, 2015), pp. 1–5
23. E. Arruti, M. Mendicute, M. Barrenechea, QoS in industrial wireless networks using LDM, in *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)* (IEEE, Piscataway, 2017), pp. 1–6

Chapter 10
Wide Band THz Antenna Design Using Salp Swarm Algorithm for 6G Communications Systems

Sotirios K. Goudos and Mohammad Abdul Matin

10.1 Introduction

The future 6G wireless communication will expand the 5G capabilities even further. Specifically, 6G is expected to have higher bitrates (up to Tbps) and lower latency (less than 1ms) [26]. Such capabilities will enable the introduction of new services and applications. These services may include ultra-massive machine-type communications, holographic telepresence, wellness and e-health applications, pervasive connectivity in smart environments, pervasive edge intelligence, extremely reliable low-latency communications, augmented reality (AR), and virtual reality (VR) [7]. These services will need more efficient and effective wireless communications than ever before. Thus, 6G wireless communication systems should deliver broadband, near-instant, and dependable communication to enable huge data transmission at many frequencies, as well as a wide range of technologies. Such increased data rates demand and the fact that the spectral efficiency of the 5G wireless links is approaching its fundamental limits motivate researchers to explore and exploit higher-frequency bands. Hence, wireless communications in the terahertz (THz) band is considered to be an attractive candidate for 6G wireless networks. THz communications have the potential to provide an order of magnitude capacity improvements [2]. These 6G networks will outperform 5G in terms of latency and speed, and they will satisfy the above described emerging applications and services.

S. K. Goudos
ELEDIA@AUTH, Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
e-mail: sgoudo@physics.auth.gr

M. A. Matin (✉)
Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh
e-mail: mohammad.matin@northsouth.edu

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0_10

Moreover, recently the Institute of Electrical and Electronics Engineers (IEEE) introduced the first standard for low THz bands, namely, the IEEE 802.15.3d [1]. This defines global wireless personal area network (WPAN) application at frequencies from 252 to 325 GHz. Thus, this is the first global wireless communications standard for the 250–350 GHz frequency band. This will have channel bandwidths ranging from 2 to 70 GHz and a nominal PHY data throughput of 100 Gbps.

Taking into account all the above, the research in 6G wireless systems is a highly challenging topic. Antennas are one of the most important parts of any communication system. Moreover, antennas perform the task of receiving and transmitting electromagnetic waves. Because of the high transmission loss of low THz electromagnetic waves in free space, it becomes necessary to design low THz (more specifically in the frequency band of 252–325 GHz) antennas having very high gain to make up for the high path loss values. As a result, the THz communication system's range is extended, and the power limitation of the THz source generator is overcome. Antennas of this type can be used in a variety of ways, for example, high-speed short distance communication [24]. One of the key characteristics of antenna design for such communication systems is circular polarization (CP). CP has the benefit of freeing the receiver from the direction of arrival dependence in an input electromagnetic wave. Moreover, circular polarization has the potential to enhance the antenna's efficiency and eliminate the mismatch between the transmitter and the receiver caused by misalignment. In a multipath environment, CP plays an important role because it can reduce the delay spread [19].

To this end, new antennas must be designed and evaluated for 6G wireless systems. These antennas should meet all the forthcoming requirements of 6G networks, including the operation in the terahertz band and circular polarization. Considering the complexity of these antenna systems for terahertz communications, optimization algorithms are a well-promising design technique for these systems, offering comparative advantages compared to traditional antenna design techniques.

Microstrip antennas or generally printed antennas with different shapes like E-shape are low-cost options, and they will be evaluated. Antenna design necessitates the optimization of numerous geometrical parameters at the same time. A appropriate strategy for solving this problem is to use an optimization method or technique. Several patch antenna designs utilizing various evolutionary algorithms (EAs) as the optimization method can be found in the literature. These design strategies include the following: particle swarm optimization (PSO) [15], differential evolution (DE) [22, 25], genetic algorithms (GA) [13, 23], and teaching-learning optimization (TLBO)[11].

In this context, we use the salp swarm algorithm (SSA), a recently introduced algorithm [20]. SSA is a mathematical model that simulates salps' swarming behavior while navigating and hunting in oceans. This is the first time, to the best of the authors' knowledge, that SSA has been used to solve a design problem in THz antennas. The use of SSA is further motivated by the fact that it is parameter-free, which means it does not require parameter fine-tuning to perform well.

In this chapter, we use the SSA for 6G antenna design. We design a half E-shaped patch antenna for operation at 300 GHz. This antenna design can be used for operation in the whole low THz frequency band from 252 to 325 GHz as defined in [1]. We design this antenna on a Rogers RT/duroid 5880 laminate with 2.2 dielectric constant and $t = 0.127$ mm thickness. A single microstrip feed is used, which simplifies the design process even more. The suggested antenna is circularly polarized with an acceptable operating frequency bandwidth. Additionally, in order to achieve wide bandwidth, we introduce an additional design parameter by using not a full ground plane but rather a partial one. The final obtained antenna presents an ultra-wide band characteristic that covers the entire low THz frequency range from 252 to 325 GHz.

It is worth mentioning here that the presented work is based on previous design in [10], where the author designs a patch antenna for operation in 5G networks at a resonant frequency of 39 GHz using whale optimization algorithm (WOA). This work uses a different optimization algorithm (SSA) to design a patch antenna at a higher frequency of 300 GHz and also includes the concept of partial ground plane to increase antenna bandwidth.

10.2 Related Work

Patch antennas are an appealing antenna design option for millimeter-wave frequency bands and above. This is due to their benefits, which include ease of fabrication, low profile, and, more importantly, low cost. In the literature, several researchers have designed and manufactured antennas that work well in millimeter-wave frequency bands [3, 18]. The authors in [3] present a dual-band circularly polarized slotted patch antenna for operation at 28 and 38 GHz.

The polarization bandwidths (i.e., axial ratio value below 3 dB) reported by the authors are about 200 and 400 MHz on 28 and 38 GHz band, respectively. Moreover, the authors in [18] present a 1×2 compact wide band rectangular patch antenna for operation at IEEE 802.16a (5–6 GHz) and 802.16b (2–11 GHz). The antenna design uses the concept of partial ground to achieve wide band behavior. In the literature, the E-shaped patch antenna is a popular shape among the antenna designers [11, 15, 25]. The authors in [15] propose an E-shaped coaxial-fed patch antenna for operation in two frequency bands (1.8 and 2.4 GHz). The design method is a parallel PSO/FDTD algorithm, and the antenna is fabricated on a substrate with dielectric constant 2.2. Additionally, the E-shaped antenna is used by the authors in [25] to present a metallic antenna on air substrate for operation at the 5.2 and 5.8 GHz. The authors use a combined design method based on a numerical method (method of moments) and an evolutionary algorithm (differential evolution). Moreover, the authors in [11] extend the previous designs for 5G operation. They propose an aperture-coupled E-shaped patch antenna for dual band operation at 25 and 37 GHz. The design method used is TLBO.

