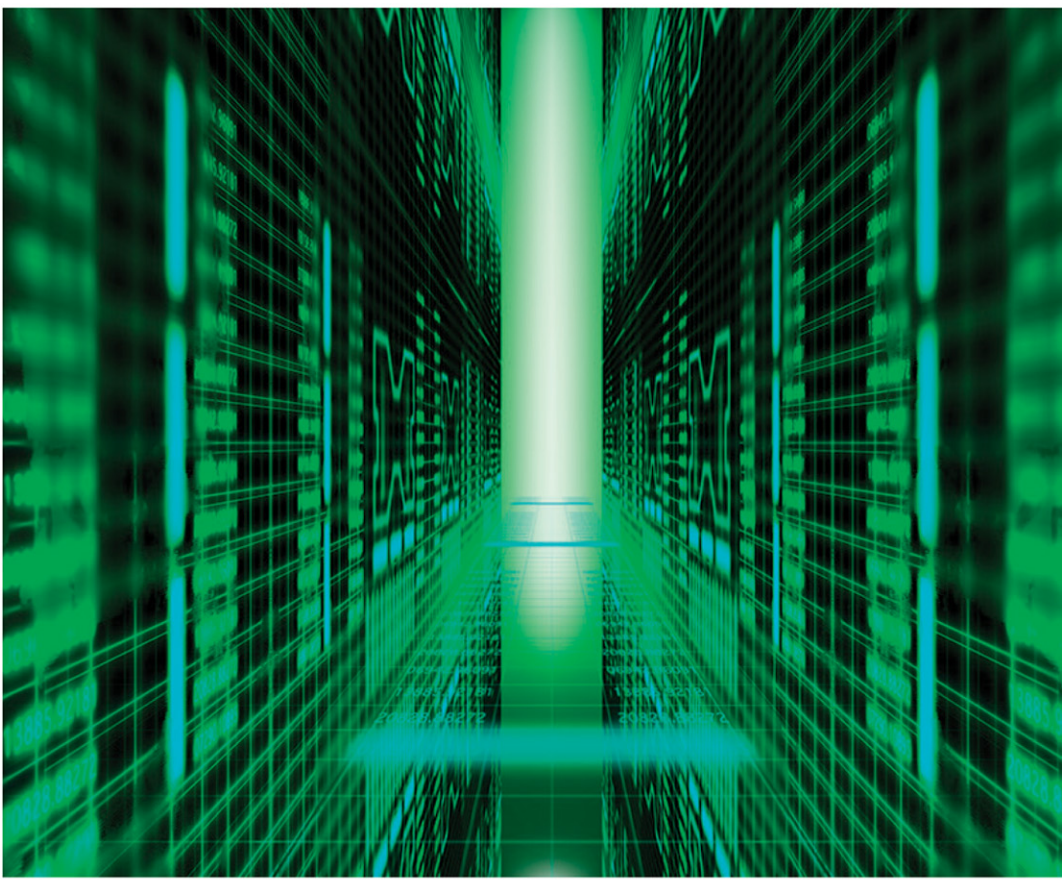
The Institution of Engineering and Technology

Green Communications for Energy-Efficient Wireless Systems and Networks

Edited by
Himal A. Suraweera, Jing Yang, Alessio Zappone and John Thompson



IET TELECOMMUNICATIONS SERIES 91

Green Communications for Energy-Efficient Wireless Systems and Networks

Other volumes in this series:

Volume 9 **Phase Noise in Signal Sources** W.P. Robins
Volume 12 **Spread Spectrum in Communications** R. Skaug and J.F. Hjelmstad
Volume 13 **Advanced Signal Processing** D.J. Creasey (Editor)
Volume 19 **Telecommunications Traffic, Tariffs and Costs** R.E. Farr
Volume 20 **An Introduction to Satellite Communications** D.I. Dalgleish
Volume 26 **Common-Channel Signalling** R.J. Manterfield
Volume 28 **Very Small Aperture Terminals (VSATs)** J.L. Everett (Editor)
Volume 29 **ATM: The broadband telecommunications solution** L.G. Cuthbert and J.C. Sapanel
Volume 31 **Data Communications and Networks, 3rd Edition** R.L. Brewster (Editor)
Volume 32 **Analogue Optical Fibre Communications** B. Wilson, Z. Ghassemlooy and I.Z. Darwazeh (Editors)
Volume 33 **Modern Personal Radio Systems** R.C.V. Macario (Editor)
Volume 34 **Digital Broadcasting** P. Dambacher
Volume 35 **Principles of Performance Engineering for Telecommunication and Information Systems** M. Ghanbari, C.J. Hughes, M.C. Sinclair and J.P. Eade
Volume 36 **Telecommunication Networks, 2nd Edition** J.E. Flood (Editor)
Volume 37 **Optical Communication Receiver Design** S.B. Alexander
Volume 38 **Satellite Communication Systems, 3rd Edition** B.G. Evans (Editor)
Volume 40 **Spread Spectrum in Mobile Communication** O. Berg, T. Berg, J.F. Hjelmstad, S. Haavik and R. Skaug
Volume 41 **World Telecommunications Economics** J.J. Wheatley
Volume 43 **Telecommunications Signalling** R.J. Manterfield
Volume 44 **Digital Signal Filtering, Analysis and Restoration** J. Jan
Volume 45 **Radio Spectrum Management, 2nd Edition** D.J. Withers
Volume 46 **Intelligent Networks: Principles and applications** J.R. Anderson
Volume 47 **Local Access Network Technologies** P. France
Volume 48 **Telecommunications Quality of Service Management** A.P. Oodan (Editor)
Volume 49 **Standard Codecs: Image compression to advanced video coding** M. Ghanbari
Volume 50 **Telecommunications Regulation** J. Buckley
Volume 51 **Security for Mobility** C. Mitchell (Editor)
Volume 52 **Understanding Telecommunications Networks** A. Valdar
Volume 53 **Video Compression Systems: From first principles to concatenated codecs** A. Bock
Volume 54 **Standard Codecs: Image compression to advanced video coding, 3rd Edition** M. Ghanbari
Volume 59 **Dynamic Ad Hoc Networks** H. Rashvand and H. Chao (Editors)
Volume 60 **Understanding Telecommunications Business** A. Valdar and I. Morfett
Volume 65 **Advances in Body-Centric Wireless Communication: Applications and state-of-the-art** Q.H. Abbasi, M.U. Rehman, K. Qaraqe and A. Alomainy (Editors)
Volume 67 **Managing the Internet of Things: Architectures, theories and applications** J. Huang and K. Hua (Editors)
Volume 68 **Advanced Relay Technologies in Next Generation Wireless Communications** I. Krikidis and G. Zheng
Volume 69 **5G Wireless Technologies** A. Alexiou (Editor)
Volume 70 **Cloud and Fog Computing in 5G Mobile Networks** E. Markakis, G. Mastorakis, C.X. Mavromoustakis and E. Pallis (Editors)
Volume 71 **Understanding Telecommunications Networks, 2nd Edition** A. Valdar
Volume 72 **Introduction to Digital Wireless Communications** H.-C. Yang
Volume 73 **Network as a Service for Next Generation Internet** Q. Duan and S. Wang (Editors)
Volume 74 **Access, Fronthaul and Backhaul Networks for 5G Beyond** M.A. Imran, S.A.R. Zaidi and M.Z. Shakir (Editors)
Volume 76 **Trusted Communications with Physical Layer Security for 5G and Beyond** T.Q. Duong, X. Zhou and H.V. Poor (Editors)

Volume 77 **Network Design, Modelling and Performance Evaluation** Q. Vien
Volume 78 **Principles and Applications of Free Space Optical Communications** A.K. Majumdar, Z. Ghassemlooy, A.A.B. Raj (Editors)
Volume 79 **Satellite Communications in the 5G Era** S.K. Sharma, S. Chatzinotas and D. Arapoglou
Volume 80 **Transceiver and System Design for Digital Communications, 5th Edition** S.R. Bullock
Volume 81 **Applications of Machine Learning in Wireless Communications** R. He and Z. Ding (Editors)
Volume 83 **Microstrip and Printed Antenna Design, 3rd Edition** R. Bancroft
Volume 84 **Low Electromagnetic Emission Wireless Network Technologies: 5G and beyond** M.A. Imran, F. Héliot and Y.A. Sambo (Editors)
Volume 86 **Advances in Communications Satellite Systems Proceedings of the 36th International Communications Satellite Systems Conference (ICSSC-2018)** I. Otung, T. Butash and P. Garland (Editors)
Volume 89 **Information and Communication Technologies for Humanitarian Services** M.N. Islam (Editor)
Volume 92 **Flexible and Cognitive Radio Access Technologies for 5G and Beyond** H. Arslan and E. Başar (Editors)
Volume 93 **Antennas and Propagation for 5G and Beyond** Q. Abbasi, S.F. Jilani, A. Alomainy and M.A. Imran (Editors)
Volume 95 **ISDN Applications in Education and Training** R. Mason and P.D. Bacsich

This page intentionally left blank

Green Communications for Energy-Efficient Wireless Systems and Networks

Edited by
Himal A. Suraweera, Jing Yang,
Alessio Zappone and John S. Thompson

The Institution of Engineering and Technology

Published by The Institution of Engineering and Technology, London, United Kingdom

The Institution of Engineering and Technology is registered as a Charity in England & Wales (no. 211014) and Scotland (no. SC038698).

First published 2020

The Institution of Engineering and Technology
Michael Faraday House
Six Hills Way, Stevenage
Herts, SG1 2AY, United Kingdom

www.theiet.org

British Library Cataloguing in Publication Data
A catalogue record for this product is available from the British Library

ISBN 978-1-83953-067-8 (hardback)
ISBN 978-1-83953-068-5 (PDF)

Typeset in India by MPS Limited
Printed in the UK by CPI Group (UK) Ltd, Croydon

Contents

About the editors

Himal A. Suraweera is a senior lecturer in the Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka. He is in the editorial boards of *IEEE Transactions on Communications and Transactions on Green Communications and Networking*. His research interests include cooperative communications systems, energy harvesting communications, full-duplex wireless techniques, massive MIMO systems, communication theory and machine learning. He is a senior member of the IEEE.

Jing Yang is an assistant professor in the Department of Electrical Engineering at the Pennsylvania State University, USA. She is an editor of the *IEEE Transactions on Wireless Communications and IEEE Transactions on Green Communications and Networking*. Her research interests lie in wireless communications and networking, statistical learning and inference, and information theory.

Alessio Zappone is a tenured professor at the University of Cassino and Southern Lazio, Cassino, Italy. Previously he was with TU Dresden and CentraleSupélec in Paris, France. His research focuses on communication theory, energy-efficient communications, and radio resource allocation. He serves as a senior area editor for the *IEEE Signal Processing Letters* and has served as a guest editor for the *IEEE Journal on Selected Areas on Communications*.

John S. Thompson is a professor at the School of Engineering in the University of Edinburgh, United Kingdom. He leads UK research projects that study new concepts for fifth-generation wireless communications. He specializes in antenna array processing, cooperative communications systems, and energy-efficient wireless communications. In January 2016, he was elevated to the Fellow of the IEEE for contributions to antenna arrays and multi-hop communications.

This page intentionally left blank

Chapter 1

Introduction

Himal A. Suraweera[1], Jing Yang[2], Alessio Zappone[3] and John S. Thompson[4]

Inventions made in the last century laid the ground-work for the development of wireless communications; one of the largest sectors of the telecommunications industry. At present, there are more mobile connections than there are people on Earth, while wireless systems and devices such as smartphones have penetrated all sectors of the society at an unprecedented scale. As such, energy consumption has become a significant concern for green wireless systems operation [1]. Traditionally, wireless system's design has focused on performance optimization such as maximizing the spectral efficiency, throughput and minimizing the end-to-end communication latency. On the other hand, energy efficiency (EE) of wireless communications, which was mostly overlooked in the operation of previous generations of wireless systems, is now a key figure of merit [2]. Over the past few years, telecommunication operators across the world have seen their revenues eroding, while infrastructure, operation and maintenance costs have increased. At the same time, research projects have found that the information and communications technology (ICT) industry is responsible for a major percentage of greenhouse gas emissions such as carbon dioxide. As more wireless networks and devices get connected every day, pollution levels will further rise, and it is essential to reduce harmful emissions to acceptable levels in order to act on the threat of global warming and climate change.

Over the years, all segments of the ICT industry, academia and other stakeholders have collaborated to find breakthrough solutions for increasing the network EE through several research initiatives such as the GreenTouch Consortium, Energy Aware Radio and neTwork tecHnologies (EARTH), Cognitive radio and Cooperative strategies for POWER saving in multi-standard wireless devices (C2POWER), sustainable cellular networks harvesting ambient energy (SCAVENGE) and

[1]Department of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya, Sri Lanka
[2]School of Electrical Engineering and Computer Science, The Pennsylvania State University, State College, PA, USA
[3]Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy
[4]Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, UK

energy-autonomous portable access points for infrastructure-less networks (PAINLESS). The mission of GreenTouch founded in 2010 was to deliver the architecture, specifications and road map to increase the EE by a factor of 1,000 compared to the 2010 reference network [3]. The consortium announced its final results in 2015. The European-funded EARTH project from 2010 to June 2012 looked at the EE in base stations and focused on 3GPP mobile broadband technologies in particular Long Term Evolution (LTE) [4]. The aim of the C2POWER project was to investigate on cognitive and cooperative strategies that can be extended to decrease the overall power consumption of mobile devices [5]. The SCAVENGE project* focuses on sustainable design of architectures, protocols and algorithms for 5G cellular systems and their components such as base stations, mobile devices and sensors by taking advantage of renewable sources. The goal of the PAINLESS project† is to demonstrate energy-neutral and infrastructure-less operation by adopting solutions such as holistic optimization of energy harvesting, optimized access and backhauling techniques and unmanned aerial vehicle (UAV) access points for the future generations of wireless networks.

Energy consumption is responsible for 20%–40% of the network operational expenditure, of which the majority is due to electricity consumption [6]. The primary power source of legacy networks is grid-supplied energy with backup power supplied from diesel generators. Current commercial grids operate with electricity produced using coal and other types of fossil fuels. In order to address the capacity, maintenance and upgrade issues of existing grids, modern smart grids are being rolled out. Further, in order to reduce the reliance on mains power, operators are increasingly moving toward green renewable technologies such as photovoltaic (PV) panels and fuel-cell generators. Other factors such as carbon market volatility, emission regulation, cost reduction in renewables and environmental change continue to accelerate this conversion. Green base station development is a major step toward implementing future sustainable wireless networks [7]. Equipped with local PV or wind generation capabilities, green base stations use battery storage systems, smart DC controllers, etc. to enable flexible and energy-efficient operation of radio equipment. Air cooled, outdoor and light-weight base stations can remove the need for air conditioning and, coupled with renewable sources, are an energy solution, in particular to provide connectivity in rural areas.

Power models help one to estimate the realistic energy consumption of individual components within different types of base stations (e.g., macro, micro, pico and femto). Power amplifier (PA), signal processing circuits, analog-to-digital (A/D) converter (ADC), antenna, feeders, power sources and cooling are responsible for different power consumption figures [8]. Furthermore, power consumption depends on the base station mode (operational or idle) and, if operational, also on conditions such as high/low traffic. Figure 1.1 shows the power consumption breakdown of macro and micro base stations at maximum load [9]. At variable load, base station power consumption mainly depends on the PA consumption. The scaling overload is

*http://www.scavenge.eu/
†http://painless-itn.com/

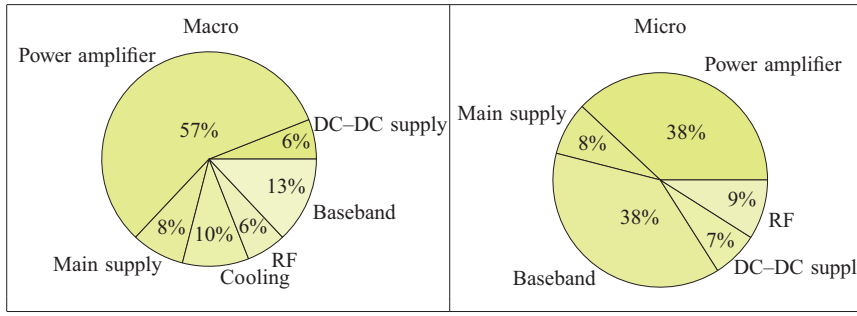


Figure 1.1 Base station power consumption breakdown at maximum load

significant especially for macro base stations as the PA is responsible for 57% of the overall consumed power at full load.

As connectivity demands continue to increase at an exponential rate, new services pose more constraints on the performance that end users expect. Between 2020 and 2030, mobile traffic will rise by 55% annually, reaching 607 exabytes in 2025 and 5,016 exabytes in 2030 [10]. LTE is currently the fastest mobile technology available, which can only support peak data rates of up to 300 Mbit/s. On the other hand, 5G wireless networks featuring innovative technologies such as infrastructure densification, antenna densification and the use of frequency bands in the millimeter wave (mmWave) range, which promise to achieve 1,000× higher data-rates. Lower end-to-end latency of 5G will also make it ideal for supporting a wide range of real-time use cases such as critical Internet of Things (IoT). Such a huge data-rate increase and low latency must be achieved in an energy-sustainable way [11]. EE was already identified as a main target of 5G wireless networks, and indeed the goal was for 5G to provide 2,000× higher energy efficiencies.