Table 10.1 Patch antenna designs using different optimization methods

Antenna type	Design method	Design frequency	Polarization	Reference
Slotted patch line-fed	N/A	28, 38 GHz	CP	[3]
1 × 2 wide band rectangular patch, line fed, partial ground	N/A	2–11 GHz	LP	[18]
E-shaped coaxial-fed patch	Parallel PSO/FDTD	5.2, 5.8 GHz	LP	[15]
E-shaped coaxial-fed metallic patch	DE/MoM	1.8, 2.4 GHz	LP	[25]
E-shaped aperture-coupled patch	TLBO	25 GHz, 37 GHz	LP	[11]
Half E-shaped patch coaxial-fed	PSO	2.4 GHz	CP	[16]
Half E-shaped patch line-fed	Jaya	25 GHz	CP	[12]
Half E-shaped patch line-fed	Hybrid GWO-Jaya	Two antennas at 3.7 and 26 GHz	CP	[9]

The E-shaped design was chosen to improve the initial simple rectangular patch's characteristics like the functionality and the bandwidth by incorporating slots inside the patch. This antenna's major characteristics include wide band functioning and linear polarization (LP). In [16], the authors present a half E-shaped patch antenna that has an added metal part, namely, a shorting bar. The initial E-shaped antenna design has been modified in this way. In the frequency of interest, this half E-shaped antenna is circularly polarized. The study provided a two-layered, coaxial-fed, half E-shaped antenna for operating at 2.4 GHz. The design algorithm used is PSO. The authors in [9] extend the half E-shaped patch for operation in 5G at 25 GHz using the Jaya algorithm. The antenna is excited using a feed line, whose dimensions are also optimized. Additionally, the authors in [12] describe half E-shaped dual-band patch antenna designs for 5G frequencies (3.7 and 26 GHz). A hybrid GWO-Jaya algorithm is used to create the antenna designs. Table 10.1 briefly reports the above presented references.

10.3 Algorithm Description

The SSA mathematically models salp swarming behavior when moving and foraging in oceans [20]. Each salp in SSA represents a D-dimensional solution vector. Additionally, the food position is represented by the solution vector that obtains the best fitness value discovered in each iteration. The behavior of salps is modeled in SSA by dividing the population into two groups. There are two types of vectors in this algorithm: leaders and followers. The leaders are the salps in front of the chain, while the remainder are known as followers.

This indicates that the algorithm considers leaders to be the solution vectors with the best fitness values, while followers are the solution vectors with the worst ones.

The SSA divides the population into two equal-sized groups, of size $\frac{NP}{2}$, where NP is the population size.

The updating rule for the leader's position is as follows:

$$x_{G+1,m,k} = \begin{cases} Food_k + C_1 \times \left[rand_{2k}\left(x_{k,U} - x_{k,L}\right) + x_{k,L}\right] & rand_{3k} < 0.5 \\ Food_k - C_1 \times \left[rand_{2k}\left(x_{k,U} - x_{k,L}\right) + x_{k,L}\right] & rand_{3k} \geq 0.5 \end{cases} \tag{10.1}$$

where $x_{k,U}, x_{k,L}$ denote the upper and lower bounds in the kth, dimension respectively, and the position of the m-th leader salp for $m < NP/2$ in the kth dimension for generation $G + 1$ is denoted by $x_{G+1,m,k}$. Additionally, the best vector in the population or the food source position in the kth dimension is denoted by $Food_k$. Moreover, C_1 is an important parameter for SSA and is given by

$$C_1 = 2 \times e^{-\left(\frac{4G}{G\max}\right)^2} \tag{10.2}$$

where G denotes the current iteration and G_{max} is the maximum number of iterations. The importance of the C_1 parameter, according to [20], is that it balances exploration with exploitation. Furthermore, $rand_{2k}$ and $rand_{3k}$ are uniformly distributed random numbers in the range [0, 1]. These random numbers aid the algorithm in determining whether the next point in the kth dimension will be positive or negative infinity, as well as the step size. For the remaining $\frac{NP}{2}$ vectors, SSA defines a separate position update rule, which is stated as:

$$x_{G+1,m,k} = \frac{\left(x_{G,m,k} + x_{G,m-1,k}\right)}{2} \tag{10.3}$$

The SSA for antenna design is briefly described in the Algorithm 1. Among others, the main SSA advantages include the ability to simple write the source code of the algorithm in any programming language. However, the SSA source code is available from the authors in MATLAB, Python, and R. As stated earlier, an additional advantage is the fact that SSA is parameter free. On the other hand, if we consider the SSA disadvantages, we will include the fact that the algorithm may lead to stagnation to local optima.

The SSA has been employed for several optimization problems in wireless communications. These include the combined power allocation and user association problem in non-orthogonal multiple access networks (NOMA) [8]. This type of problem is non-convex and very difficult to solve with any other method. Additionally, SSA was also employed in [6] for 5G antenna design. The authors applied SSA for a bow-tie MIMO antenna design that resonates at 28 GHz for use in Frequency Range 2 (FR2) in 5G cellular networks.

Algorithm 1: SSA algorithm for antenna design

Initialize the maximum number of generations G_{max}
Initialize the random initial population of size NP
Calculate antenna fitness value for population
Set the antenna design with the best fitness to be the initial food $Food$
$k = 1$
while $k < G_{max}$ **do**
    Calculate new C_1 value using (10.2)
    for r=1 to NP **do**
        if $r <= NP/2$ **then**
            Update the leaders positions using (10.1)
        else
            Update the followers positions using (10.3)
        end if
    end for
    Sort all the antenna design vectors according to fitness value in descending order
    If a better antenna design is found, update the food vector
    $k++$
end while

10.4 6G Patch Antenna Design

The E-shaped patch antennas are linearly polarized and wide band (LP). Patch antennas can be used for a variety of purposes, including high-speed short-distance communication. Furthermore, circular polarization is frequently a prerequisite for such communication systems (CP). The fundamental advantage of CP is that it frees the receiver from the direction of arrival dependency in an incoming electromagnetic wave. Misalignment causes a mismatch between the transmitter and receiver, which is eliminated by CP. Thus, CP potentially increases the antenna's efficiency. Therefore, it is sometimes necessary to design CP antennas for some applications like for THz operation in future 6G networks. The authors of [16] create a half E-shaped antenna for Wi-Fi operation (2.4 GHz), which is circularly polarized. The proposed design in [16] consists of two layers, and it is probe-fed. As a result, designing CP antennas is occasionally necessary for particular applications. The authors of [16] create a half E-shaped antenna that resonates at 2.4 GHz. This antenna design is circularly polarized. Moreover, the proposed design in [16] is probe-fed and consists of two layers.

The E-shaped antenna can be used as a starting shape to create the half E-shaped CP antenna. Figure 10.1 shows the shape evolution from the E-shaped LP antenna to the half E-shaped CP antenna.

We can see from these diagrams that a shorting bar of width W_b is required for the half E-shaped CP antenna. Currents in the y direction are induced by this shorting bar. As a result, the antenna can operate in CP mode. The shorting bar position is one major feature that influences the axial ratio (AR) bandwidth. As a result, another unknown variable is the proper shorting bar position.