In the literature, several approaches have been listed as useful for increasing the EE [3]. These approaches can be organized into four categories as follows:

Energy-efficient resource allocation: This technique refers to optimally allocating radio resources such as power and bandwidth to strike a balance between performance metrics and the required energy consumption. In many cases, by optimally allocating power, high EE can be achieved at the expense of moderate performance degradation.

Network design and deployment: According to this technique, network infrastructure is deployed to jointly optimize the covered area and energy usage. Since base station is a key point of network design, in order to promote EE, base station on–off techniques, antenna muting and antenna down tilting can be used.

Energy harvesting communications: While today's networks largely offer services to mobile users, 5G and beyond systems will also feature device-to-device (D2D) communications that will connect millions of sensor-like devices operating in diverse and harsh environments. A sustainable solution for powering up low cost, low energy sensors is to adopt energy harvesting [12]. Moreover, renewable energy

sources such as solar can be used in parts of the world where off-grid or poor-grid (in poor grids frequent system outages occur) base stations are predicted to grow in number.

Efficient hardware design: Hardware components should be designed to minimize the power consumption. A major source of power inefficiency in wireless transceivers is the PA. Some solutions to increase the PA efficiency aim at the device level, while other techniques such as Doherty amplifiers, envelope tracking, digital pre-distortion and peak power reduction schemes are also possible. Additionally, transceiver design with system-on-chip architectures is a low power consumption approach.

The first four sections of this chapter will deal with these four themes in turn. Finally, Section 1.5 concludes the chapter with an overview of the remaining chapters of this book.

1.1 Energy-efficient resource allocation

1.1.1 Energy-efficient performance metrics

Several performance metrics have been proposed to define the EE of a complex multiuser network where multiple nodes interact and interfere with one another. Nevertheless, two general approaches have been identified as far as defining the EE of a communication network is concerned.

- **Network benefit–cost ratio.** This approach defines the network EE as an energy-efficient benefit–cost ratio, wherein the cost is represented by the total power consumed in the network, whereas the benefit is represented by any measure that quantifies the reliable transfer of information in the network, e.g., the network capacity/achievable rate, bit error rate and outage capacity. The resulting metric is called global EE (GEE) and is the EE metric with the strongest physical meaning from a network-wide perspective.
- **Multi-objective approach.** One drawback of the GEE is that it does not allow tuning the individual energy efficiencies of the different network nodes. To address this issue, an alternative approach is to regard the EE of each individual node in the network as a different objective to maximize, thus performing a multi-objective resource allocation and maximizing a combination of all the EE values for the network. Several combining functions have been proposed, among which the most widely used are the weighted sum EE, defined as a weighted sum of the different EEs in the network, the weighted product EE, defined as an exponentially weighted product of the different EEs in the network, and the weighted minimum energy efficiency (WMEE), defined as a weighted minimum of the different EEs in the network. All three combining functions are able to describe (at least parts of) the energy-efficient Pareto boundary of the system, by varying the choice of the weights, with the WMEE being able to describe the complete Pareto boundary, by sweeping the weights of the combination [13].

1.1.2 Energy-efficient resource allocation methods

Compared to traditional resource allocation schemes, energy-efficient radio resource allocation involves the maximization of a fractional performance function, which, in turn, requires the use of a specific branch of optimization theory, named *fractional programming* [13]. Fractional programming provides a framework to maximize a fractional function with a concave numerator and a convex denominator, subject to convex constraints, by means of standard convex optimization methods. Direct use of fractional programming for EE maximization has been used in several instances of wireless communication networks. However, this method typically requires a prohibitive complexity to operate in interference-limited networks, because the presence of multiuser interference typically means that the numerators of the EE functions are not concave in terms of the resources to allocate. Since multiuser interference is a peculiar trait of present and future wireless networks, energy-efficient resource allocation must cope with the presence of multiuser interference. One simple approach is to employ orthogonal transmission schemes and/or interference neutralization techniques to fall back to the noise-limited regime. However, these approaches are not practical in large networks where many users must be served. Instead, a more useful approach is the development of energy-efficient optimization frameworks that extend fractional programming by coupling it with specific methods to handle the presence of multiuser interference. In this context, the most widely used framework is that of *sequential fractional programming*, which merges fractional programming with the tool of sequential optimization, also known as successive convex approximation, or majorization–minimization. Sequential fractional programming provides a systematic approach to extend fractional programming to interference-limited networks with affordable complexity while enjoying optimality properties. Sequential fractional programming has been successfully used to optimize the EE of many wireless networks employing 5G and beyond 5G technologies, e.g., cloud radio access network (C-RAN) and coordinated multipoint (also with multi-carrier transmissions) [14], multicell massive multiple-input–multiple-output (MIMO) systems [15], full-duplex systems [16], D2D communications [17], cell-free systems [18] and systems employing physical-layer security [19].

Other practical optimization methods for energy-efficient resource allocation in interference-limited systems consider similar successive approximation approaches, such as the successive pseudo-convex method from [20], which can reformulate a wide class of EE problems as a sequence of pseudo-convex programs. Similarly, in [21], fractional programming is merged with the weighted minimum mean square error algorithm to develop a practical energy-efficient resource allocation method.

As already mentioned, all previous approaches trade off optimality with complexity. Instead, recently, a novel global optimization framework has also been proposed that can converge to the globally best solution of a wide class of EE maximization problems, with a complexity that is significantly lower than general-purpose global optimization methods [22,23]. While this approach is not fast enough for online resource allocation, it provides an effective and efficient way of computing offline the maximum EE of a complex network where many users reuse the same resource block.

On the other hand, optimal energy-efficient resource allocation can be performed in dense networks exploiting stochastic geometry arguments [24].

Recently, the issue of energy-efficient resource allocation has become a major point also in the context of wireless networks empowered by reconfigurable intelligent surfaces (RISs) [25]. The use of RISs is a recent technological breakthrough that holds the potential to revolutionize the traditional approach to wireless network design and operation. RISs are planar structures made of special materials, known as metamaterials, which are not bound by conventional reflection and diffraction laws, but that instead can modify the phase and direction of the radio wave impinging on them in a fully customizable way. This enables an RIS to control the phase of the reflected/refracted signal. Moreover, the electromagnetic properties of the surface can be dynamically reconfigured, so that the effect of the RIS on the incoming radio waves can be adapted in real time in response to the sudden changes in the network and/or in the traffic demands. RISs can be deployed on the walls of buildings or can be used to coat objects in the environment between the communicating devices, which effectively makes the wireless channel a new variable to be optimized, besides the design of the transmitters and receivers. Compared to traditional antenna arrays, RISs have the advantage of granting a large amount of degrees of freedom to exploit, in a more energy-efficient and cost-efficient way, being composed of cheap and nearly passive reflecting elements. In [26], it is shown that an RIS is much more energy-efficient than traditional relaying schemes. Similar results are obtained in [27] with reference to the power consumption of RIS-based wireless networks.

On the other hand, the design of an RIS-based wireless network is more involved because it requires the design of the RIS. In other words, the use of RISs provides new free parameters that can be optimized to improve the performance, but this comes at the expense of a more difficult resource allocation problem to be solved, especially if a large number of RIS are employed. A promising approach that might ease the computational burden of the design of RIS-based wireless networks is the use of artificial intelligence-based resource allocation. Indeed, artificial neural networks have been shown to be able to learn the map between the parameters of a neural network and the corresponding optimal resource allocation to employ. The joint use of artificial neural networks and RISs forms the concept of smart radio environment, a paradigm that has recently emerged for future wireless networks [28,29].

1.2 Network design and deployment

Innovative solutions at network design and deployment stages can improve the EE of wireless systems.

1.2.1 Dense networks

In order to address the issue of explosively increasing number of 5G connections, wireless operators are moving toward densely deployed infrastructure [30]. For a long time, cellular system design has been based on the principles of hexagonal cell

structure, and capacity improvements were realized by uniformly splitting a macrocell into a number of small areas covered with small base stations. On the other hand, in cell densification architectures, a large number of infrastructure nodes per unit area which are connected to a centralized processing unit with optical fiber-based backhaul will be deployed. Moreover, heterogeneous nodes such as macro base stations, small cell base stations and relays will be activated to meet the traffic demands in different geographical areas of the network.

In principle, densely deployed infrastructure reduces the distance between the transmitters and mobile users considerably, and high data rates can be achieved with low transmit power, since the effect of path loss on signal attenuation can be minimized. Also, node cooperation and infrastructure sharing among different network operators may lead to substantial energy savings. However, there are several challenges related to the EE of dense networks [31]. Any uncoordinated dense deployment of transmitters at different sites can create interference and introduce a ceiling for the achievable EE with increased densification. Also, modeling the positions of densely deployed base stations and mobiles to capture real deployments is complicated in general. A powerful tool used in the literature to model dense networks is based on stochastic geometry, and mathematical models of *point processes* such as the Poisson point process, Matérn hard-core point process and cluster processes have been found useful. Such analytical models allow accurate understanding of the impact of interference, and thus network optimization in terms of energy consumption to deliver a certain performance becomes possible. Dense deployment of infrastructure demands the installation of a large number of nodes at different points in the network. Consequently, overall power consumption could increase. In addition, nodes should be connected with many miles of cables, and losses in them will increase the overall energy budget. Interconnections among users, remote radio heads (RRHs) and core network functions made possible through fronthaul and backhaul connections will also require additional signaling and processing power and will increase the energy consumption. To this end, mobile edge computing is a recently considered approach to reduce the core network traffic as the technology allows storing of frequently accessed information at selected local nodes. Moreover, integrated access and backhaul is a complement for dense deployment of street-level radio nodes or in sites where fiber access is not available or is cost inefficient.

1.2.2 Base station on/off switching

The base station is a major source energy usage in a cellular network. Many base stations are powered by off-grid diesel generators, while grid connections are expensive and unreliable. To this end, in order to reduce the energy usage, a cell can be put into sleep by switching off of lightly loaded base stations or by discontinuous transmission (DTX). In such situations, traffic can be divided to other base stations as required. Even if there is no user data in the cell, cellular standards require the base stations to transmit signaling information. For example, Wideband Code Division Multiple Access (WCDMA) systems continuously transmit a common pilot channel and a common control channel even in the idle mode and PA utilization remains high [32]. With

5G, there are new tools to reduce energy such as devices that are allowed to spend most of their time in idle or deep sleep states and provide different levels of monitoring granularity (short and long DTX configurations). In addition, wake-up schemes can reduce the power consumption of DTX even further [33]. In wake-up schemes, the receiver periodically monitors a wake-up signal to activate itself. The wake-up signal can be optimized to achieve a desired delay-energy consumption trade-off.

Turning off underutilized base stations can preserve energy while guaranteeing quality-of-service requirements of the mobile users [34]. Antenna muting can also be used to reduce the power consumption of networks. In particular, depending on the user requirements and channel conditions, some antennas can be switched off at the cost of marginal performance degradation. Electrical or mechanical antenna down tilting can reduce the interference on neighboring cells for improved EE. Cell zooming or *cell breathing* is a technique that can be applied to fill the coverage gaps when some base stations are switched off [35]. It adjusts the cell size according to the prevalent traffic conditions in the network. To find the ideal cell size, cell zooming increases the transmit power of the active base stations or adjusts the antenna height and tilt angle. Accordingly, consumed energy of the active base stations in the network increases; however, additional energy savings can be achieved when cell zooming and sleeping strategies are used in combination.

1.2.3 Massive MIMO

MIMO technology has been integrated into modern wireless systems such as it can provide multiplexing and diversity gains. In particular, installing multiple antennas at the base station due to available space is possible, and beamforming techniques can be used to improve the performance. A fairly new development in MIMO is the introduction of massive MIMO technology for 5G [36]. Massive MIMO exploits the large number of degrees of freedom available to the system due to the use of a massive antenna array to serve multiple users under favorable propagation conditions. Moreover, it has been shown that even with linear signal processing significant capacity gains can be achieved. A characteristic of massive MIMO is its ability to form "pencil-sharp" beams, and thus energy can be beamformed into user locations with minimal inter-user inference. Therefore, massive MIMO is widely regarded as a green solution to realize modern wireless systems. With imperfect channel knowledge at the transmitter and for a fixed rate, a single-cell massive MIMO system can reduce the radiated power by a factor proportional to the square root of the number of antennas [37]. Figure 1.2 shows the massive MIMO testbed at Lund University, Sweden, which was the first real-time experimental platform demonstrating the technology.

However, massive MIMO system implementation comes with several challenges. The large number of required antennas and transceiver chains could increase the hardware-consumed power. Moreover, pilot contamination is an issue to be concerned with at the design phase. EE of massive MIMO systems under various system/channel parameters and with different uplink and downlink processing schemes has been studied in many papers. One needs to construct an accurate power consumption model when optimizing the energy consumption in massive MIMO systems. For example, several optimization variables, namely number of base station antennas, active users,

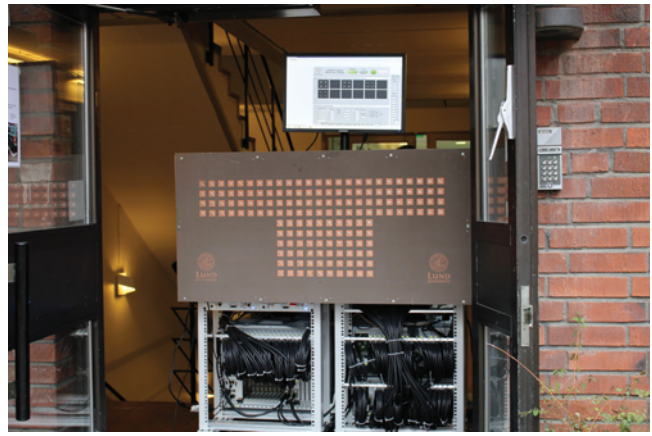


Figure 1.2 The massive MIMO testbed at Lund University, Sweden has been used to validate reciprocity-based beamforming using 100 antennas and up to 12 users, in both static and mobile scenarios (Image credit: Emil Björnson)

transmit power, base station density and pilot reuse factor, influence the EE of massive MIMO deployment. Optimal design of energy-efficient multiuser MIMO is investigated in [38] and contrary to the belief that the transmit power should decrease, it has been shown that the transmit power should increase when the number of antennas is increased, since the increase in circuit power can be compensated for.

1.2.4 mmWave cellular systems

There are opportunities to harness abundant spectrum at mmWave frequencies for increasing 5G data rates and capacity. For example, 26 and 28 GHz bands suit dense 5G small cell networks, while the multi-gigabit Wi-Fi technology, 802.11, can utilize the 60 GHz band and the lightly licensed E-band (70/80 GHz) is suitable for HetNet backhaul/fronthaul deployment. At such high frequencies, higher path loss, rain absorption and blockage effects from hand, body, walls, etc. adversely influence the propagation conditions. Factors such as amplifier efficiency, antenna complexity, A/D and digital-to-analog converters pose challenges in achieving the required link budget at mmWave frequencies, while beamforming and directional architectures allow increased gain. Techniques such as device-assisted power saving, reducing blind decoding of control channels using carrier aggregation and wake-up signals based on beamforming can also be employed for improving the power efficiency in 5G mmWave devices. As an example, mobile devices can provide additional information on battery level, antenna, orientation of the mobile, etc. to select optimal power and enable efficient beamforming/switching. In the literature, the suitability of the mmWave and massive MIMO combination has been studied [39]. However, implementing mmWave massive MIMO based on a fully digital architecture is an energy

inefficient and costly affair. In the near term, the hybrid analog/digital beamforming approach is a potential solution as energy-efficient fully digital architectures will take time to develop.

1.2.5 Cloudification and virtualization

The concept behind network cloudification in the form of the C-RAN architecture was first proposed by the China Mobile Research Institute as C-RAN in 2010. C-RAN is a centralized cloud computing architecture with an integrated baseband unit (BBU) pool and distributed RRH equipment connected by high-speed optical fiber. The BBU pool capable of processing baseband signals functions as a cloud data center solution, while RRHs perform signal modulation and amplification and are located at the cell sites. Since early 2019, solutions such as the Nokia AirScale C-RAN has been deployed. Cloudification ushers in new opportunities to improve the network EE [40]. For example, the mix of traditional and cloud base stations can be optimized for cost-effective, energy-aware high-performance deployments in the context of 5G. Furthermore, network function virtualization allows clouds to be virtualized to create end-to-end network slices consolidated in common physical infrastructure. Each network slice can be customized from an energy-efficient point of view to deliver heterogeneous services in parallel for new deployment cases such as industrial IoT, intelligent transportation systems, public safety and positioning.