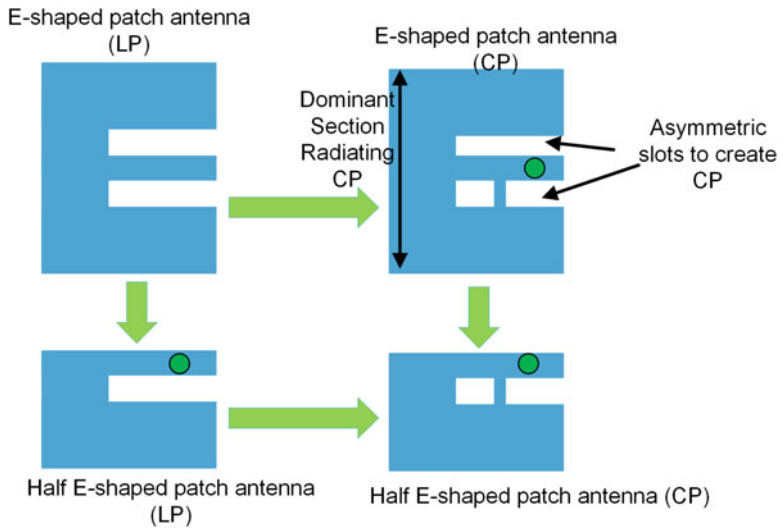


Fig. 10.1 E-shaped antenna shape evolution for CP [16]

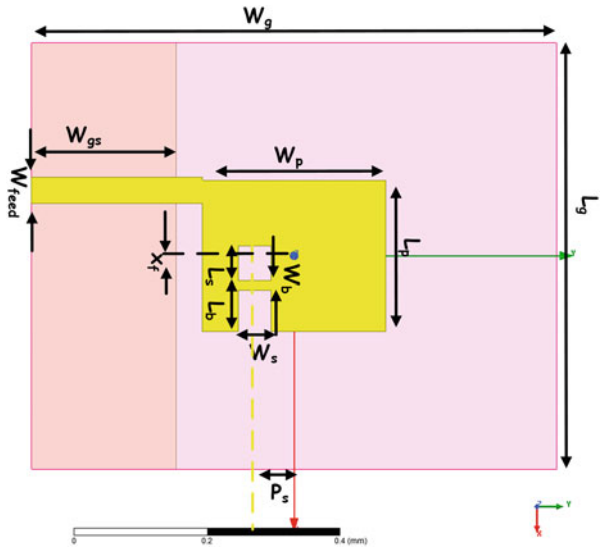


Fig. 10.2 Antenna geometry top view

Figure 10.2 shows the geometry of a modified half E-shaped patch antenna. The antenna geometry is rather complex, as one can see. It is made up of 12 geometrical design parameters. As a result, estimating the effect of each design parameter in order to obtain the specified antenna performance will be extremely difficult, if not impossible.

As a result, the apparent solution to this design difficulty is an optimization strategy. Additionally, because we need a circular polarized antenna, we need to have axial ratio (AR) below 3 dB in the frequency band of interest.

Additionally, in order to obtain ultra-wide band (UWB) behavior, a modification in antenna ground plane is required. Toward this end, it is common in the literature to use a partial ground plane or ground planes that have additional slots [14, 17]. The basic operation of partial grounds is the reduction of the energy stored in the substrate. This reduction results in a quality factor decrease, which, in turn, causes the bandwidth to increase. Partial ground planes also have back radiation, which results in more radiation loss.

Thus, there are two antenna design goals. One is to have the S_{11} magnitude values below -10 dB at the working frequency, and the other is to keep the axial ratio below 3 dB. The following formulation is used to express the design problem [12]

$$F(\bar{v}) = S_{11}(\bar{v}) + \xi \times \left|\left|S_{11}(\bar{v})\right| - |L_{dB}|\right| + \\ +\xi \times \left|AR(\bar{v}) - L_{AR}\right| \tag{10.4}$$

where $\bar{v}$ denotes the vector of the unknown antenna geometry parameters, the S_{11} magnitude is denoted by S_{11}, and the axial ratio at the design frequency is denoted by AR. Additionally, we denote S_{11} dB limit as L_{dB} and AR dB limit as L_{AR}, while ξ represents a penalty factor. For both design cases, we select $L_{dB} = -10$ dB and $L_{AR} = 3$ dB.

The usage of a full-wave numerical method is required to compute an objective function like (10.4). A commercial full-wave electromagnetic (EM) software can be used to model the half E-shaped patch antenna. ANSYS HFSS [4] is a possible software to employ for this task. Furthermore, integrating an evolutionary algorithm's in-house source code with an EM solver necessitates the usage of a wrapper program. The authors in [5] also describe the construction of a separate optimizer that invokes external software so that the simulations for the evaluation of the fitness function are run. The whole process can be described in Algorithm 2. The wrapper application can use the HFSS MATLAB API [21] to combine the source code of an evolutionary algorithm written in MATLAB with ANSYS HFSS. As a result, the SSA algorithm can be combined with the wrapper program to call the EM solver in our scenario. Algorithm 2 can be used to describe the entire procedure.

Algorithm 2: Wrapper procedure for antenna design

1: Given a new geometry vector $\bar{v}$ as input
2: Use the MATLAB to create a HFSS Visual Basic script file
3: Call HFSS from MATLAB using the new script file
4: Use HFSS GUI to generate antenna geometry
5: Run simulations to compute the antenna indices (e.g., S_{11}, AR) at the desired operating frequency
6: Generate text files that include the simulation results
7: Based on the output files, compute the fitness function value
8: Exit from the routine and return the fitness function value to SSA

Table 10.2 Geometrical parameters for antenna design

Parameter	Value (mm)	Parameter	Value (mm)
L_p	v_1	W_{feed}	$v_7 \times (W_p/2)$
W_p	v_2	W_{gs}	$v_8 \times W_g$
L_g	$(1 + v_3) \times L_p$	P_s	$v_9 \times W_p$
W_g	$(1 + v_4) \times W_p$	W_s	$v_{10} \times W_p$
L_b	$v_5 \times L_p$	L_s	$v_{11} \times L_p$
W_b	$v_6 \times W_p$	x_f	$v_{12} \times (L_p/2)$

We need to specify the upper and lower limits for each antenna geometrical parameter in the preceding example. This is a complicated problem, which has a strong frequency dependence. If we establish the geometrical parameter limits, they are only valid for the given design and design frequency. The difficulty now is how to improve the performance of another antenna with the same shape but operating at different frequencies. For each geometrical parameter, one would have to find the right bounds all over again. This process necessitates a lot of computer resources and takes a long time because multiple tests must be run in order to discover the most appropriate limits.

As a result, a frequency-independent design technique can be used. The primary idea is to set the first two starting parameters, namely, the patch width W_p and patch length L_p, with acceptable frequency restrictions, and then set the remaining parameters as a perchance of these two. The remaining unknowns' limits will then be inside the range of [0, 1]. In the above-described scenario, the twelve optimization variables $(v_1, v_2, \ldots, v_{12})$ are considered. Table 10.2 lists the geometrical parameters for antenna design and their representation in [0, 1]. Moreover, L_{feed} parameter is equal to $(W_g/2) - (W_p/2)$.

If we need to construct a new design, we merely need to calculate the bounds of the first two optimization variables for the new frequencies. Setting the boundaries of an optimization variable inside the range of [0, 1] reduces the search space, making it easier for an evolutionary algorithm to provide results.

10.5 Numerical Results

In this section, we present the antenna design for future 6G mobile networks. As it is reported in [1], the low THz band for WPAN use defines the frequencies from 252 to 325 GHz. In this book chapter, we will use the 300 GHz as the design frequency of our work. It must be pointed out that a similar design case is reported in [12], where the optimization method is a hybrid Jaya-GWO and the design frequency is 26 GHz.

Figures 10.2 and 10.3 show the antenna geometry for this case. The antenna is made from a single layer of dielectric, and the feeding is done via a 50ohm feed line. We have selected for this design the dielectric Rogers RT/duroid 5880 laminate ($\epsilon_r = 2.21$) with 0.127 mm thickness. This kind of material and thickness is suitable for operation in 300 GHz.

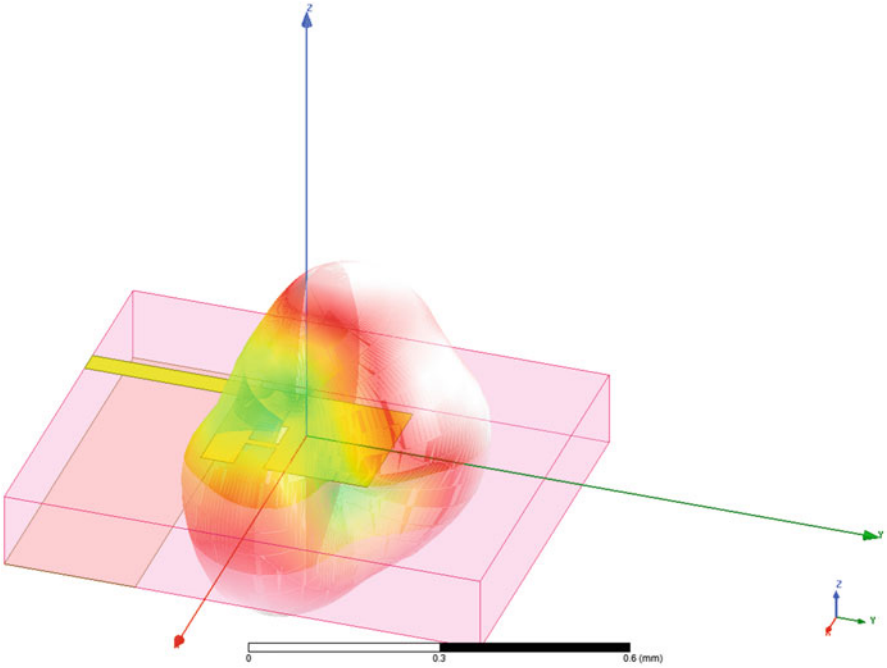


Fig. 10.3 Antenna geometry in 3D view

Table 10.3 Parameter values of the best achieved antenna design at 300 GHz

Parameter	Value (mm)	Parameter	Value (mm)
W_p	0.28	W_{feed}	0.04
L_p	0.23	W_{gs}	0.22
W_g	0.80	P_s	0.06
L_g	0.65	W_s	0.05
L_b	0.05	L_s	0.13
W_b	0.01	x_f	−0.10

As it is described in the previous section, there are 12 geometrical parameters to be fine-tuned in order to design this antenna. We used the SSA as the main optimization method to design this antenna. Due to computationally intensive operation while running the HFSS, we need to set the population size to a small value like 20. Moreover, the maximum number of generations is preset at 2000. We, then, run the SSA for ten independent trials and obtain the best result. SSA does not need any setting of other control parameters, which is an additional advantage of the algorithm.

Table 10.3 lists all the decision variables or the antenna geometrical parameters that form the best design found. In order to show the advantage of using a partial ground, we compare the results with the same design but using a full ground. Initially, we illustrate the 3D radiation patterns for both designs at 300 GHz in

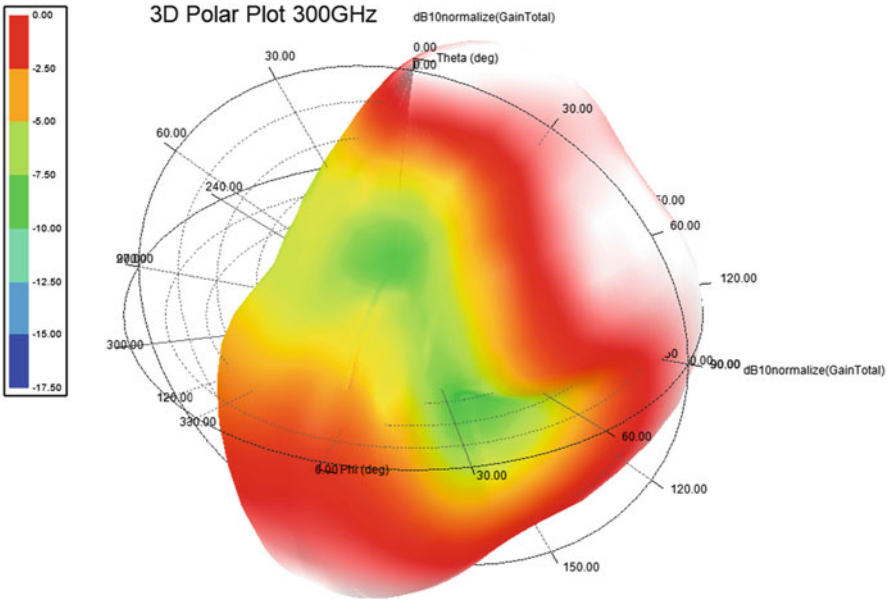


Fig. 10.4 Partial ground case: 3D gain pattern of the best design found by SSA at 300 GHz

Figs. 10.4 and 10.5. The initial observation shows that in both cases, the radiation pattern is quite near to omni-directional.

Additionally, we compare the surface current distribution for both ground cases at 300 GHz in Figs. 10.6 and 10.7. The figures illustrate the surface current distribution for the patch and the ground. We notice that in both cases, the current values present a higher value in the area near the shorting bar and in a small area near the feed line. Moreover, the surface current values of the patch in the full ground case are generally higher than in the partial ground case. This can be expected because in the partial ground case, there is more radiation loss and the antenna quality factor is reduced. This is shown also in the surface current distribution of the partial ground where the higher current values are not only around the patch as in full ground but also near the ground edge.

Figures 10.8, 10.9 and 10.10 show radiation patterns for the partial ground case at 300 GHz for three distinct plane cuts. The corresponding radiation patterns for the same plane cuts of the full ground cases are illustrated in Figs. 10.11, 10.12 and 10.13. We notice that the radiation patterns for the full ground case are close to being uniform in space. Additionally, the radiation patterns of the partial ground case are more directive than the full ground case.