In addition to the network, computing and storage servers, cloud management system and appliances (operating systems, platforms, applications used by end users) that make up a cloud computing architecture can benefit from energy-efficient features, including dynamic voltage and frequency scaling, server enclosure design, scheduling and load balancing of physical and virtual machines and the development of energy-efficient software [41]. When migrating to cloud computing systems, open access models such as The Cloud Energy and Emissions Research Model[‡] developed by Lawrence Berkeley National Laboratory and Northwestern University are useful for assessing the energy savings.

1.2.6 Offloading techniques

In the context of 5G, offloading will allow operate to free up network capacity and increase the EE. Mobile devices continue to become more capable and already they can connect to multiple cellular, Wi-Fi, Bluetooth networks as required. Hence, macro cellular traffic can be offloaded when users are in areas of small cells or indoor coverage. In addition to popularly used Wi-Fi, D2D communication and visible light communication (VLC) are also helpful to implement offloading techniques in wireless networks. D2D communication enables the possibility of direct transmission between two devices with low transmit power instead of using a cellular connection. VLC uses high-speed light-emitting diodes and photodiodes at the transmitter and receiver, respectively, for direct modulation of light [42]. LiFi, a form of VLC, can transmit at multiple gigabits offering latency values several times lower than Wi-Fi. Base stations

[‡]http://cleermodel.lbl.gov/

or access points in terrestrial networks are fixed entities, and there are many situations where they cannot be optimally used for offloading tasks. In this light, a promising solution that has received significant attention is the use of UAVs for communication and computing [43]. The mobility and high maneuverability of drones make them ideal candidates to be deployed in crowded areas such as stadiums, concerts or in disaster recovery applications to off-load cellular traffic. Moreover, encouraged by the availability of strong line-of-sight air-to-ground links, a new aerial edge computing paradigm where computational tasks of terrestrial users can be off-loaded to overhead drones has emerged. However, drones have limited on-board energy, and all offloading tasks should be completed within a limited period through solutions such as optimum trajectory planning and resource allocation.

Off-loading of computation, communication and storage operations from the main processor of a mobile device is also a way of improving its battery life. Off-loading at the processor level can be performed using approaches such as running of lightweight applications in secondary processors on behalf of the main processor, offloading storage to a less resource constrained neighboring device and opportunistic offload into a cloud server.

1.3 Energy harvesting communications

Sensor networks equipped with energy harvesting devices have attracted great attention recently. Compared with conventional sensor networks powered by batteries, energy harvesting capabilities of nodes make sustainable and environment-friendly sensor networks possible. However, the random, scarce and nonuniform energy supply features also necessitate a completely different approach to energy management.

According to the source types, energy harvesting communications can be divided into the following two groups:

Ambient energy harvesting: This technique refers to harvesting energy from natural sources such as solar, wind and ocean waves. The intermittent and seasonal nature of such natural energy sources pose challenges for energy-efficient system design due to uncertainty of key parameters.

RF energy harvesting: In these techniques, radio waves in the environment or dedicated transmitters such as power beacons are used to transfer energy. In this context, interference normally considered as an unwanted signal can also provide benefits for energy harvesting. Further, information-carrying signals can be designed for energy harvesting, resulting in the paradigm of simultaneous wireless information and power transfer (SWIPT) [44].

The main element of an RF energy harvesting node is the rectifying antennas (rectennas), while the end-to-end efficiency of a wireless-powered system depends on several individual components such as the antenna, the RF-to-DC rectifier and the power management circuit [45]. Two main architectures for designing RF-power and information receivers are the antenna-switching architecture and the colocated architecture. In the antenna-switching architecture, separate antennas are used for RF power harvesting and the information reception, while in the colocated receiver,

both the tasks are accomplished using a single-shared antenna. The colocated receiver architecture can be designed based on the time-switching or power-splitting model. A time-switching receiver switches between the energy harvester and the information receiver, while a power-splitting receiver separates the received RF signal into two streams for the energy harvesting and the information receiver.

A single-node energy harvesting communication system is illustrated in Figure 1.3. It is equipped with an energy harvesting module that converts ambient energy to electrical energy, which is stored in a rechargeable battery, and will be used to power the communication operations of the sensor. Therefore, energy harvesting communications are subject to the so-called *energy causality constraint* imposed by the energy harvesting process, i.e., energy cannot be used before it is harvested.

The performance analysis and optimization of systems harvesting energy from natural sources or RF signals require an accurate model of the energy harvesting process. In the literature, recorded data has been fitted to Gamma, Weibull and log-normal distribution functions to provide statistical models for wind and solar energy [46]. In the case of wireless power transfer, deterministic models in the form of simple linear models and piecewise-linear models were proposed in the early literature on the topic, while advanced nonlinear models based on the parametric functions such as logistic (sigmoidal) have been proposed recently. The nonlinear models are better suited to model sensitivity and saturation issues of real hardware.

In the following, we briefly summarize several research directions on the energy management of energy harvesting communication networks, as well as the information theoretic limits of energy harvesting communications.

1.3.1 Information-theoretic characterization of energy harvesting channels

In energy harvesting communication channels, the cumulative energy expended cannot exceed the total energy harvested at each channel use. This is in contrast to the classical information theory setting, where a single average power constraint is imposed for the entire code word. Instead, the energy causality constraint imposes n power constraints on the code word, which dramatically increases the difficulty of characterizing the corresponding channel capacity. Here, n refers to the code word length measured in channel uses.

The channel capacity of the additive white Gaussian noise (AWGN) energy harvesting channel is characterized in [47]. It has shown that the capacity of the AWGN channel with an infinite-sized battery subject to energy harvesting constraints is equal

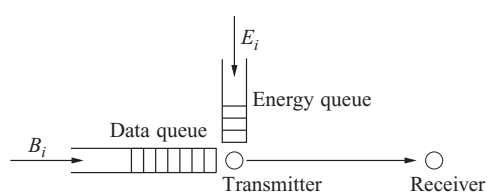


Figure 1.3 A single-node energy harvesting communication system

to the capacity of the same channel with an average power constraint equal to the average recharge rate of the battery. In particular, *save-and-transmit* and *best-effort* are proposed as capacity-achieving transmission strategies.

For a fixed tolerable error probability, [48] has performed a finite blocklength analysis of save-and-transmit proposed in [47] and obtained a non-asymptotic achievable rate for the AWGN energy harvesting channel. The first-, second- and third-order terms of the non-asymptotic achievable rate are also presented. Reference [49] has further improved the second-order term. For the block energy arrival model where the length of each energy block L grows sublinearly in n [49,50], [49] has proved that save-and-transmit achieves the optimal second-order scaling.

In [51], the first finite blocklength analysis of the best-effort scheme is provided for the AWGN energy harvesting channel. In addition, this work obtains a new non-asymptotic achievable rate for save-and-transmit, which outperforms the state-of-the-art result for save-and-transmit [49] in the high signal-to-noise ratio regime.

1.3.2 Offline energy management for throughput maximization

In the offline optimization framework, it is assumed that the energy harvesting profile is predictable and known in advance for the whole duration of operation. With such assumptions, energy has been managed to optimize the throughput of various communication systems. Noncausal knowledge of the energy harvesting process allows the optimal policy to be obtained through a one-shot optimization problem. The structural properties of the optimal policies have also been explicitly characterized in the literature.

For single-user energy harvesting communication channels, the optimal solution can be obtained through a *directional water-filling* [52] algorithm. It requires walls at the points of energy arrival with the right permeable water taps. The water taps allow water to flow only to the right, which implements the energy causality constraint. In addition, if a finite battery constraint ($E_{\max}$) is imposed, these taps allow at most $E_{\max}$ amount of water to flow to the right. The directional water-filling algorithm is based on the Karush–Kuhn–Tucker (KKT) optimality conditions for the corresponding convex optimization problem.

For broadcast channels with an energy harvesting transmitter, [53] shows that the optimal total transmit power is the same as the optimal single-user counterpart. Moreover, the distribution of the optimum total transmit power among users exhibits a cutoff structure, i.e., only the amount that is above this cutoff level is allocated to the weaker user. In [54], a *generalized iterative backward water-filling* algorithm is proposed to obtain the maximum throughput region for a multiple access channel with multiple energy harvesting transmitters. The algorithm combines directional water-filling together with generalized iterative water-filling in [55] to solve the optimization problem.

Similar approaches have been applied to obtain the optimal energy management schemes in interference channels in [56], systems with battery imperfections or processing costs [57].

1.3.3 Online energy management for performance optimization

In contrast to the offline setting, in the online optimization framework, it is assumed that the system knows the past realizations of the energy harvesting process but has only statistical knowledge of their future evolution. The major approach is to formulate the optimal energy management problem as a stochastic control problem, with the objective to determine the optimal decision rules so that the expected reward of the decisions is maximized. The reward could be data throughput, channel capacity, sensing utility, etc. With this approach, the energy harvesting process and/or the data arrival process are usually modeled as Markov processes, while the online problem can be cast under the powerful framework of Markov decision processes and solved numerically with standard dynamic programming tools, see, e.g., [52,58,59].

Modeling the energy replenishing process as a Markov process, [60] aims to maximize the time average reward by making decisions regarding whether to transmit or discard a packet based on the current energy level. The optimal policy is shown to have a threshold structure. Reference [61] studies the utility maximization problem of an energy harvesting sensor with finite battery capacity and data buffer. It proposes an adaptive energy management policy that decides the energy spent in each time slot based on the instantaneous battery level and data queue length and shows that it is asymptotically optimal.

In [62], the online sensing scheduling problem for an energy harvesting sensor is studied. The objective is to strategically select the sensing time such that the long-term time-average sensing performance is optimized, where the sensing performance depends on the time durations between two consecutive sensing epochs. It shows that when the battery size is infinite, a best-effort uniform sensing policy is optimal. For the finite battery case, an energy-aware adaptive sensing scheduling policy, which dynamically chooses the next sensing epoch based on the battery level at the current sensing epoch, is shown to be asymptotically optimal. When multiple energy harvesting sensor nodes are considered, [63] studies the dynamic activation of sensors with a unit battery in order to maximize the sensing utility. In [64], it is assumed that the sensing utility is a concave function of the number of active sensing nodes in each time slot. It shows that a randomized myopic policy, which aims to select a number of sensors with the highest energy levels to perform the sensing task in each slot, maximizes the long-term average utility and thus is an optimal approach.

A different online energy management approach is introduced in [65] and then extended for various system models. The policy uses a fixed fraction of the available energy for transmission in each time slot and is shown to perform within a constant gap from the optimal online policy.

1.3.4 Routing and resource allocation in multi-hop energy harvesting networks

While the aforementioned works focus on algorithms and protocols for point-to-point links or small-scale systems powered by energy harvesting, it is of paramount

importance to study the effects of energy harvesting on large-scale networks from the perspective of network design and operation. One critical aspect is routing and resource allocation methods that must cope with fluctuating energy harvesting processes in multiple-hop networks. For such problems, one common approach is to model the wireless network by a directed graph, where the set of vertices represents the sensor nodes and the set of edges represents the communication links between them. While the energy harvesting process at each node is modeled as a random process, the objective is to design online scheduling, routing and resource allocation schemes to manage the available energy and to optimize the network utility [66–69]. Commonly used tools include the standard dual decomposition and the subgradient methods [70] and the Lyapunov optimization technique [71].

In [66], an online algorithm named E-WME (Energy-opportunistic Weighted Minimum Energy) is proposed. The basic idea of the algorithm is to assign a cost to each node, which is an exponential function in its residual energy, and then use the shortest path routing with respect to this metric. It is shown that the E-WME achieves the asymptotically optimal competitive ratio. In [67], an algorithm called QuickFix is proposed to compute the data sampling rate and routes. The algorithm is based on the dual decomposition and subgradient method [70]. In order to cope with the fluctuations in the recharging process, a local algorithm called SnapIt is designed to adapt the sampling rate and maintain the battery charge at a target level. Reference [68] leverages the Lyapunov optimization technique [71] to design an online algorithm called the energy-limited scheduling algorithm (ESA). ESA keeps track of the amount of energy left at the network nodes and makes power allocation decisions for packet transmissions. It shows that ESA achieves a utility that is within $O(\epsilon)$ of the optimal, for any $\epsilon > 0$, while ensuring that the network congestion and the required capacity of the energy storage devices are deterministically upper bounded by bounds of size $O(1/\epsilon)$. Similar approaches have been adopted in [69] to jointly control the data queue and battery buffer to maximize the long-term average sensing rate of an energy harvesting wireless sensor network.

1.4 Efficient hardware design

Wireless networks contain diverse hardware components, e.g., power sources, batteries, electronics, antennas and RF chains. Moreover, recent trends such as network densification and IoT have led to a major increase in the number of hardware components deployed. Different EE levels of hardware components pose challenges for the reduction of network wide total power [72].

A major part of the macro base station power budget is consumed by the RF PA. Two practical parameters of PAs are the linearity and efficiency. Transmit signals of modern wireless systems are based on multicarrier modulation, and they exhibit a high peak-to-average power ratio. Therefore, PAs should operate in the linear region by selecting a back-off to avoid out-of-band emissions. As a result of a large back-off, average output power is reduced and the PA efficiency is adversely affected. Consequently, power-added efficiency (PAE) that is a measure of how well

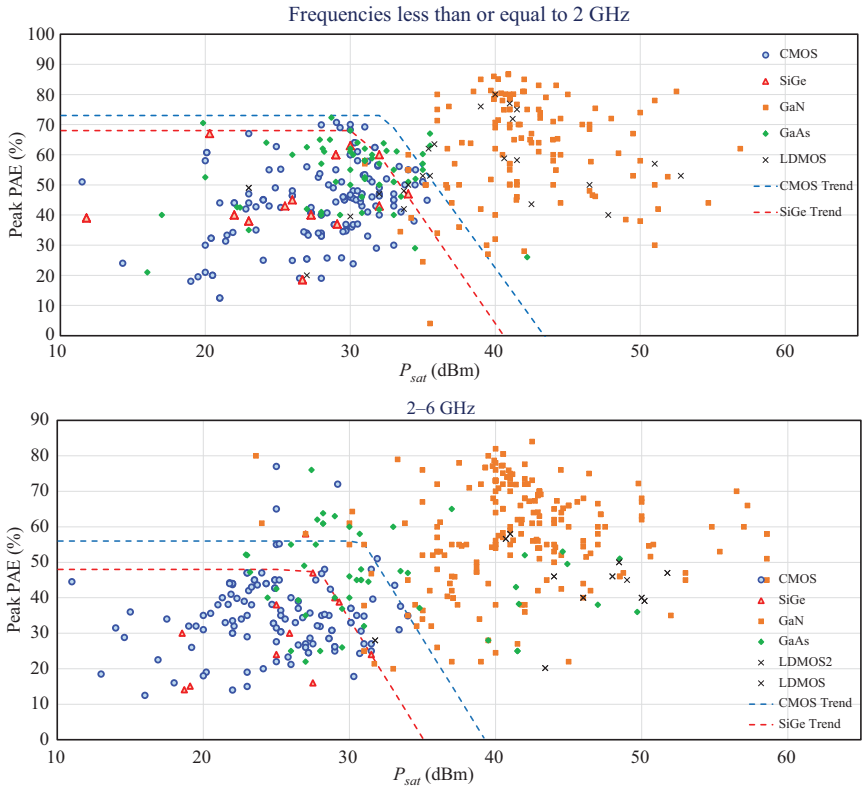


Figure 1.4 Peak PAE versus saturated output power (P_{sat}) for frequencies up to 6 GHz. Adapted from [73]

a PA can convert DC power into RF/microwave power can be improved using a technology such as complementary metal–oxide–semiconductor (CMOS), gallium nitride, silicon-germanium bipolar plus CMOS. Figure 1.4 shows the peak PAE versus saturated output power for different PA technologies in the frequency range up to 6 GHz [73].

Attention has also given to design simplified transmitter and receiver architectures, including ones that use coarse signal quantization. For example, in massive MIMO systems and mmWave systems, the use of 1-bit ADCs is an effective approach to increase the hardware EE. Coupling high-resolution ADCs with 1-bit ADCs, a mixed-ADC architecture for massive MIMO systems is proposed in [74]. The paper shows that the proposed energy-efficient solution can achieve a large fraction of the channel capacity of conventional architecture even with relatively a small number of high-resolution ADCs. There are also other cases where implementation of different hardware solutions within a system becomes a prospective solution to achieve a trade-off between the EE and the system performance. In mmWave systems, hybrid analog/digital beamforming is a solution to achieve a low energy cost. Unlike a fully

connected architecture that employs a large number of high-resolution phase shifters, the sub-connected architecture with low bit resolution ADC is advantageous from the power consumption point of view [75].

Most wireless protocols to date are implemented on custom hardware. However, programmable options offer potential advantages, and both professional and open source software are increasingly used to design and deploy wireless systems to multiple hardware targets. Moreover, software-defined radio (SDR) architectures realize multiple radio standards using reduced hardware at the RF front-end and baseband software. Therefore, hardware/software codesign continues to play a major role in the design of many energy-efficient wireless applications in cellular systems, sensor networks, etc. In low-power SDR platforms, hardware support for operations such as floating-point arithmetic is crucial to achieve a higher EE. Further, programmable features of SDR solutions combined with low power hardware are an attractive prospect to implement power management function in energy harvesting circuits.