The distribution of radiated power, on the other hand, is deemed satisfactory since the gain remains near maximum inside wide spans of space, allowing the antenna to successfully transmit and receive signals from a wide range of directions. In the case of mobile communications, this performance is desired. The realized gains

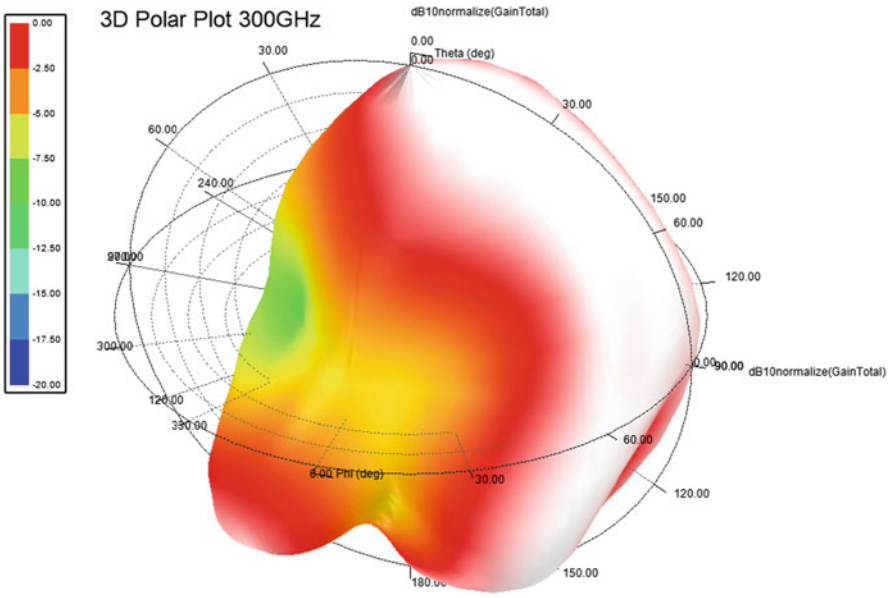


Fig. 10.5 Full ground case: 3D gain pattern of the best design found by SSA at 300 GHz

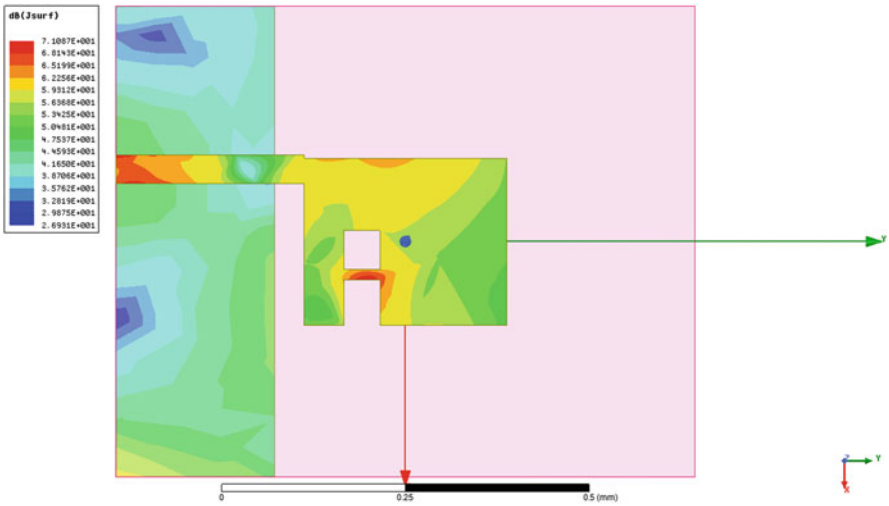


Fig. 10.6 Partial ground case: Simulated surface current distribution of the best design at 300 GHz

at 300 GHz are 3.48 and 6.84 dBi for the partial ground and the full ground case, respectively. Also, the radiation efficiency at 300 GHz is about 98 and 97% for the partial ground and the full ground case, respectively.

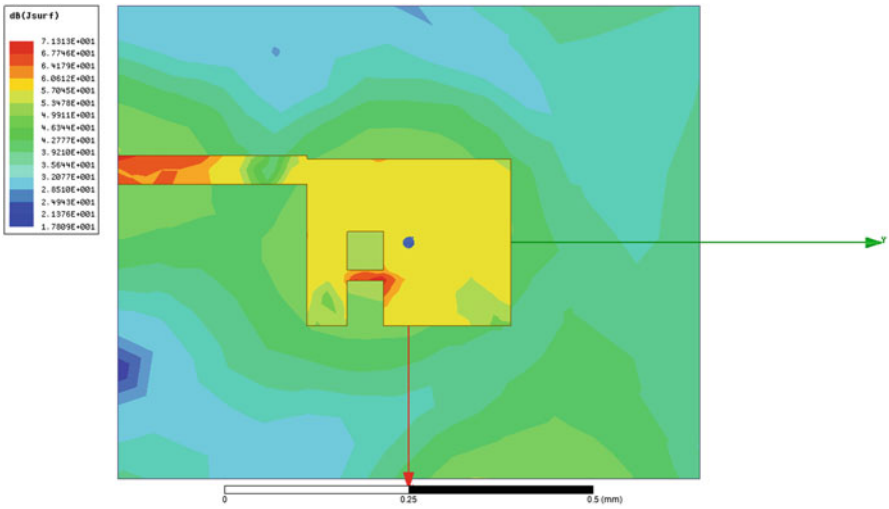


Fig. 10.7 Partial ground case: Simulated surface current distribution of the best design at 300 GHz

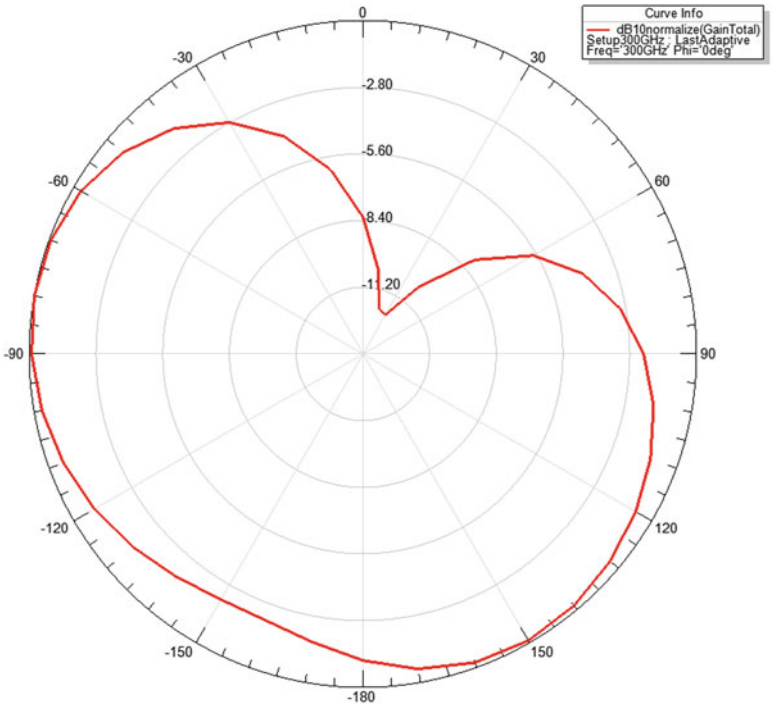


Fig. 10.8 Partial ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\phi = 0°$

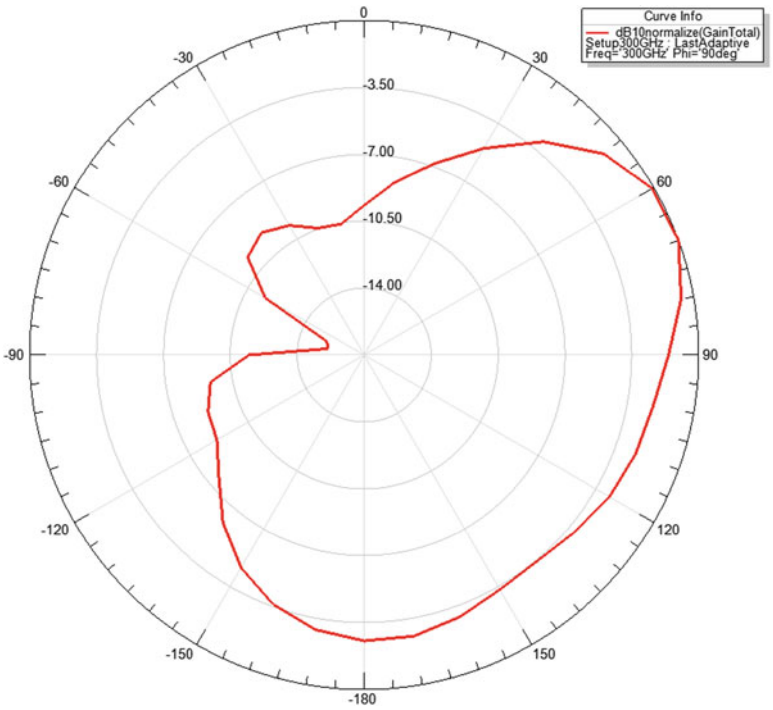


Fig. 10.9 Partial ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\phi = 90°$

The S_{11} frequency response of the best antenna design is shown in Fig. 10.14. We may observe that the best design antenna exhibits an ultra-wide band behavior starting from 250 to 325 GHz. Thus, it covers the whole low THz frequency band defined by the [1]. The antenna resonates at 300 GHz with a very low S_{11} value below -70 dB. Additionally, we notice that the full ground case presents a worse behavior with S_{11} below -10 dB in very limited bandwidth and levels that do not become less than -10 dB.

The axial ratio frequency response plot is presented in Fig. 10.15. We notice that the full ground case obtains axial ratio values quite higher than 3 dB and thus does not present circular polarization. Moreover, the partial ground case exhibits circular polarization from about 294–311 GHz. Therefore, the circular polarization bandwidth is about 17 GHz.

Overall, SSA was able to obtain a solution to the antenna design problems that fully covers the strict design requirements for both 6G operation and circular polarization.

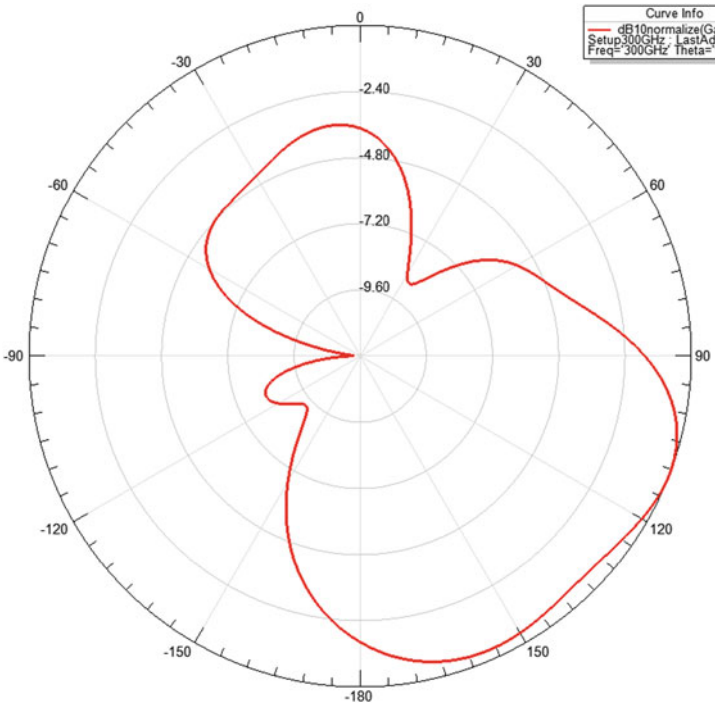


Fig. 10.10 Partial ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\theta = 90°$

10.6 Conclusion

In this book chapter, we presented a wide band antenna design framework for low THz band for future 6G wireless communications. The optimization algorithm used was SSA, which proved to be very effective in antenna design. The obtained antenna presents an ultra-wide band behavior that covers the whole band from 250 to 325 GHz. The $S_{11} - AR$ bandwidth in this case is about 17 GHz. Moreover, the design is at a frequency of 300 GHz, the antenna gain is 3.48 dBi, and the radiation efficiency is about 98%, which can be considered satisfactory.