There is also scope to optimize hardware usage for low power consumption in cloud-based network deployments. Since C-RAN allows virtualized radio functions to run in a distributed cloud, the technology offers flexibility to deploy hardware in the most energy-efficient manner while meeting the application needs. Moreover, energy harvesting at different nodes and locations is an sustainable approach for C-RAN. Since energy demands and availability vary across different geographical locations, energy storage using batteries will be the answer to balance intermittent generation and transportation of energy from one place to another. Engineers continue to pioneer ways of improving the battery technology. For a long time, lead–acid battery has served as the standard, while valve-regulated lead–acid batteries, lithium-ion batteries are used in telecom sites. However, high capacity, long-life batteries that can power base stations and mobile devices are yet to be developed. Innovative solutions with the potential to replace existing battery technologies, namely, lithium–sulfur, sodium-ion and graphene, hold promise for better performance and low environmental impact. Fuel cells have also received attention as a backup power source. A hydrogen fuel cell converts chemical energy stored in the element to electricity with water and heat being the only by-products. They are more effective than batteries, require limited maintenance and cause less pollution. However, the distribution of hydrogen is hazardous and prone to accidents. Hence, more stable fuels containing hydrogen, such as methanol, are being closely examined as alternatives.

1.5 Overview of the textbook

The remainder of this book is organized thematically into three major sections to capture major research advances in this field. To begin with, the focus is on new mathematical tools that are being developed to analyze green communications systems. In future networks, it will become increasingly important to find good solutions to very complex optimization problems that describe the behavior of communications systems and networks. The second chapter by *Matthiesen and Jorswieck* describes

the state of the art in this important area and presents a promising method entitled Mixed Monotonic Programming that can yield significant improvements for finding energy-efficient solutions. Recent advances in deep neural networks have shown promise for solving complex and ill-posed problems in many fields. The third chapter by *Zappone et al.* shows how machine learning can tackle the increasing complexity of wireless networks and complement existing model-based techniques for operating and controlling wireless networks. The fourth chapter by *Rusu and Thompson* explores efficient management techniques to try to minimize energy consumption in a wireless network. The approach is based on the concept of sensor management, which is an algorithmic technique for determining how to activate or switch off different base stations in order to achieve different performance objectives.

The second theme of the book deals with the use of renewable energy or energy harvesting techniques to make communications systems more sustainable. The chapter by *Meo and Renga* considers in detail how renewable energy sources can be used to power future wireless networks. They carry out extensive studies to show how issues of intermittent energy availability can be overcome, and new approaches to match energy supply and demand are described. SWIPT will also be a key technology for future IoT networks, and this topic is studied by *Kishk et al.* They use mathematical tools from stochastic geometry to analyze the performance of two networks: one is a single network using power and information transfer; the other involves power/information transfer network sharing resources with a second secure wireless networks. Next, the chapter by *Ko et al.* describes the related technology of backscatter communications, where a transmitter generates a carrier signal that powers a low energy tag and allows it to communicate back to the sender. A detailed description of these systems is provided, highlighting the current state of the art and future research directions. Finally, the age of information is currently a hot topic in the research community, which helps system designers to understand how timely information can be provided in a communication network. The chapter by *Arafa et al.* provides a detailed discussion of communications systems that relies on energy harvesting and how such systems may be designed to minimize the age property of data in a wireless network.

The third and final part of this book deals with research advances in EE relating to specific communications technologies. Massive MIMO technologies are a key enabler for 5G wireless networks, and assessing EE is crucial for developing sustainable future communications systems. The first chapter in this part by *Pizzo et al.* studies the EE of 5G in detail, developing increasingly realistic energy models that give insight into key performance trade-offs. The use of large-scale or massive MIMO antenna arrays is also a key technology for future wireless systems operating at high frequency millimeter wave bands (above 6 GHz). The chapter by *Buzzi and D'Andrea* studies the spectral efficiency and EE trade-offs and explores what is the best overall transceiver architecture for such systems. The next chapter by *Nguyen et al.* studies C-RAN where the baseband signal processing is separated from base station antennas, using optical fiber connections. Optimal beamforming strategies across all the antennas in a C-RAN system are typically too complex, so this chapter presents several suboptimal but promising approaches that can be applied in practical radio networks.

The chapter by *da Silva et al.* studies the performance of full-duplex interference cancellation techniques that can be implemented in a single transceiver system. A number of different network scenarios are evaluated, and the authors highlight under what conditions full-duplex cancellation can provide improvements in energy and spectrum efficiency. Another promising technology for 5G wireless networks is non-orthogonal multiple access (NOMA), which enables two or more users to share the same radio channel simultaneously, increasing the overall capacity. The chapter by *Wei et al.* studies new mathematical approaches that enable NOMA systems to operate in a highly energy-efficient manner. The final technical chapter by *Elgala et al.* discusses recent research on VLCs, which can offer a low-cost and energy-efficient alternative to radio frequency wireless communications. This chapter describes recent research advances on this technology to enable higher data rates and novel approaches of machine learning technology to achieve advanced transmitter and receiver designs. Chapter 15 concludes the book and also discusses future research directions.

References

[1] Hasan Z, Boostanimehr H, and Bhargava VK. Green Cellular Networks: A Survey, Some Research Issues and Challenges. IEEE Communications Surveys & Tutorials. 2011;13(4):524–540.

[2] Chen Y, Zhang S, Xu S, *et al.* Fundamental Trade-Offs on Green Wireless Networks. IEEE Communications Magazine. 2011;49(6):30–37.

[3] Buzzi S, Chih-Lin I, Klein TE, *et al.* A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. IEEE Journal on Selected Areas in Communications. 2016;34(4):697–709.

[4] Gruber M, Blume O, Ferling D, *et al.* EARTH – Energy Aware Radio and Network Technologies. In: 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications; 2009. pp. 1–5.

[5] Rodriguez J, Marques P, Radwan A, *et al.* Cognitive Radio and Cooperative Strategies for Power Saving in Multi-Standard Wireless Devices. In: 2010 Future Network & Mobile Summit; 2010. pp. 1–8.

[6] The GSM Association. Energy Efficiency: An Overview; 2019. Available from: https://www.gsma.com/futurenetworks/wiki/energy-efficiency-2/.

[7] Chamola V and Sikdar B. Solar Powered Cellular Base Stations: Current Scenario, Issues and Proposed Solutions. IEEE Communications Magazine. 2016;54(5):108–114.

[8] Arnold O, Richter F, Fettweis G, *et al.* Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks. In: 2010 Future Network Mobile Summit; 2010. pp. 1–8.

[9] Auer G, Blume O, Giannini V, *et al.* EARTH Deliverable D2.3: Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown; 2012.

[10] ITU. IMT Traffic Estimates for the Years 2020 to 2030. Rep ITU-R M2370-0; 2015. pp. 1–49.

[11] Wu Q, Li GY, Chen W, *et al.* An Overview of Sustainable Green 5G Networks. IEEE Wireless Communications. 2017;24(4):72–80.
[12] Lu X, Wang P, Niyato D, *et al.* Wireless Networks With RF Energy Harvesting: A Contemporary Survey. IEEE Communications Surveys & Tutorials. 2015;17(2):757–789.
[13] Zappone A and Jorswieck E. Energy Efficiency in Wireless Networks via Fractional Programming Theory. Foundations and Trends in Communications and Information Theory. 2015;11(3–4):185–396.
[14] Zappone A, Sanguinetti L, Bacci G, *et al.* Energy-Efficient Power Control: A Look at 5G Wireless Technologies. IEEE Transactions on Signal Processing. 2016;64(7):1668–1683.
[15] Zappone A, Björnson E, Sanguinetti L, *et al.* Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks. IEEE Transactions on Signal Processing. 2017;65(11):2844–2859.
[16] Nguyen D, Tran LN, Pirinen P, *et al.* Precoding for Full Duplex Multiuser MIMO Systems: Spectral and Energy Efficiency Maximization. IEEE Transactions on Signal Processing. 2013;61(16):4038–4050.
[17] Zappone A, Matthiesen B, and Jorswieck EA. Energy Efficiency in MIMO Underlay and Overlay Device-to-Device Communications and Cognitive Radio Systems. IEEE Transactions on Signal Processing. 2017;65(4):1026–1041.
[18] Buzzi S, D'Andrea C, Zappone A, *et al.* User-Centric 5G Cellular Networks: Resource Allocation and Comparison With the Cell-Free Massive MIMO Approach. IEEE Transactions on Wireless Communications. 2020;19(2):1250–1264.
[19] Zappone A, Lin PH, and Jorswieck EA. Optimal Energy-Efficient Design of Confidential Multiple-Antenna Systems. IEEE Transactions on Information Forensics and Security. 2018;13(1):237–252.
[20] Yang Y, Pesavento M, Chatzinotas S, *et al.* Energy Efficiency Optimization in MIMO Interference Channels: A Successive Pseudoconvex Approximation Approach. IEEE Transactions on Signal Processing. 2018;67(15):4107–4121.
[21] He S, Huang Y, Yang L, *et al.* Coordinated Multicell Multiuser Precoding for Maximizing Weighted Sum Energy Efficiency. IEEE Transactions on Signal Processing. 2014;62(3):741–751.
[22] Matthiesen B and Jorswieck EA. Efficient Global Optimal Resource Allocation in Non-Orthogonal Interference Networks. IEEE Transactions on Signal Processing. 2019;67(21):5612–5627.
[23] Matthiesen B, Zappone A, Besser K, *et al.* A Globally Optimal Energy-Efficient Power Control Framework and Its Efficient Implementation in Wireless Interference Networks. IEEE Transactions on Signal Processing. 2020;68:3887–3902.
[24] Di Renzo M, Zappone A, Lam TT, *et al.* System-Level Modeling and Optimization of the Energy Efficiency in Cellular Networks—A Stochastic Geometry Framework. IEEE Transactions on Wireless Communications. 2018;17(4):2539–2556.

[25] Di Renzo M, Zappone A, Debbah M, *et al.* Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and Road Ahead. IEEE Journal on Selected Areas in Communications. 2020.

[26] Huang C, Zappone A, Alexandropoulos GC, *et al.* Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communications. IEEE Transactions on Wireless Communications. 2019;18(8):4157–4170.

[27] Wu Q and Zhang R. Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming. IEEE Transactions on Communications. 2019;18(11):5394–5409.

[28] Zappone A, Di Renzo M, and Debbah M. Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both? IEEE Transactions on Communications. 2019;67(10):7331–7376.

[29] Zappone A, Di Renzo M, Debbah M, *et al.* Model-Aided Wireless Artificial Intelligence: Embedding Expert Knowledge in Deep Neural Networks for Wireless System Optimization. IEEE Vehicular Technology Magazine. 2019;14(3):60–69.

[30] Bhushan N, Li J, Malladi D, *et al.* Network Densification: The Dominant Theme for Wireless Evolution Into 5G. IEEE Communications Magazine. 2014;52(2):82–89.

[31] Ge X, Tu S, Mao G, *et al.* 5G Ultra-Dense Cellular Networks. IEEE Wireless Communications. 2016;23(1):72–79.

[32] Frenger P, Moberg P, Malmodin J, *et al.* Reducing Energy Consumption in LTE With Cell DTX. In: 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring); 2012. pp. 1–5.

[33] Rostami S, Lagen S, Costa M, *et al.* Optimized Wake-Up Scheme With Bounded Delay for Energy-Efficient MTC. In: 2019 IEEE Global Communications Conference (GLOBECOM); 2019. pp. 1–6.

[34] Feng M, Mao S, and Jiang T. Base Station ON-OFF Switching in 5G Wireless Networks: Approaches and Challenges. IEEE Wireless Communications. 2017;24(4):46–54.

[35] Piovesan N, Gambin AF, Miozzo M, *et al.* Energy Sustainable Paradigms and Methods for Future Mobile Networks: A Survey. Computer Communications. 2018;119:101–117.

[36] Larsson EG, Edfors O, Tufvesson F, *et al.* Massive MIMO for Next Generation Wireless Systems. IEEE Communications Magazine. 2014;52(2): 186–195.

[37] Ngo HQ, Larsson EG, and Marzetta TL. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. IEEE Transactions on Communications. 2013;61(4):1436–1449.

[38] Björnson E, Sanguinetti L, Hoydis J, *et al.* Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer? IEEE Transactions on Wireless Communications. 2015;14(6):3059–3075.

[39] Gao X, Dai L, Han S, *et al.* Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays. IEEE Journal on Selected Areas in Communications. 2016;34(4):998–1009.

[40] Alnoman A, Carvalho GHS, Anpalagan A, *et al.* Energy Efficiency on Fully Cloudified Mobile Networks: Survey, Challenges, and Open Issues. IEEE Communications Surveys & Tutorials. 2018;20(2):1271–1291.

[41] Mastelic T and Brandic I. Recent Trends in Energy-Efficient Cloud Computing. IEEE Cloud Computing. 2015;2(1):40–47.

[42] Komine T and Nakagawa M. Fundamental Analysis for Visible-Light Communication System Using LED Lights. IEEE Transactions on Consumer Electronics. 2004;50(1):100–107.

[43] Mozaffari M, Saad W, Bennis M, *et al.* A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. IEEE Communications Surveys & Tutorials. 2019;21(3):2334–2360.

[44] Ding Z, Zhong C, Ng DWK, *et al.* Application of Smart Antenna Technologies in Simultaneous Wireless Information and Power Transfer. IEEE Communications Magazine. 2015;53(4):86–93.

[45] Wagih M, Weddel AS, and Beeby S. Rectennas for RF Energy Harvesting and Wireless Power Transfer: A Review of Antenna Design. IEEE Antennas and Propagation Magazine. 2019.

[46] Afzaal MU, Sajjad IA, Awan AB, *et al.* Probabilistic Generation Model of Solar Irradiance for Grid Connected Photovoltaic Systems Using Weibull Distribution. Sustainability. 2020;12(6):2241. Available from: http://dx.doi.org/10.3390/su12062241.

[47] Ozel O and Ulukus S. Achieving AWGN Capacity Under Stochastic Energy Harvesting. IEEE Transactions on Information Theory. 2012;58(10):6471–6483.

[48] Fong SL, Tan VYF, and Yang J. Non-Asymptotic Achievable Rates for Energy-Harvesting Channels Using Save-and-Transmit. IEEE Journal on Selected Areas in Communications. 2016;34(12):3499–3511.

[49] Fong SL, Tan VYF, and Özgür A. On Achievable Rates of AWGN Energy-Harvesting Channels With Block Energy Arrival and Non-Vanishing Error Probabilities. IEEE Transactions on Information Theory. 2018;64(3):2038–2064.

[50] Shaviv D and Özgür A. Online Power Control for Block i.i.d. Energy Harvesting Channels. IEEE Transactions on Information Theory. 2018;64(8):5920–5937.

[51] Fong SL, Yang J, and Yener A. Non-Asymptotic Achievable Rates for Gaussian Energy-Harvesting Channels: Save-and-Transmit and Best-Effort. IEEE Transactions on Information Theory. 2019;65(11):7233–7252.

[52] Ozel O, Tutuncuoglu K, Yang J, *et al.* Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies. IEEE Journal on Selected Areas in Communications. 2011;29(8):1732–1743.

[53] Ozel O, Yang J, and Ulukus S. Optimal Transmission Schemes for Parallel and Fading Gaussian Broadcast Channels With an Energy Harvesting Rechargeable Transmitter. Computer Communications. 2013;36(12):1360–1372.

[54] Yang J and Ulukus S. Optimal Packet Scheduling in a Multiple Access Channel With Energy Harvesting Transmitters. Journal of Communications and Networks. 2012;14(2):140–150.

[55] Kaya O and Ulukus S. Achieving the Capacity Region Boundary of Fading CDMA Channels via Generalized Iterative Waterfilling. IEEE Transactions on Wireless Communications. 2006;5(11):3215–3223.

[56] Tutuncuoglu K and Yener A. Sum-Rate Optimal Power Policies for Energy Harvesting Transmitters in an Interference Channel. Journal of Communications and Networks. 2012;14(2):151–161.

[57] Devillers B and Gundüz D. A General Framework for the Optimization of Energy Harvesting Communication Systems With Battery Imperfections. Journal of Communications and Networks. 2012;14(2):130–139.

[58] Ho CK and Zhang R. Optimal Energy Allocation for Wireless Communications With Energy Harvesting Constraints. IEEE Transactions on Signal Processing. 2012;60(9):4808–4818.

[59] Mao S, Cheung MH, and Wong VWS. An Optimal Energy Allocation Algorithm for Energy Harvesting Wireless Sensor Networks. In: Proc. IEEE ICC; 2012.

[60] Lei J, Yates R, and Greenstein L. A Generic Model for Optimizing Single-Hop Transmission Policy of Replenishable Sensors. IEEE Transactions on Wireless Communications. 2009;8(2):547–551.

[61] Srivastava R and Koksal CE. Basic Performance Limits and Tradeoffs in Energy-Harvesting Sensor Nodes With Finite Data and Energy Storage. IEEE/ACM Transactions on Networking. 2013;21(4):1049–1062.

[62] Yang J, Wu X, and Wu J. Optimal Online Sensing Scheduling for Energy Harvesting Sensors With Infinite and Finite Batteries. IEEE Journal on Selected Areas in Communications. 2016;34(5):1578–1589.

[63] Kar K, Krishnamurthy A, and Jaggi N. Dynamic Node Activation in Networks of Rechargeable Sensors. IEEE/ACM Transactions on Networking. 2006;14(1):15–26.

[64] Yang J, Wu X, and Wu J. Optimal Scheduling of Collaborative Sensing in Energy Harvesting Sensor Networks. IEEE Journal on Selected Areas in Communications. 2015;33(3):512–523.

[65] Shaviv D and Özgür A. Universally Near Optimal Online Power Control for Energy Harvesting Nodes. IEEE Journal on Selected Areas in Communications. 2016;34(12):3620–3631.

[66] Lin L, Shroff NB, and Srikant R. Asymptotically Optimal Energy-Aware Routing for Multihop Wireless Networks With Renewable Energy Sources. IEEE/ACM Transactions on Networking. 2007;15(5):1021–1034.

[67] Liu R, Sinha P, and Koksal CE. Joint Energy Management and Resource Allocation in Rechargeable Sensor Networks. In: 2010 Proceedings IEEE INFOCOM; 2010. pp. 1–9.

[68] Huang L and Neely MJ. Utility Optimal Scheduling in Energy-Harvesting Networks. IEEE/ACM Transactions on Networking. 2013;21(4):1117–1130.

[69] Mao Z, Koksal CE, and Shroff NB. Near Optimal Power and Rate Control of Multi-Hop Sensor Networks With Energy Replenishment: Basic Limitations With Finite Energy and Data Storage. IEEE Transactions on Automatic Control. 2012;57(4):815–829.

[70] Boyd S and Vandenberghe L. Convex Optimization. New York, NY: Cambridge University Press; 2004.
[71] Neely MJ. Energy Optimal Control for Time-Varying Wireless Networks. IEEE Transactions on Information Theory. 2006;52(7):2915–2934.
[72] Chatzipapas A, Alouf S, and Mancuso V. On the Minimization of Power Consumption in Base Stations Using On/Off Power Amplifiers. In: 2011 IEEE Online Conference on Green Communications; 2011. pp. 18–23.
[73] Wang H, Wang F, Li S, *et al.* Power Amplifiers Performance Survey 2000-Present; 2020. Available from: https://gems.ece.gatech.edu/PA_survey.html.
[74] Liang N and Zhang W. Mixed-ADC Massive MIMO. IEEE Journal on Selected Areas in Communications. 2016;34(4):983–997.
[75] Ahmed I, Khammari H, Shahid A, *et al.* A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives. IEEE Communications Surveys Tutorials. 2018;20(4):3060–3097.

Part I

Mathematical tools for energy efficiency

This page intentionally left blank

Chapter 2

Optimization techniques for energy efficiency

Bho Matthiesen[1] and Eduard A. Jorswieck[2]

2.1 Introduction and motivation

Classical resource allocation, power, and flow control techniques for wireless communication networks comprise general network utility maximization (NUM) problems with constraints [1]. Common assumptions regarding the utility and constraint functions include that they are defined over the non-negative numbers, continuously differentiable, non-decreasing, and concave. The typical representatives of this class of utility functions are α-fair utility functions, including achievable rates, proportional fair rates, the negative delay, and more [2]. This NUM framework exploits the concavity of the utility and constraint functions in order to apply convex optimization methods [3] to efficiently solve them.

Energy efficiency (EE) is a key performance indicator in communication networks introduced about a decade ago [4]. It is motivated by an exponential increase of connected devices, significantly increased data rates, a rapid expansion of wireless networks, and finally due to ecological and economical concerns. Several different EE metrics have been proposed, each having its own merits and applications. For example, the global energy efficiency (GEE) measures the networks EE as a whole, while individual metrics like the weighted sum EE and weighted minimum EE focus more on the performance of individual devices. The analytical properties of EE utilities differ from classical α-fair utilities and require the development of new NUM frameworks: it is usually neither convex nor concave and increasing up to one point, called the mode, after which it decreases. The resulting NUM is a non-convex problem and the standard approaches cannot be applied.

Another reason for non-convex NUM problems are novel physical and medium access control layer technologies that support transmission of different links on the same time–frequency resources, e.g., non-orthogonal multiple access systems [5]. The resulting achievable rate expressions are neither increasing nor concave in the transmit power and, therefore, lead to even more involved non-convex resource allocation problems.

[1]Department of Communications Engineering, University of Bremen, Bremen, Germany
[2]Department of Information Theory and Communication Systems, Technische Universität Braunschweig, Braunschweig, Germany

The wireless engineer is interested in both global optimum solutions for non-convex NUM problems and suboptimal efficient algorithms. The first ones serve as ultimate benchmarks and upper bounds on the achievable performance, while the latter ones are compared with these bounds to justify their real-world implementations. This chapter provides the methods, tools, and algorithms to serve this demand. Please note that the monograph [6] covers the methodological background material of fractional programming theory and provides the basis for the motivating example and the review on fractional programming. However, the global solution to the non-convex NUM problems in interference networks is based on more recent results.

2.1.1 Motivating single-link examples

Let us motivate the development of the novel NUM framework by a simple example of a point-to-point link whose EE should be maximized. We introduce the EE of a point-to-point link with the general α-fair utility function for $\alpha \in [0, \infty)$ as

$$\mathrm{EE}(R) = \begin{cases} \frac{R^{(1-\alpha)}}{(1-\alpha)\cdot(\mu p + P_c)}, & \alpha \geq 0,\ \alpha \neq 1, \\ \frac{\log(R)}{\mu p(R) + P_c}, & \alpha = 1, \end{cases} \tag{2.1}$$

where R is the achievable rate, μ is the power amplifier inefficiency, $p(R)$ is the transmit power necessary to support a rate of R, e.g., $p(R) = 2^R - 1$, and $P_c > 0$ is the constant power consumption. As special cases, we obtain for $\alpha = 0$ the EE as defined in [7]. For $\alpha = 1$, we obtain the proportional fair EE [8].

Please note the following properties of the EE utility function in (2.1) for a fixed $\alpha \geq 0$: (1) the numerator of the EE in (2.1) is by assumption differentiable and concave with respect to R. (2) The denominator (2.1) is positive, differentiable, and convex with respect to R. The following statement follows from [6, Proposition 2.9].

Proposition 2.1. *For fixed $\alpha \geq 0$, the EE defined in (2.1) is a* pseudo-concave function *with respect to the achievable rate R.*

Pseudo-concave functions have important properties that allow one to develop algorithms for efficiently finding their global optimum [9,10]. Consider the following fractional programming problem:

$$\max_{R \geq 0} \frac{R^{(1-\alpha)}}{(1-\alpha)\cdot(\mu(2^R - 1) + P_c)}. \tag{2.2}$$

Since the utility function in (2.2) is pseudo-concave the Karush–Kuhn–Tucker (KKT) optimality conditions are necessary and sufficient for the global optimum. Further, there exist at least two simple ways to transform this fractional program. The first is based on Dinkelbach's Algorithm [9], where the problem in (2.2) is transformed into the following parametric program for $\lambda \geq 0$:

$$F(\lambda) = \max_{R \geq 0} \left(R^{(1-\alpha)} - \lambda(1-\alpha)\cdot(\mu(2^R - 1) + P_c) \right). \tag{2.3}$$

Then, the global optimum solution R^* of (2.2) corresponds to the argument of the maximization in (2.3) where $F(\lambda) = 0$. The second approach is the non-parametric approach outlined in [7].

In the simple case of only scalar fractional programming in (2.2), we can compute the global optimum by the KKT conditions in closed form. However, this does not generalize to multiple parallel point-to-point links to multi-carrier or to multiple antenna links.

2.1.2 Interference networks with treating interference as noise

Let us motivate the contributions of the following chapter further by a slightly more involved scenario in which two links operate on the same frequency and at the same time. Denote their channel coefficients as d_1 and d_2 for the two direct channels and as c_1 and c_2 for their cross-channels from transmitters 1 and 2, respectively. The achievable rate by treating interference as noise for link $i = \{1, 2\}$ is given by $j \neq i$:

$$R_i(p_1, p_2) = \log\left(1 + \frac{d_i p_i}{1 + c_j p_j}\right). \tag{2.4}$$

Using matrix inversion, the linear system of two equations (2.4) can be solved for $[p_1, p_2]$, which gives:

$$p_1(\tilde{R}_1, \tilde{R}_2) = \frac{\tilde{R}_1(c_2\tilde{R}_2 + d_2)}{d_1 d_2 - \tilde{R}_1\tilde{R}_2 c_1 c_2} \quad \text{and} \quad p_2(\tilde{R}_1, \tilde{R}_2) = \frac{\tilde{R}_2(c_1\tilde{R}_1 + d_1)}{d_1 d_2 - \tilde{R}_1\tilde{R}_2 c_1 c_2}, \tag{2.5}$$

with $\tilde{R}_i = 2^{R_i} - 1$. The following problem is to be solved for fixed $\alpha \geq 0$ to maximize the GEE:

$$\max_{R_1, R_2 \geq 0} \frac{\frac{R_1^{1-\alpha}}{1-\alpha} + \frac{R_2^{1-\alpha}}{1-\alpha}}{P_c + \mu \frac{(2^{R_1}-1)(2^{R_2}-1)(c_1+c_2)+(2^{R_1}-1)d_2+(2^{R_2}-1)d_1}{d_1 d_2 - (2^{R_1}-1)(2^{R_2}-1)c_1 c_2}}. \tag{2.6}$$

Even though the numerator is still jointly concave in R_1 and R_2, the denominator is a complicated function in R_1 and R_2 that is neither concave nor convex. We can transform problem (2.6) into the power domain using (2.4) and write it as the equivalent problem:

$$\max_{p_1, p_2 \geq 0} \frac{\sum_{i=1, j\neq i}^{2} \log\left(1 + \frac{d_i p_i}{1 + c_j p_j}\right)^{1-\alpha}}{(1-\alpha)(P_c + \mu p_1 + \mu p_2)}. \tag{2.7}$$

In (2.7), the denominator is a linear function in the optimization variables p_1, p_2. However, the numerator is neither concave nor convex in p_1, p_2.

If one looks at the problem from the information theoretic point of view, then the corresponding rate region achievable under a sum power constraint is neither convex

nor concave [11]. Hence, the combined rate $\mathbf{R} = [R_1, R_2]$ and power $\mathbf{p} = [p_1, p_2]$ allocation problem

$$\max_{\mathbf{R}\geq 0, \mathbf{p}\geq 0} \frac{\sum_{i=1}^{n} \frac{R_i^{1-\alpha}}{1-\alpha}}{P_c + \mu \sum_{i=1}^{n} p_i} \quad \text{s.t.} \quad R \in \mathscr{R}(\mathbf{p}), \tag{2.8}$$

where $\mathscr{R}(\mathbf{p})$ is the achievable rate region described by inequalities, is also a complicated non-convex programming problem.

In conclusion, it is not possible to directly apply classical fractional programming approaches that rely on convex optimization even to the considered simple two-user interference network. This applies for the optimization both in the rate domain and in the power domain. Therefore, we need new methods and tools in order to solve the corresponding global and individual EE problems. For a methodical approach to the system design, we are interested in the global optimal solution as well as suboptimal efficient algorithms. This chapter will develop the required method and tools and explain the corresponding optimization algorithms.

2.1.3 Overview and outline

After introducing the necessary notations, symbols, and terms, we will start by reviewing classical fractional programming theory with a main application to resource and power allocation problems in interference networks. We have already introduced terms such as pseudo-concavity that needs proper definition. Dinkelbach's Algorithm and its generalization are introduced. Standard convex programming approaches and feasibility checks are explained. Finally, the global extension of Dinkelbach to monotonic programming for non-convex utility functions is discussed. This will motivate the section on general global optimization methods, including branch-and-bound (BB), monotonic programming, the novel approach of mixed monotonic programming (MMP), and the successive incumbent transcending (SIT) scheme. For the development of suboptimal efficient algorithms, this chapter will conclude with first-order optimal algorithms. As a result, the reader knows the classical approach to fractional programming and its combination with global programming. The novel methods MMP and SIT and their algorithmic implementation are learned as well. Thereby, the reader will obtain a rich toolbox of modeling and programming algorithms for solving a wide class of relevant optimization problems focused on, but not limited to, EE maximization.

2.1.4 Notation

Boldface upper case and lower case letters denote matrices and vectors, respectively. $\|\mathbf{x}\|$, $\mathbf{x}^T$, $\mathbf{x}^H$ denote Euclidean norm, transpose, and conjugate transpose of the n-dimensional column vector $\mathbf{x} = \{x_i\}_{i=1}^n$. $\mathbf{0}_n$ and $\mathbf{1}_n$ denote an all-0 and an all-1 n-dimensional vector, respectively. Component-wise vector ordering is used, i.e., $\mathbf{x} \geq \mathbf{y}$ means $x_i \geq y_i$, for all $i = 1, \ldots, N$. The sets $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denote the sets of real and real non-negative numbers, respectively.

A function $f : \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}$ is *increasing* if $f(\mathbf{x}') \leq f(\mathbf{x})$ whenever $\mathbf{x}' \leq \mathbf{x}$, and *decreasing* if $-f$ is increasing. The gradient of f is denoted as ∇f.

Sets are typeset in calligraphic letters $\mathscr{A}$, and $\mathbb{R}$ denotes the set of real numbers. A set $\mathscr{G} \subseteq \mathbb{R}^n_{\geq 0}$ is said to be *normal* if for $0 \leq \mathbf{x}' \leq \mathbf{x}, \mathbf{x} \in \mathscr{G} \Rightarrow \mathbf{x}' \in \mathscr{G}$. A set $\mathscr{H} \subseteq \mathbb{R}^n_{\geq 0}$ is called *conormal* if $\mathbf{x} + \mathbb{R}^n_{\geq 0} \subseteq \mathscr{H}$ whenever $\mathbf{x} \in \mathscr{H}$ [12, Section 11.1.1], where $\mathbf{x} + \mathbb{R}^n_{\geq 0} = \{\mathbf{x} + \mathbf{a} \mid \mathbf{a} \in \mathbb{R}^n_{\geq 0}\}$. The diameter of $\mathscr{A}$ is denoted as $\operatorname{diam}(\mathscr{A})$ and is the maximum Euclidean distance of two points in $\mathscr{A}$. The projection of $\mathscr{A}$ onto the $\mathbf{x}$ coordinates is $\operatorname{proj}_{\mathbf{x}} \mathscr{A} = \{\mathbf{x} \mid (\mathbf{x}, \mathbf{y}) \in \mathscr{A} \text{ for some } \mathbf{y}\}$.

2.2 Fractional programming theory

Let us start this section by formally introducing the terms *pseudo-concavity* and *pseudo-convexity*. The interested reader is referred to [6, Chapter 2] for further explanations and information.

2.2.1 Pseudo-concavity

Definition 2.1 (pseudo-concavity). *Let $\mathscr{C} \subseteq \mathbb{R}^n$ be a convex set. Then $r : \mathscr{C} \to \mathbb{R}$ is* pseudo-convex *if and only if, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{C}$, it is differentiable and*

$$r(\mathbf{x}_2) < r(\mathbf{x}_1) \;\Rightarrow\; \nabla(r(\mathbf{x}_2))^T(\mathbf{x}_1 - \mathbf{x}_2) > 0. \tag{2.9}$$

Remark 2.1. *In a similar way, it is possible to define pseudo-convexity. In particular, if r is* pseudo-concave*, then $-r$ is pseudo-convex. Moreover, if r is both pseudo-concave and pseudo-convex, then it is called pseudo-linear.*

The strict version of the earlier-mentioned definition for strict pseudo-concavity replaces the first inequality in (2.9) with "$\leq$" and has to hold for all $\mathbf{x}_1 \neq \mathbf{x}_2$. The following property is important for maximization of pseudo-concave functions.

Proposition 2.2. *Let $r : \mathscr{C} \to \mathbb{R}$ be a pseudo-concave function.*

1. *If $\mathbf{x}^*$ is a stationary point for r, then it is a global maximizer for r.*
2. *If r is strictly pseudo-concave, then a unique global maximizer exists.*

For our application to fractional programming problems as motivated earlier, the following statements are very helpful and characterize a rich class of pseudo-concave functions.

Proposition 2.3. *Let $r(\mathbf{x}) = \dfrac{f(\mathbf{x})}{g(\mathbf{x})}$, with $f : \mathscr{C} \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : \mathscr{C} \subseteq \mathbb{R}^n \to \mathbb{R}_+$. Then:*

1. *If f is non-negative, differentiable, and concave, while g is differentiable and convex, then r is pseudo-convex. If g is affine, the non-negativity of f can be relaxed.*
2. *If f and g are affine, then r is pseudo-linear.*

2.2.2 Specific fractional programming problems

Definition 2.2 (concave–convex fractional program). *A concave–convex fractional program is stated as the optimization problem*

$$\begin{cases} \max\limits_{\mathbf{x}} & \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{s.t.} & c_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, I \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, J, \end{cases} \tag{2.10}$$

with f concave, differentiable, and non-negative, g convex, differentiable, and positive, c_i convex for all $i = 1, \ldots, I$, and h_j affine for all $j = 1, \ldots, J$.