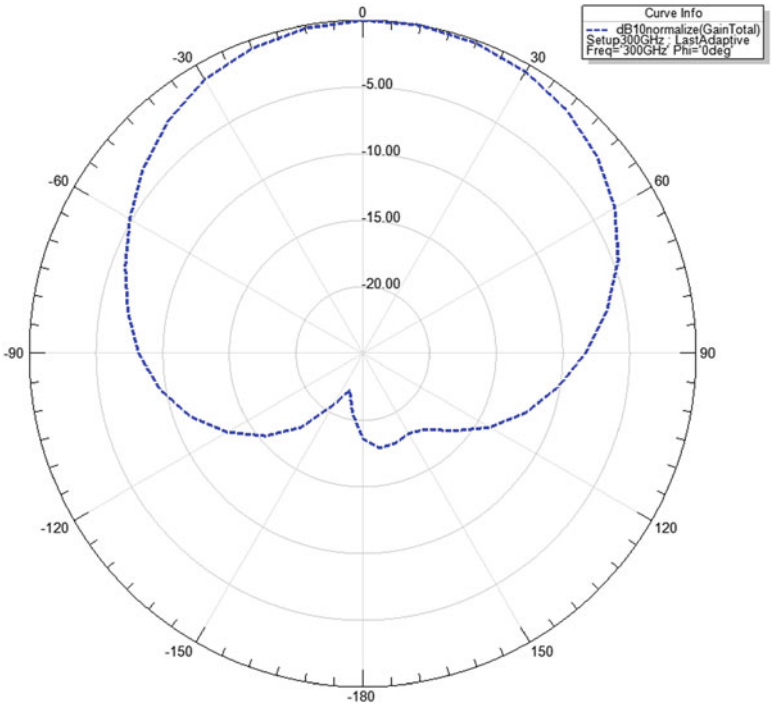


Fig. 10.11 Full ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\phi = 0°$

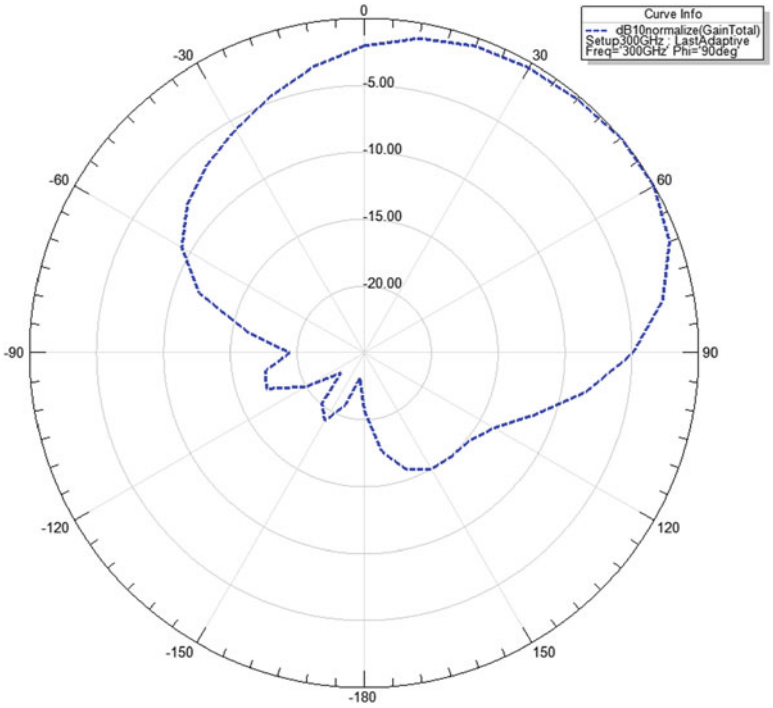


Fig. 10.12 Full ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\phi = 90°$

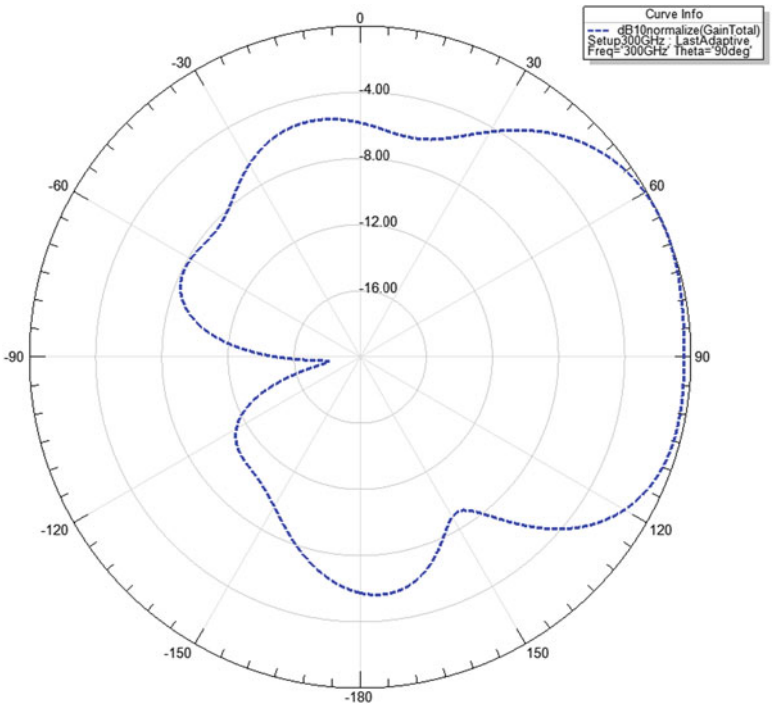


Fig. 10.13 Full ground case: Radiation pattern of the best obtained design by SSA at 300 GHz. $\theta = 90°$

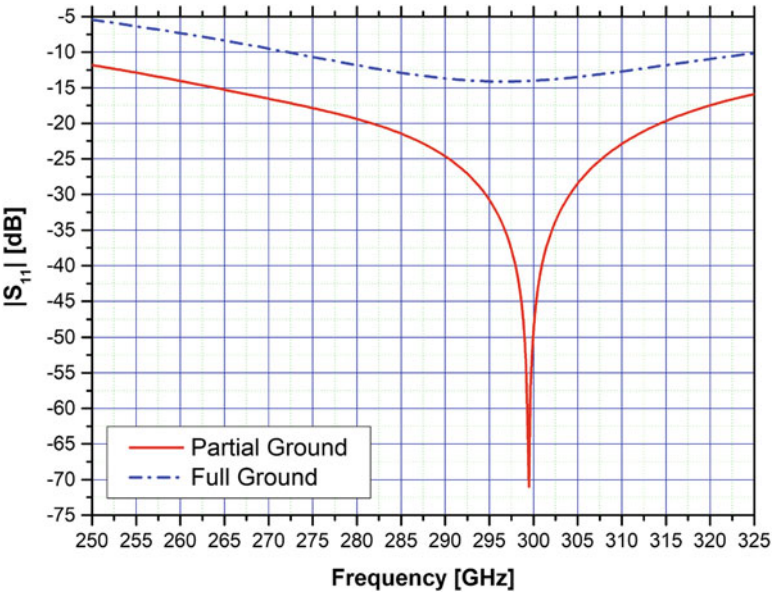


Fig. 10.14 S_{11} plot of the best antenna best design found by SSA

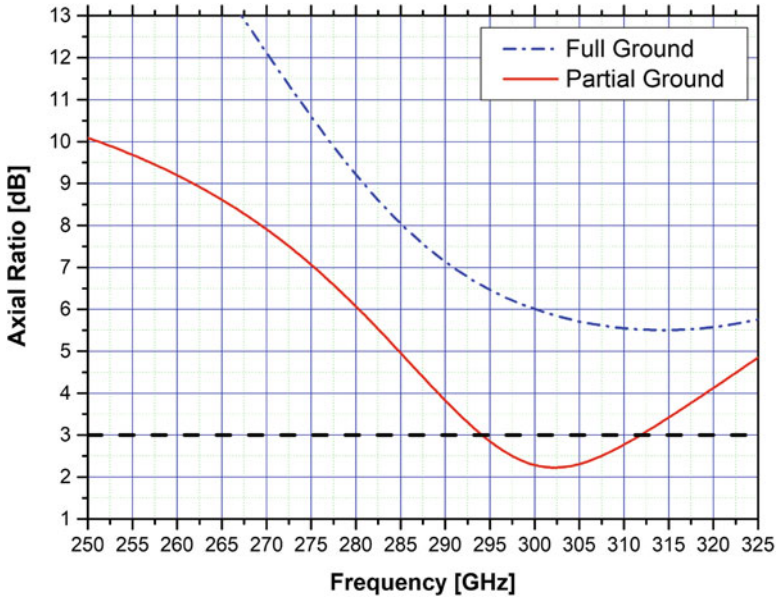


Fig. 10.15 Axial ratio plot of the best design found by SSA

References

1. IEEE Standard for High Data Rate Wireless Multi-Media Networks–Amendment 2: 100 gb/s Wireless Switched Point-to-Point Physical Layer. IEEE Std 802.15.3d-2017 (Amendment to IEEE Std 802.15.3-2016 as amended by IEEE Std 802.15.3e-2017) (2017) pp. 1–55 . https://doi.org/10.1109/IEEESTD.2017.8066476
2. I.F. Akyildiz, C. Han, S. Nie, Combating the distance problem in the millimeter wave and terahertz frequency bands. IEEE Commun. Mag. **56**(6), 102–108 (2018). https://doi.org/10.1109/MCOM.2018.1700928
3. H. Aliakbari, A. Abdipour, R. Mirzavand, A. Costanzo, P. Mousavi, A single feed dual-band circularly polarized millimeter-wave antenna for 5G communication, in *2016 10th European Conference on Antennas and Propagation, EuCAP 2016* (2016)
4. ANSYS: Electomagnetics Suite, ANSYS: User's Guide, Version 16.1 (2015)
5. D.J. Bekers, S. Monni, S.M. van den Berg, A.M. van de Water, B.J. Morsink, C. Alboin, V. Ducros, M. Celikbas, J. Blanche, N. Fiscante, G. Gerini, J.P. Martinaud, M. Rochette, G.H.C. van Werkhoven, Optimization of phased arrays integrated with fss and feeding elements based on parametric models, in *The Second European Conference on Antennas and Propagation, 2007. EuCAP* (2007), pp. 1–7
6. A.D. Boursianis, S.K. Goudos, T.V. Yioultsis, K. Siakavara, P. Rocca, Mimo antenna design for 5G communication systems using salp swarm algorithm, in *2020 International Workshop on Antenna Technology (iWAT)* (2020), pp. 1–3. https://doi.org/10.1109/iWAT48004.2020.1570618331
7. M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, M. Zorzi, Toward 6g networks: use cases and technologies. IEEE Commun. Mag. **58**(3), 55–61 (2020). https://doi.org/10.1109/MCOM.001.1900411

8. S. Goudos, Joint power allocation and user association in non-orthogonal multiple access networks: an evolutionary approach. Phys. Commun. **37** (2019). https://doi.org/10.1016/j.phycom.2019.100841
9. S. Goudos, T. Yioultsis, K. Dalidou, K. Siakavara, A low cost wide band and circularly polarized modified half e-shaped patch antenna for 5G mobile communications, in *IET Conference Publications, EuCAP 2018*, vol. 2018 (2018)
10. S.K. Goudos, *Application of the Whale Optimization Algorithm to Antenna Design for mm-Wave 5G Communications Systems* (Springer International Publishing, Cham, 2021), pp. 251–267.. https://doi.org/10.1007/978-3-030-74311-6_8
11. S.K. Goudos, A. Tsiflikiotis, D. Babas, K. Siakavara, C. Kalialakis, G.K. Karagiannidis, Evolutionary design of a dual band e-shaped patch antenna for 5G mobile communications, in *2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)* (2017), pp. 1–4
12. S.K. Goudos, T.V. Yioultsis, A.D. Boursianis, K.E. Psannis, K. Siakavara, Application of new hybrid jaya grey wolf optimizer to antenna design for 5G communications systems. IEEE Access **7**, 71061–71071 (2019)
13. R.L. Haupt, Antenna design with a mixed integer genetic algorithm. IEEE Trans. Antenn. Propag. **55**(3 I), 577–582 (2007)
14. M.M. Islam, M.T. Islam, M.R.I. Faruque, R.W. Aldhaheri, M. Samsuzzaman, Design of a compact UWB antenna with a partial ground plane on epoxy woven glass material. Sci. Eng. Composite Mat. **24**(1), 73–79 (2017). https://doi.org/10.1515/secm-2014-0297
15. N. Jin, Y. Rahmat-Samii, Parallel particle swarm optimization and finite-difference time-domain (PSO/FDTD) algorithm for multiband and wide-band patch antenna designs. IEEE Trans. Antenn. Propag. **53**(11), 3459–3468 (2005). https://doi.org/10.1109/tap.2005.858842
16. J.M. Kovitz, H. Rajagopalan, Y. Rahmat-Samii, Circularly polarised half e-shaped patch antenna: a compact and fabrication-friendly design. IET Microw. Antenna Propag. **10**(9), 932–938 (2016)
17. J. Kumar, S. Shirgan, Compact partial ground plane 1x2 patch antennas, in *2014 International Conference on Computational Intelligence and Communication Networks* (2014), pp. 33–37. https://doi.org/10.1109/CICN.2014.19
18. K. Mak, H. Lai, K. Luk, C. Chan, Circularly polarized patch antenna for future 5G mobile phones. IEEE Access **2**, 1521–1529 (2014)
19. T. Manabe, Y. Miura, T. Ihara, Effects of antenna directivity and polarization on indoor multipath propagation characteristics at 60 GHz. IEEE J. Sel. Areas Commun. **14**(3), 441–447 (1996)
20. S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, Mirjalili, S.M.: Salp Swarm algorithm: A bio-inspired optimizer for engineering design problems. Adv. Eng. Softw. (2017). https://doi.org/10.1016/j.advengsoft.2017.07.002
21. V. Ramasami, A HFSS API to Control HFSS from Matlab (2020). https://github.com/yuip/hfss-api/. Accessed 19 July 2021
22. P. Rocca, G. Oliveri, A. Massa, Differential evolution as applied to electromagnetics. IEEE Antenna. Propag. Mag. **53**(1), 38–49 (2011). https://doi.org/10.1109/MAP.2011.5773566
23. F.J. Villegas, T. Cwik, Y. Rahmat-Samii, M. Manteghi, A parallel electromagnetic genetic-algorithm optimization (EGO) application for patch antenna design. IEEE Trans. Antenna. Propag. **52**(9), 2424–2435 (2004)
24. R. Xu, S. Gao, B.S. Izquierdo, C. Gu, P. Reynaert, A. Standaert, G.J.Gibbons, W. Bösch, M.E. Gadringer, D. Li, A review of broadband low-cost and high-gain low-terahertz antennas for wireless communications applications. IEEE Access **8**, 57615–57629 (2020). https://doi.org/10.1109/ACCESS.2020.2981393
25. L. Zhang, Z. Cui, Y.C. Jiao, F.S. Zhang, Broadband patch antenna design using differential evolution algorithm. Microw. Opt. Technol. Lett. **51**(7), 1692–1695 (2009)
26. Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G.K. Karagiannidis, P. Fan, 6G wireless networks: vision, requirements, architecture, and key technologies. IEEE Vehic. Technol. Mag. **14**(3), 28–41 (2019). https://doi.org/10.1109/MVT.2019.2921208

Index

M. A. Matin (ed.), *A Glimpse Beyond 5G in Wireless Networks*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-13786-0