The objective of (2.10) is pseudo-concave, and, as a consequence, each stationary point of the objective is a global maximizer. Moreover, the KKT conditions are necessary and sufficient for optimality. In addition, the numerator f of the objective is required to be non-negative.

Definition 2.3 (sum-of-ratios program). *A sum-of-ratios program is stated as the optimization problem*

$$\begin{cases} \max\limits_{\mathbf{x}} & \displaystyle\sum_{k=1}^{K} \frac{f_k(\mathbf{x})}{g_k(\mathbf{x})} \\ \text{s.t.} & c_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, I \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, J, \end{cases} \tag{2.11}$$

with f_k, concave and non-negative for all $k = 1, \ldots, K$, $\{g_k\}_{k=1}^{K}$ convex, and positive for all $k = 1, \ldots, K$, c_i convex for all $i = 1, \ldots, I$, and h_j affine for all $j = 1, \ldots, J$.

The sum of pseudo-concave functions is not necessarily a pseudo-concave function. Thus, contrary to the single-ratio fractional program mentioned earlier, the objective of (2.11) is, in general, not pseudo-concave and problem (2.11) might have multiple local extrema. Indeed, sum-of-ratio programs are considered to be among the most challenging continuous global optimization problems and known to be essentially NP-complete [13, Theorem 2].

2.2.3 Dinkelbach's Algorithm

Fortunately, there exists an elegant algorithm to approach single-ratio fractional programs that work very efficiently for pseudo-concave functions. It is known as Dinkelbach's Algorithm [9] and will be discussed in the following.

Dinkelbach's Algorithm has been introduced in [9,14]. It belongs to the class of parametric algorithms, whose basic idea is to tackle a concave–convex fractional program by solving a sequence of easier problems that converges to the global solution of the original concave–convex fractional program. The fundamental result upon

which Dinkelbach's Algorithm is built is the relation between the concave–convex fractional program and the function:

$$F(\lambda) = \max_{\mathbf{x}\in\mathscr{S}} \{f(\mathbf{x}) - \lambda g(\mathbf{x})\}, \tag{2.12}$$

with real variable λ.

Lemma 2.1. *Assume f and g are continuous, g is positive, and $\mathscr{S}$ is compact. Then, the function F enjoys the following properties:*

1. *F is convex for $\lambda \in \mathbb{R}$.*
2. *F is strictly monotone decreasing for $\lambda \in \mathbb{R}$.*
3. *$F(\lambda)$ has a unique root λ_0.*
4. *For any $\tilde{\mathbf{x}} \in \mathscr{S}$, $F(\lambda_{\tilde{x}}) \geq 0$, with $\lambda_{\tilde{x}} = \frac{f(\tilde{\mathbf{x}})}{g(\tilde{\mathbf{x}})}$, with equality when $\tilde{\mathbf{x}} = \arg\max_{\mathbf{x}\in\mathscr{S}} \{f(\mathbf{x}) - \lambda_{\tilde{x}} g(\mathbf{x})\}$.*

As a consequence of this statement, solving a fractional problem is equivalent to finding the unique zero of the auxiliary function $F(\lambda)$. Dinkelbach's Algorithm allows one to accomplish this. It is important to note that Algorithm 2.1 can be employed to solve any single-ratio fractional program regardless of convexity properties as long as the maximization step in line 3 can be implemented correctly, i.e., it must be solved globally in order to guarantee convergence and performance. If the optimization problem in line 3 can be solved efficiently, the iterative algorithm converges very fast. Indeed, it has super-linear convergence. It always converges to an ε'-optimal solution, i.e., the obtained $\mathbf{x}^*$ has a function value ε', close to the global optimum, $\frac{f(\mathbf{x}^*)}{g(\mathbf{x}^*)} + \varepsilon' \geq \max_{\mathbf{x}\in\mathscr{S}} \frac{f(\mathbf{x})}{g(\mathbf{x})}$, where ε' is related to the stopping criterion ε.

It is also possible to generalize Dinkelbach's Algorithm in order to solve min-max fractional programming problems [6, Section 3.3.1]. However, the sum-of-ratios fractional program described earlier cannot be solved by any Dinkelbach-type algorithm. In particular, the algorithm proposed in [15] does not converge to the global optimal solution [16].

Algorithm 2.1: Dinkelbach's Algorithm

1: $\varepsilon > 0$; $n = 0$; $\lambda_n = 0$;
2: **while** $F(\lambda_n) > \varepsilon$ **do**
3: $\mathbf{x}_n^* = \arg\max_{\mathbf{x}\in\mathscr{S}} \{f(\mathbf{x}) - \lambda_n g(\mathbf{x})\}$;
4: $F(\lambda_n) = f(\mathbf{x}_n^*) - \lambda_n g(\mathbf{x}_n^*)$;
5: $\lambda_{n+1} = \dfrac{f(\mathbf{x}_n^*)}{g(\mathbf{x}_n^*)}$;
6: $n = n + 1$;
7: **end while**

2.2.4 Variants of Dinkelbach's Algorithm

Depending on the type of inner optimization problem in line 3 of Algorithm 2.1, different variants are known. The easiest is when the numerator is concave and the denominator is convex. Then, the inner programming problem

$$\max_{\mathbf{x}\in\mathscr{S}} f(\mathbf{x}) - \lambda g(\mathbf{x}), \tag{2.13}$$

is convex if the constraint set $\mathscr{S}$ is a convex set. It can be solved with polynomial complexity under very mild assumptions using standard convex optimization tools [17].

Instead, if the same problem (2.13) does not fulfill the concavity and convexity constraint, e.g., like the objective function in the two-user interference channel example (2.7), it is still possible to apply Dinkelbach's Algorithm in combination with a global programming method like monotonic programming. This combination was first proposed in [18] and subsequently developed into a framework in [19]. The general approach is illustrated in Figure 2.1.

In Figure 2.1, the upper left side represents the classical fractional concave–convex program that can be efficiently solved by applying the Dinkelbach's Algorithm shown on the upper right hand side. If the function f is not concave, as is the case in the interference channel with treating interference as noise, the resulting fractional program can be still solved by Dinkelbach's Algorithm.

Consider then the following inner programming problem,

$$\max_{\mathbf{x}\in[\mathbf{a},\mathbf{b}]} f_1(\mathbf{x}) - f_2(\mathbf{x}) - \lambda g(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathscr{S}, \tag{2.14}$$

where f_1, f_2, g are monotonically increasing functions in $\mathbf{x}$ and $\mathbf{a} \leq \mathbf{b}$ and $\mathscr{S} = \mathscr{G} \cap \mathscr{H}$ with $\mathscr{G}$ normal and $\mathscr{H}$ conormal. In typical communications scenarios, such as the interference channel mentioned earlier, we can write the signal-to-interference-and-noise ratio from the log-term as a difference of two log-terms. Therefore, a

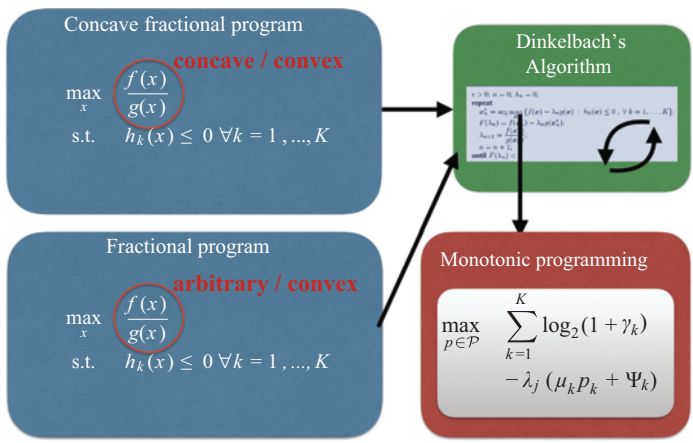


Figure 2.1 Illustration of Dinkelbach's Algorithm with monotonic programming

representation as in (2.14) is possible. In the EE expression, the function g is usually linear in the transmit power with non-negative coefficients.

Starting from the representation in (2.14), it is possible to introduce an auxiliary variable t and move the decreasing term $f_2 + \lambda g$ to the constraint set [12, Section 11.1.3], i.e.,

$$\begin{cases} \max\limits_{t,\mathbf{x}} & f_1(\mathbf{x}) + t \\ \text{s.t.} & f_2(\mathbf{x}) + \lambda g(\mathbf{x}) + t \leq 0 \\ & -f_2(\mathbf{b}) - \lambda g(\mathbf{b}) \leq t \leq -f_2(\mathbf{a}) - \lambda g(\mathbf{a}) \\ & \mathbf{x} \in \mathscr{S} \cap [\mathbf{a}, \mathbf{b}]. \end{cases} \tag{2.15}$$

This is a monotonic optimization problem and usually solved with the Polyblock algorithm [12,20,21]. Thus, the inner optimization problem in Dinkelbach's Algorithm can be solved with global optimality by monotonic programming as illustrated in Figure 2.1 on the lower right hand side. This approach is known as *Fractional Monotonic Programming* [19].

While this approach integrates nicely into existing algorithms and is coherent with the overall development and trends in global optimal resource allocation, it has the downside of having very high computational complexity. This is because for each outer Dinkelbach iteration a complete inner monotonic programming problem has to be solved. In addition, the complexity is even further increased due the additional auxiliary variable introduced in (2.15). Finally, the inner monotonic programming algorithm might not converge in finite time and, in that case, does not produce an ε-optimal solution.

Another option for combining Dinkelbach's Algorithm with another solution approach for general fractional programs is to apply a suboptimal solution technique like sequential convex programming. This combination was proposed in [22] to solve resource allocation problem for general interference networks and is discussed in Section 2.5. For this combination, it is very important to note again that the convergence of the Dinkelbach method is only guaranteed if the inner problem is solved globally. Therefore, the sequential convex approximation has to be performed outside, while the Dinkelbach's Algorithm runs as the inner optimization of the resulting concave–convex fractional program.

In Figure 2.2, the combination of outer convex sequential programming and inner Dinkelbach's Algorithm is illustrated. The general fractional program on the lower left-hand side is solved by iteratively computing convex lower bounds for the numerator, illustrated on the lower right-hand side, and then each convex–concave fractional program is solved by Dinkelbach's Algorithm shown on the right upper corner.

In this section, we have reviewed three ways to handle the inner optimization problems applying Dinkelbach's Algorithm for fractional problems. However, for the global optimum solution, we could not present an efficient implementation. This motivates the search for alternative global programming methods in the next section.

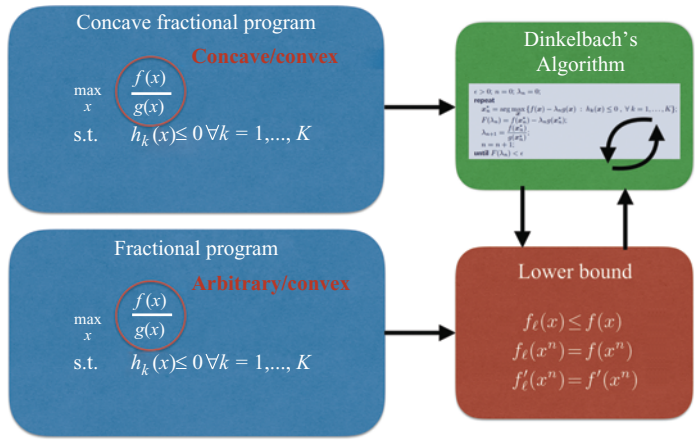


Figure 2.2 Illustration of combining Dinkelbach's Algorithm with convex sequential programming

2.3 Global optimization

Consider the optimization problem:

$$\max_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x}), \tag{2.16}$$

where $f : \mathscr{D} \to \mathbb{R}$ is a continuous function and $\mathscr{D}$ is a compact, non-empty subset of $\mathbb{R}^n$. A point $\mathbf{x}^*$ satisfying $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{D}$ is called a global maximizer of (2.16). The existence of such a point is guaranteed by the extreme value theorem [23, Theorem 4.16] under the assumptions shown earlier. If a point $\mathbf{x}'$ satisfies the optimality condition only within an open ε-neighborhood for some $\varepsilon > 0$, i.e., $f(\mathbf{x}') \geq f(\mathbf{x})$ for all $\mathbf{x} \in \{\mathbf{x} \in \mathscr{D} : \|\mathbf{x} - \mathbf{x}'\| < \varepsilon\}$, it is called a local optimizer.

Practical algorithms that solve (2.16) are, in general, unable to obtain the exact solution $\mathbf{x}^*$ within reasonable, or even finite, time. The standard approach is to compute an η-optimal solution, i.e., a point $\bar{\mathbf{x}}$ that satisfies:

$$f(\bar{\mathbf{x}}) \geq f(\mathbf{x}) - \eta \quad \text{for all } \mathbf{x} \in \mathscr{D}. \tag{2.17}$$

The issue with global optimization is that all known solution algorithms for (2.16) with polynomial computational complexity in the number of variables are, at most, capable of finding a local optimizer. For convex optimization problems, i.e., with f concave on $\mathscr{D}$ and $\mathscr{D}$ a convex set, and some generalized convex problems, this is not a problem because every local optimizer also solves (2.16) globally.

Global optimization theory deals with problems where this local–global property does not hold. This requires completely different algorithmic approaches than known from convex optimization theory. These usually scale with exponential complexity in the number of variables. Indeed, many non-convex optimization problems are known to be NP-hard, including many resource allocation problems in interference

networks [24]. Global optimization algorithms can be roughly categorized into two different classes: outer approximation (OA) algorithms and BB methods. The former approximates the feasible set by a sequence of simpler problems with a relaxed feasible set. A prominent instance is the Polyblock algorithm [20] mentioned earlier, which solves monotonic optimization problems. Instead, the general idea of BB algorithms is to relax and then successively partition the feasible set such that upper bounds on the objective value can be computed for each partition element.

Experience shows that BB procedures often outperform OA algorithms, especially for monotonic optimization in the context of energy-efficient resource allocation in interference networks. The reasons are manifold. First, and foremost, is the fact that the underlying data structures for BB can be implemented much more efficiently than for the Polyblock algorithm. Another equally important reason is that BB methods are much more flexible. The Polyblock algorithm and monotonic optimization in general take a very specific problem formulations as a basis, i.e., the canonical monotonic optimization problem:

$$\max_{\mathbf{x}\in[\mathbf{a},\mathbf{b}]} f(\mathbf{x}) \quad \text{s.t.} \quad g(\mathbf{x}) \leq 0 \leq h(\mathbf{x}), \tag{2.18}$$

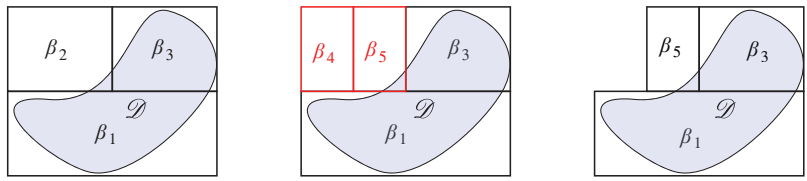
with f, g, h increasing functions. Casting a general optimization problem with hidden monotonicity into this form often requires cumbersome transformations, auxiliary variables (cf. (2.15)), and additional algorithms (e.g., Dinkelbach's Algorithm). Modifying such an OA algorithms is usually very hard or even impossible since it heavily relies on problem-specific characteristics. Instead, BB procedures are modular and one can often easily replace inefficient parts by approaches better suited to the problem at hand.

2.3.1 Branch and bound

BB algorithms are one of the most versatile and widely used tools in global optimization. The core idea is to relax the feasible set $\mathscr{D}$ and subsequently partition it. The partitioning method is chosen such that upper bounds on the objective value can be determined efficiently for each partition element. During the course of the algorithm, available or easily computable feasible points are evaluated, and the partition of $\mathscr{D}$ is successively refined. The algorithm terminates when each partition element is either proven to be empty, i.e., it does not contain any feasible points, or it does not contain any points with objective value greater than the current best known feasible solution. Part of this process is illustrated in Figure 2.3.

There are three basic types of subdivision: simplicial, conical, and rectangular. The choice of a subdivision procedure mostly depends on the bounding procedure and the structure of the problem to be solved. Rectangular subdivision is especially well suited if the primary bounding mechanism relies on exploiting monotonicity. This is because the extrema of an increasing function over the hyperrectangle:

$$[\mathbf{r}, \mathbf{s}] = \{\mathbf{x} \in \mathbb{R}^n : r_i \leq x_i \leq s_i, \text{ for all } i = 1, \ldots, n\}, \tag{2.19}$$



(a) (b) (c)

Figure 2.3 Illustration of rectangular subdivision in a BB procedure: (a) after 2 iterations, (b) in iteration 3, and (c) after iteration 3. The feasible set is denoted as $\mathscr{D}$, the bounds are β_i, and in (b) the upper left box is selected for branching

are at the lower left and upper right cornerpoints, respectively, i.e.,

$$\arg\min_{\mathbf{x}\in[\mathbf{r},\mathbf{s}]} f(\mathbf{x}) = \mathbf{r} \qquad \arg\max_{\mathbf{x}\in[\mathbf{r},\mathbf{s}]} f(\mathbf{x}) = \mathbf{s}. \tag{2.20}$$

Correctly exploiting this very simple property leads to bounds that can often be computed with low computational effort. As already discussed in Section 2.2.4, many utility functions encountered in energy-efficient resource allocation have exactly these hidden monotonicity properties. Therefore, we focus our exposition on rectangular BB methods and refer the interested reader to [12] for a treatment of the other two subdivision approaches.

Rectangular subdivision procedures start with a hyperrectangle $\mathscr{M}_0$ that contains the feasible set $\mathscr{D}$. Then, this initial rectangle is subdivided successively to create a partition of the feasible set. The subdivision of a rectangle $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ is defined by a tuple $(\mathbf{v}, j)$ where $\mathbf{v}$ defines the point the division is made through and j is the dimension of the cut, i.e., it results in the two rectangles:

$$\begin{aligned} \mathscr{M}^- &= \{\mathbf{x} \mid r_j^k \le x_j \le v_j^k,\ r_i^k \le x_i \le s_i^k\ (i \ne j)\} \\ \mathscr{M}^+ &= \{\mathbf{x} \mid v_j^k \le x_j \le s_j^k,\ r_i^k \le x_i \le s_i^k\ (i \ne j)\}. \end{aligned} \tag{2.21}$$

In its simplest form where $\mathbf{v} = \frac{1}{2}(\mathbf{r} + \mathbf{s})$ and j is the index of the longest side of $\mathscr{M}$, this is known as the standard bisection. It results in an exhaustive subdivision procedure that creates an infinite nested sequence of rectangles that converges to a singleton, i.e., a set with diameter zero.

A prototype rectangular BB algorithm is stated in Algorithm 2.2. In iteration k, the set $\mathscr{R}_k$ holds the partition elements of the feasible set that are not yet proven to be infeasible or suboptimal, $\bar{\mathbf{x}}^k$ and γ_k are the current best solution and value, respectively, $\mathscr{P}_k$ is an intermediate set holding the new boxes after branching, and $\beta(\mathscr{M})$ computes an upper bound on the objective values in $\mathscr{M}$, i.e., $\beta(\mathscr{M}) \ge \max_{\mathbf{x}\in\mathscr{M}} f(\mathbf{x})$. We say that this procedure is convergent if it can be infinite* only when $\eta = 0$ and

*A procedure is said to be infinite if it does not terminate.

then $\lim_{k\to\infty}\gamma_k = \max_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x})$. An exhaustive subdivision procedure is said to be consistent with branching if:

$$\beta(\mathscr{M}) - \max\{f(\mathbf{x}) \mid \mathbf{x}\in\mathscr{M}\cap\mathscr{D}\} \to 0 \text{ as } \operatorname{diam}(\mathscr{M}) \to 0. \tag{2.22}$$

We can formally establish convergence of Algorithm 2.2 under this condition.

Proposition 2.4. *If the subdivision procedure is exhaustive and the bounding operation satisfies (2.22), then Algorithm 2.2 is convergent and the final $\bar{\mathbf{x}}^k$ is a global η-optimal solution of (2.16).*

Proof. Please refer to [12, Proposition 6.1], [25, Proposition 3.7], or [26, Theorem 1]. □

Exhaustive rectangular subdivision is straightforward to be implemented because it is independent of the bounding operation. However, by coupling the subdivision with the bounding procedure, the convergence speed can be improved. Suppose that for each rectangle $\mathscr{M}_k$, two points $\mathbf{x}^k, \mathbf{y}^k \in \mathscr{M}_k$ are known such that:

$$\mathbf{x}^k \in \mathscr{M}_k \cap \mathscr{D} \quad \text{and} \quad \beta(\mathscr{M}_k) - f(\mathbf{y}^k) \to 0 \text{ as } \left\|\mathbf{x}^k - \mathbf{y}^k\right\| \to 0. \tag{2.23}$$

Then, we can subdivide $\mathscr{M}_k$ adaptively via $(\mathbf{v}^k, j_k)$ where $\mathbf{v}^k = \frac{1}{2}(\mathbf{x}^k + \mathbf{y}^k)$ and $j_k \in \arg\max_j |y_j^k - x_j^k|$. For example, consider the bounding operation $\beta(\mathscr{M}^k) = \max_{\mathbf{x}\in\mathscr{M}_k\cap\mathscr{D}} \bar{f}(\mathbf{x};\mathscr{M}_k)$ where $\bar{f}(\mathbf{x};\mathscr{M}_k)$ is an upper bound of f on $\mathscr{M}_k$ tight at some point $\mathbf{z}^k \in \mathscr{M}^k$, i.e., $\bar{f}(\mathbf{x}) \geq f(\mathbf{x})$ for all $\mathbf{x}\in\mathscr{M}_k\cap\mathscr{D}$ and $\bar{f}(\mathbf{z}^k) = f(\mathbf{z}^k)$. Then, choosing $\mathbf{y}^k = \mathbf{z}^k$ and $\mathbf{x}^k \in \arg\max_{\mathbf{x}\in\mathscr{M}_k\cap\mathscr{D}} \bar{f}(\mathbf{x};\mathscr{M}_k)$ satisfies (2.23). Instead of trying to minimize the size of each box, this rule strives to bring the bounding point $\mathbf{z}^k$ close to

Algorithm 2.2: Rectangular BB algorithm

Step 0 **(Initialization)** Choose $\mathscr{M}_0 \supseteq \mathscr{D}$ and $\eta > 0$. Let $k = 1$ and $\mathscr{R}_0 = \{\mathscr{M}_0\}$. If available or easily computable, find $\bar{\mathbf{x}}^0 \in \mathscr{D}$ and set $\gamma_0 = f(\bar{\mathbf{x}}^0)$. Otherwise, set $\gamma_0 = -\infty$.

Step 1 **(Branching)** Select a box $\mathscr{M}_k \in \arg\max\{\beta(\mathscr{M}) \mid \mathscr{M}\in\mathscr{R}_{k-1}\}$ and subdivide it via $(\mathbf{v}^k, j)$ (cf. (2.21)). Let $\mathscr{P}_k = \{\mathscr{M}_k^-, \mathscr{M}_k^+\}$.

Step 2 **(Reduction)** For each $\mathscr{M}\in\mathscr{P}_k$, replace $\mathscr{M}$ by $\mathscr{M}'$ such that $\mathscr{M}'\subseteq\mathscr{M}$ and

$$(\mathscr{M}\setminus\mathscr{M}')\cap\{\mathbf{x}\in\mathscr{D} \mid f(\mathbf{x}) > \gamma_k\} = \emptyset. \tag{2.24}$$

Step 3 **(Bounding)** For each $\mathscr{M}\in\mathscr{P}_k$, compute $\beta(\mathscr{M}) \geq \sup_{\mathbf{x}\in\mathscr{M}\cap\mathscr{D}} f(\mathbf{x})$. Find $\mathbf{x}\in\mathscr{M}\cap\mathscr{D}$ and set $\alpha(\mathscr{M}) = f(\mathbf{x})$. If $\mathscr{M}\cap\mathscr{D} = \emptyset$, set $\alpha(\mathscr{M}) = -\infty$.

Step 4 **(Incumbent)** Let $\alpha_k = \max\{\alpha(\mathscr{M}) \mid \mathscr{M}\in\mathscr{P}_k\}$. If $\alpha_k > \gamma_{k-1}$, set $\gamma_k = \alpha_k$ and let $\bar{\mathbf{x}}^k \in \mathscr{D}$ such that $\alpha_k = f(\bar{\mathbf{x}}^k)$. Otherwise, let $\gamma_k = \gamma_{k-1}$ and $\bar{\mathbf{x}}^k = \bar{\mathbf{x}}^{k-1}$.

Step 5 **(Pruning)** Delete every $\mathscr{M}\in\mathscr{P}_k$ with $\mathscr{M}\cap\mathscr{D} = \emptyset$ or $\beta(\mathscr{M}) \leq \gamma_k + \eta$. Let $\mathscr{P}'_k$ be the collection of remaining sets and set $\mathscr{R}_k = \mathscr{P}'_k \cup (\mathscr{R}_{k-1}\setminus\{\mathscr{M}_k\})$.

Step 6 **(Termination)** Terminate if $\mathscr{R}_k = \emptyset$ or, optionally, if $\{\mathscr{M}\in\mathscr{R}_k \mid \beta(\mathscr{M}) > \gamma_k + \eta\} = \emptyset$. Return $\bar{\mathbf{x}}^k$ as a global η-optimal solution. Otherwise, update $k \leftarrow k+1$ and return to Step 1.

a feasible point and, hence, minimizes the distance between the bound and feasible value. In other words, the goal of an adaptive bisection is to drive the upper bound as fast as possible toward a feasible point. This often leads to much faster convergence than exhaustive bisection.

The following proposition formally establishes convergence for Algorithm 2.2 with adaptive bisection.

Proposition 2.5. *If the subdivision procedure is adaptive,* $\mathbf{x}^k$ *is such that* $f(\mathbf{x}^k) = \alpha(\mathscr{M}_k)$ *and (2.23) is satisfied, then Algorithm 2.2 is convergent and the final* $\bar{\mathbf{x}}^k$ *is a global* η*-optimal solution of (2.16).*

Proof. Please refer to [12, Proposition 6.2] or [25, Proposition 3.11]. □

Another way to improve convergence is the reduction in Step 2 of Algorithm 2.2. First of all, it is important to notice that this step is entirely optional since choosing $\mathscr{M}' = \mathscr{M}$ satisfies (2.24). The general idea is to shrink the box under consideration such that no candidate solutions, i.e., feasible points that might have objective value greater than γ, are lost. The smaller size of a box leads to tighter bounds and, thus, to faster convergence. However, it should only be implemented if this can be done efficiently and leads to an overall improvement in run time. Please refer to [12,26,27] for some example procedures that rely on exploiting monotonicity.

Implementing Algorithm 2.2 for a specific problem requires a bounding procedure $\beta(\mathscr{M})$ and a solution to the feasibility problem in Step 3. This will be discussed in the next two sections.

2.3.2 Bounding methods

In this section, we discuss bounding methods that exploit (hidden) monotonicity in the objective functions. The traditional approach is monotonic programming [20,27], where the objective is written as the difference of two increasing (difference of increasing (DI)) functions, i.e.,

$$f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}), \tag{2.25}$$

where f_1, f_2 are increasing functions. Computing the bound $\beta(\mathscr{M})$ over a box $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ is straightforward:

$$\max_{\mathscr{M}\cap\mathscr{D}} f(\mathbf{x}) \leq \max_{\mathscr{M}} f(\mathbf{x}) \leq \max_{\mathscr{M}} (f_1(\mathbf{x}) - f_2(\mathbf{r})) = f_1(\mathbf{s}) - f_2(\mathbf{r}) = \beta(\mathscr{M}). \tag{2.26}$$

This bound does not provide any points suitable for an adaptive subdivision and, thus, has to be combined with exhaustive subdivision. For convergence, the bound has to satisfy the condition in (2.22). As $\operatorname{diam}(\mathscr{M}) \to 0$, the cornerpoints $\mathbf{r}, \mathbf{s}$ approach a common limit point $\mathbf{y}$. Hence, at some point $\mathscr{M} \cap \mathscr{D} = \mathscr{M}$ and $\beta(\mathscr{M}) \to f_1(\mathbf{y}) - f_2(\mathbf{y}) = \max_{\mathbf{x}\in\mathscr{M}}\{f_1(\mathbf{x}) - f_2(\mathbf{x})\}$. Thus, the bound in (2.26) is consistent with branching and leads to a convergent exhaustive BB procedure.

Applying this bound to GEE maximization in interference networks, i.e.,

$$\max_{\mathbf{p}} \frac{\sum_i \log\left(1+\frac{d_i p_i}{1+\sum_{j\neq i} c_{i,j} p_j}\right)}{P_c + \sum_i \mu_i p_i} \quad \text{s.t.} \quad \mathbf{p} \in \mathscr{P}, \tag{2.27}$$

with $\mathscr{P} \subset \mathbb{R}^n_{\geq 0}$, requires the application of Dinkelbach's Algorithm. Assuming $d_i, c_{i,j} \geq 0$ for all i,j, the numerator can be cast into a DI function as

$$\begin{aligned}\log\left(1+\frac{d_i p_i}{1+\sum_{j\neq i} c_{i,j} p_j}\right) &= \log\left(\frac{1+\sum_{j\neq i} c_{i,j} p_j + d_i p_i}{1+\sum_{j\neq i} c_{i,j} p_j}\right) \\ &= \underbrace{\log\left(1+\sum_{j\neq i} c_{i,j} p_j + d_i p_i\right)}_{f_1(\mathbf{p})} - \underbrace{\log\left(1+\sum_{j\neq i} c_{i,j} p_j\right)}_{f_2(\mathbf{p})}.\end{aligned} \tag{2.28}$$

The denominator of (2.27) is also an increasing function, but this does not help in finding a DI representation of (2.27). The common approach is to employ Dinkelbach's Algorithm as discussed in Section 2.2.4. Recall that in each iteration of this algorithm, the auxiliary problem:

$$\max_{\mathbf{p}} \sum_i \log\left(1+\frac{d_i p_i}{1+\sum_{j\neq i} c_{i,j} p_j}\right) - \lambda\left(P_c + \sum_i \mu_i p_i\right) \quad \text{s.t.} \quad \mathbf{p} \in \mathscr{P} \tag{2.29}$$

has to be solved. The objective of (2.29) has the DI representation:

$$\sum_i \log\left(1+\sum_{j\neq i} c_{i,j} p_j + d_i p_i\right) - \left(\sum_i \log\left(1+\sum_{j\neq i} c_{i,j} p_j\right) + \lambda\left(P_c + \sum_i \mu_i p_i\right)\right), \tag{2.30}$$

and a suitable bound for (2.29) can be computed as in (2.26). The downside of this approach is that the non-convex inner problem (2.29) needs to be solved several times. This increases the computational complexity tremendously.

This is where the flexibility of BB methods comes in handy. The bound can easily be replaced by a different one without altering any other part of the algorithm as long as it is consistent with branching. Consider the following function

$$f(\mathbf{x}) = \frac{f_1(\mathbf{x}) - f_2(\mathbf{x})}{g(\mathbf{x})}, \tag{2.31}$$

with f_1, f_2, g increasing and $f_1 - f_2, g$ non-negative. A bound over the box $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ can be computed as

$$\max_{\mathbf{x}\in\mathscr{M}\cap\mathscr{D}} f(\mathbf{x}) \leq \frac{\max_{\mathbf{x}\in\mathscr{M}} f_1(\mathbf{x}) - f_2(\mathbf{x})}{\min_{\mathbf{x}\in\mathscr{M}} g(\mathbf{x})} \leq \frac{f_1(\mathbf{s}) - f_2(\mathbf{r})}{g(\mathbf{r})}. \tag{2.32}$$

Again, since $\mathbf{r}, \mathbf{s} \to \mathbf{y}$ as $\operatorname{diam}(\mathscr{M}) \to 0$, $\max_{\mathbf{x}\in\mathscr{M}\cap\mathscr{D}} f(\mathbf{x}) \to f(\mathbf{y})$ and $\beta(\mathscr{M}) \to \frac{f_1(\mathbf{y}) - f_2(\mathbf{y})}{g(\mathbf{y})} = f(\mathbf{y})$. Thus, this bound is consistent with branching. This approach

eliminates the need for Dinkelbach's Algorithm and directly obtains an η-optimal solution.

We can easily extend this bounding approach to multi-ratio optimization problems. For example, a consistent bounding function for the sum-of-ratios objective

$$\sum_i \frac{f_{i,1}(\mathbf{x}) - f_{i,2}(\mathbf{x})}{g_i(\mathbf{x})}, \tag{2.33}$$

is directly obtained from (2.32) as

$$\sum_i \frac{f_{i,1}(\mathbf{s}) - f_{i,2}(\mathbf{r})}{g(\mathbf{r})}. \tag{2.34}$$

Recall from Definition 2.3 and the discussion made later that this problem cannot be solved with Dinkelbach's Algorithm or its generalizations.

The bounding approach taken here can be further improved and generalized into the MMP framework [26,28]. The core idea is to find a function $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies:

$$\mathbf{x} \leq \mathbf{x}' \Rightarrow F(\mathbf{x}, \mathbf{y}) \leq F(\mathbf{x}', \mathbf{y}), \tag{2.35}$$

$$\mathbf{y} \leq \mathbf{y}' \Rightarrow F(\mathbf{x}, \mathbf{y}) \geq F(\mathbf{x}, \mathbf{y}'), \tag{2.36}$$

for all $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}' \in \mathscr{M}_0 \supseteq \mathscr{D}$. Such a function is called *mixed monotonic (MM) function*. If it satisfies:

$$F(\mathbf{x}, \mathbf{x}) = f(\mathbf{x}), \tag{2.37}$$

it is called an *MMP representation* of f and (2.16) is said to be an *MMP problem*. A bound on the objective f over $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ can be determined easily from F as

$$\max_{\mathbf{x} \in \mathscr{M} \cap \mathscr{D}} f(\mathbf{x}) \leq \max_{\mathbf{x} \in \mathscr{M}} F(\mathbf{x}, \mathbf{x}) \leq \max_{\mathbf{x}, \mathbf{y} \in \mathscr{M}} F(\mathbf{x}, \mathbf{y}) = F(\mathbf{s}, \mathbf{r}) = \beta(\mathscr{M}). \tag{2.38}$$

It is shown in [26, Theorem 1] that this bound is consistent with branching and, hence, leads to a convergent BB procedure if combined with an exhaustive rectangular subdivision. For example, an MMP representation of the GEE in (2.27) is

$$(\mathbf{x}, \mathbf{y}) \mapsto \frac{\sum_i \log\left(1 + \frac{d_i x_i}{1 + \sum_{j \neq i} c_{i,j} y_j}\right)}{P_c + \sum_i \mu_i y_i}, \tag{2.39}$$

and, thus, a bounding function to solve (2.27) with Algorithm 2.2 is, for $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$,

$$\beta(\mathscr{M}) = F(\mathbf{s}, \mathbf{r}) = \frac{\sum_i \log\left(1 + \frac{d_i s_i}{1 + \sum_{j \neq i} c_{i,j} r_j}\right)}{P_c + \sum_i \mu_i r_i}. \tag{2.40}$$

It is shown analytically in [26, Section IV] that the bound for the numerator is tighter than that obtained by the DI approach in (2.28). Figure 2.4 shows performance comparisons from [26, Section IV] for the different bounding approaches discussed in

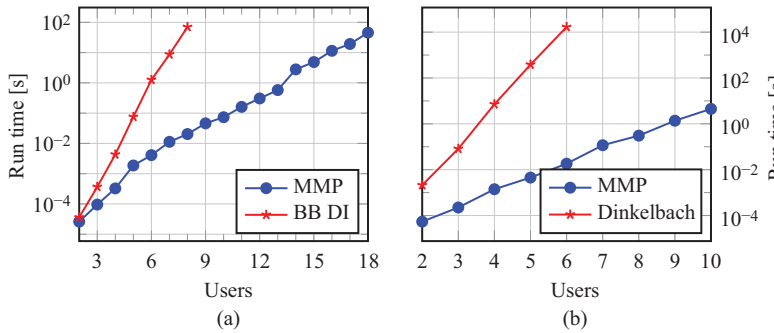


Figure 2.4 Average run time of Algorithm 2.2 with different bounding approaches. Results are averaged over 100 independent and identically distributed channel realizations. (a) Weighted sum rate maximization: MMP bound vs. DI bound, i.e., (2.40) with $\mu_i = 0$ and $P_c = 1$ and (2.28), respectively; (b) GEE maximization: MMP vs. Dinkelbach's Algorithm as in (2.40) and (2.30), respectively.

this section. Figure 2.4(a) compares bounding of the numerator of the GEE, i.e., throughput maximization. It can be observed that MMP bounding is several orders of magnitude faster than the classical DI bound. For example, MMP solves throughput maximization problems with 18 variables in the same time than classical monotonic programming requires for 8 variables. In Figure 2.4(b), run times for GEE maximization are shown. Solving this problem directly with Algorithm 2.2 and the MMP bound in (2.40) is again several orders of magnitude faster than using Dinkelbach's Algorithm with Algorithm 2.2 and DI bounding as an inner solver, e.g., for six users, MMP is five orders of magnitude faster than the state-of-the-art approach.

2.3.3 Feasibility test

The second component to implement in any BB procedure is a feasibility test. While the bounding operation deals with the objective, the feasibility test is concerned with the feasible set $\mathscr{D}$. In this section, we first discuss some special cases often encountered in EE maximization and then focus on more general methods.

2.3.3.1 Box constraints

The most simple feasible sets are simple box constraints on the optimization variables, i.e.,

$$\mathscr{D} = \{\mathbf{x} \in \mathbb{R}^n \,|\, x_{i,\min} \leq x_i \leq x_{i,\max}\} = [\mathbf{x}_{\min}, \mathbf{x}_{\max}]. \tag{2.41}$$

In this case, the initial box $\mathscr{M}_0$ should be chosen as $[\mathbf{x}_{\min}, \mathbf{x}_{\max}]$. Because Algorithm 2.2 only generates points within $\mathscr{M}_0$, every encountered point is feasible. Hence, $\mathscr{M} \cap \mathscr{D} \neq \emptyset$ in every iteration and $\alpha(\mathscr{M})$ in Step 3 is an arbitrary point in $\mathscr{M}$.

2.3.3.2 Minimum rate constraints

The most common form of Quality of Service (QoS) constraints in energy-efficient resource allocation are minimum rate constraints. In interference networks, where each user k achieves a rate:

$$r_k = \log\left(1 + \frac{d_i p_i}{1 + \sum_{j \neq i} c_{i,j} p_j}\right), \tag{2.42}$$

the resulting feasible set is

$$\mathscr{D} = \left\{ \mathbf{p} \in [\mathbf{0}, \mathbf{P}] \,\middle|\, \log\left(1 + \frac{d_i p_i}{1 + \sum_{j \neq i} c_{i,j} p_j}\right) \geq R_{i,\min}, \text{ for all } i \right\}. \tag{2.43}$$

Despite (2.42) being a non-convex function, this set can be easily transformed into a linear set as follows:

$$\log\left(1 + \frac{d_i p_i}{1 + \sum_{j \neq i} c_{i,j} p_j}\right) \geq R_{i,\min} \tag{2.44}$$

$$\Leftrightarrow \qquad \frac{d_i p_i}{1 + \sum_{j \neq i} c_{i,j} p_j} \geq 2^{R_{i,\min}} - 1 \tag{2.45}$$

$$\Leftrightarrow \qquad d_i p_i - \left(2^{R_{i,\min}} - 1\right) \sum_{j \neq i} c_{i,j} p_j \geq 2^{R_{i,\min}} - 1, \tag{2.46}$$

where the last inequality is clearly affine in the powers p_i. Hence, the feasible set

$$\mathscr{D} = \left\{ \mathbf{p} \in [\mathbf{0}, \mathbf{P}] \,\middle|\, \left(2^{R_{i,\min}} - 1\right) \sum_{j \neq i} c_{i,j} p_j - d_i p_i \leq 1 - 2^{R_{i,\min}}, \text{ for all } i \right\} \tag{2.47}$$

is linear. Thus, the check $\mathscr{M} \cap \mathscr{D} \stackrel{?}{=} \emptyset$ is a simple linear feasibility problem and can be solved by any linear programming Phase I method.

The initial box $\mathscr{M}_0$ should be chosen as tight as possible to avoid slowing down convergence unnecessarily. Thus, the obvious choice $\mathscr{M}_0 = [\mathbf{0}, \mathbf{P}]$ should be improved by choosing $\mathscr{M}_0 = [\mathbf{x}_{\min}, \mathbf{x}_{\max}]$ with

$$x_{i,\min} = \min_{\mathbf{p} \in \mathscr{D}} p_i \qquad\qquad x_{i,\max} = \max_{\mathbf{p} \in \mathscr{D}} p_i. \tag{2.48}$$

These are linear programs and can be solved at low computational cost.

2.3.3.3 General inequality constraints

Let the feasible set $\mathscr{D}$ be defined by MM functions G_i as

$$G_i(\mathbf{x}, \mathbf{x}) \leq 0, \quad \text{for all } i = 1, 2, \ldots, m. \tag{2.49}$$

The following proposition follows from the definition of MM functions.

Proposition 2.6 ([26, Proposition 1]). *Let* $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ *and* $\mathscr{D}$ *be defined by the inequalities in (2.49). Then,*

$$\forall i \in \{1, \ldots, m\} : G_i(\mathbf{r}, \mathbf{s}) \leq 0 \Rightarrow \mathscr{M} \cap \mathscr{D} = \mathscr{M} \neq \emptyset \tag{2.50}$$

$$\exists i \in \{1, \ldots, m\} : G_i(\mathbf{r}, \mathbf{s}) > 0 \Rightarrow \mathscr{M} \cap \mathscr{D} = \emptyset. \tag{2.51}$$

These are sufficient conditions for (in-)feasibility of a box $\mathscr{M}$. However, there exist boxes for which neither condition holds. In that case, it is impossible to answer the feasibility test $\mathscr{M} \cap \mathscr{D} \stackrel{?}{=} \emptyset$ conclusively solely based on the monotonicity properties.

An exception is the important special cases where the feasible set is either normal or conormal.

Proposition 2.7 ([26, Corollary 1]). *Let* $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ *and* $\mathscr{D} = \{\mathbf{x} \mid g_i(\mathbf{x}) \leq 0,$ *for all* $i = 1, \ldots, m\}$ *with* g_i *being increasing functions. Then,* $\mathscr{D} \cap \mathscr{M} \neq \emptyset$ *if and only if* $g_i(\mathbf{r}) \leq 0$ *for all* $i = 1, \ldots, m$.

Proposition 2.8 ([26, Corollary 2]). *Let* $\mathscr{M} = [\mathbf{r}, \mathbf{s}]$ *and* $\mathscr{D} = \{\mathbf{x} \mid h_i(\mathbf{x}) \geq 0,$ *for all* $i = 1, \ldots, m\}$ *with* h_i *being increasing functions. Then,* $\mathscr{D} \cap \mathscr{M} \neq \emptyset$ *if and only if* $h_i(\mathbf{r}) \geq 0$ *for all* $i = 1, \ldots, m$.

This can be generalized to MM functions that are either increasing or decreasing in each dimension. Please refer to [26, Section III-A] for details.

If neither Proposition 2.7 nor 2.8 is applicable, we can modify Algorithm 2.2 to work with general MM constraints and the result in Proposition 2.6. In particular, we only require a feasible solution in Step 3 if one is available and prune boxes in Step 5 if (2.51) is satisfied. The resulting procedure is stated in Algorithm 2.3. These

Algorithm 2.3: Rectangular BB algorithm for general DI constraints

Step 0 **(Initialization)** Choose $\mathscr{M}_0 \supseteq \mathscr{D}$ and $\eta > 0$. Let $k = 1$ and $\mathscr{R}_0 = \{\mathscr{M}_0\}$. If available or easily computable, find $\bar{\mathbf{x}}^0 \in \mathscr{D}$ and set $\gamma_0 = f(\bar{\mathbf{x}}^0)$. Otherwise, set $\gamma_0 = -\infty$.

Step 1 **(Branching)** Select a box $\mathscr{M}_k \in \arg\max\{\beta(\mathscr{M}) \mid \mathscr{M} \in \mathscr{R}_{k-1}\}$ and subdivide it via $(\mathbf{v}^k, j)$ (cf. (2.21)). Let $\mathscr{P}_k = \{\mathscr{M}_k^-, \mathscr{M}_k^+\}$.

Step 2 **(Reduction)** For each $\mathscr{M} \in \mathscr{P}_k$, replace $\mathscr{M}$ by $\mathscr{M}'$ such that $\mathscr{M}' \subseteq \mathscr{M}$ and

$$(\mathscr{M} \setminus \mathscr{M}') \cap \{\mathbf{x} \in \mathscr{D} \mid f(\mathbf{x}) > \gamma_k\} = \emptyset. \tag{2.52}$$

Step 3 **(Bounding)** For each $\mathscr{M} \in \mathscr{P}_k$, compute $\beta(\mathscr{M}) \geq \sup_{\mathbf{x} \in \mathscr{M} \cap \mathscr{D}} f(\mathbf{x})$. Let $\mathbf{x} \in \mathscr{M}$ (e.g., $\mathbf{x} = \mathbf{r}$). If $\mathbf{x} \in \mathscr{D}$, set $\alpha(\mathscr{M}) = f(\mathbf{x})$. Otherwise, set $\alpha(\mathscr{M}) = -\infty$.

Step 4 **(Incumbent)** Let $\alpha_k = \max\{\alpha(\mathscr{M}) \mid \mathscr{M} \in \mathscr{P}_k\}$. If $\alpha_k > \gamma_{k-1}$, set $\gamma_k = \alpha_k$ and let $\bar{\mathbf{x}}^k \in \mathscr{D}$ such that $\alpha_k = f(\bar{\mathbf{x}}^k)$. Otherwise, let $\gamma_k = \gamma_{k-1}$ and $\bar{\mathbf{x}}^k = \bar{\mathbf{x}}^{k-1}$.

Step 5 **(Pruning)** Delete every $\mathscr{M} \in \mathscr{P}_k$ with $\beta(\mathscr{M}) \leq \gamma_k + \eta$ or $G_i(\mathbf{r}, \mathbf{s}) > 0$ for some i. Let $\mathscr{P}'_k$ be the collection of remaining sets and set $\mathscr{R}_k = \mathscr{P}'_k \cup (\mathscr{R}_{k-1} \setminus \{\mathscr{M}_k\})$.

Step 6 **(Termination)** Terminate if $\mathscr{R}_k = \emptyset$ or, optionally, if $\{\mathscr{M} \in \mathscr{R}_k \mid \beta(\mathscr{M}) > \gamma_k + \eta\} = \emptyset$. Return $\bar{\mathbf{x}}^k$ as a global η-optimal solution. Otherwise, update $k \leftarrow k + 1$ and return to Step 1.

changes seem minor, but the consequence is that convergence in a finite number of iterations cannot be guaranteed for Algorithm 2.3. This is established in the following proposition.

Proposition 2.9. *If Algorithm 2.3 terminates, then problem (2.16) with MM constraints as in (2.49) is either infeasible or $\bar{\mathbf{x}}^k$ is an η-optimal solution of (2.16). If it is infinite, at least one accumulation point of the infinite sequence $\{\mathbf{r}^k\}$ is a globally optimal solution of (2.16).*

Proof. Please refer to [25, Theorem 6.11] or [27]. □

While an infinite algorithm often converges in a finite number of iterations, this cannot be guaranteed. Moreover, the convergence is often slow. In the next section, we discuss solution approaches for this issue.

2.4 Successive incumbent transcending scheme

To ease the exposition in this section, consider the special case of DI constraints instead of the general MM constraints from Section 2.3.3.3, i.e., let $G_i(\mathbf{x}, \mathbf{x}) = g_i^+(\mathbf{x}) - g_i^-(\mathbf{x})$ in (2.49) with g_i^+, g_i^- being increasing functions for all i. A widely accepted method to obtain a finite algorithm from Algorithm 2.3 is to accept any point $\mathbf{x}$ satisfying

$$g_i^+(\mathbf{x}) - g_i^-(\mathbf{x}) \leq \varepsilon, \quad \text{for all } i = 1, 2, \ldots, m, \tag{2.53}$$

for some small $\varepsilon > 0$ as feasible point in Step 3. Because only strictly infeasible boxes are pruned in Step 5, a gap between the pruning rule and acceptable points exists. Algorithmically, this is similar to the motivation behind the η-optimality concept and results in finite convergence. The result of this algorithm is an η-optimal point that is *almost* feasible, also known as an (ε, η)-approximate optimal solution. Since this point $\bar{\mathbf{x}}$ is almost feasible for small ε, it should converge to a feasible solution as $\varepsilon \to 0$. Moreover, it is expected that $f(\bar{\mathbf{x}})$ is also close to the optimal value for a sufficiently small $\varepsilon < \varepsilon_0$. While this is usually true, the problem is that ε_0 is unknown in general, and choosing ε too large can result in completely wrong "solutions." This is illustrated in the following example from [29].

Example 2.1. *Consider the optimization problem:*

$$\begin{cases} \min\limits_{\mathbf{p} \in \mathbb{R}^2} & p_1 \\ \text{s.t.} & \log_2(1 + \alpha_1 p_1 + \alpha_2 p_2) \geq Q \\ & \log_2(1 + \beta_1 p_1) + \log_2(1 + \beta_2 p_2) \leq L \\ & \mathbf{p} \in [\mathbf{0}, \mathbf{P}], \end{cases} \tag{2.54}$$

where the constants α_1, α_2, β_1, β_2, Q, L are all positive. This problem could be motivated as transmit power minimization in a communication system where the first

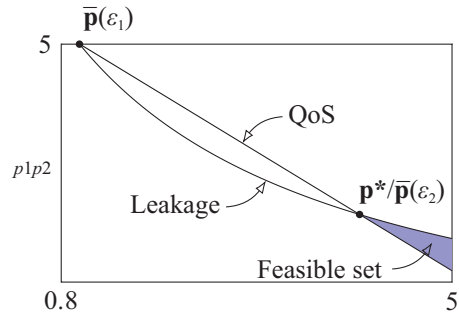


Figure 2.5 Feasible set of (2.54) with optimal solution $\boldsymbol{p}^*$ and ε-approximate solutions $\bar{\boldsymbol{p}}(\varepsilon_1)$ and $\bar{\boldsymbol{p}}(\varepsilon_2)$ for $\varepsilon_1 = 10^{-3}$ and $\varepsilon_2 = 10^{-4}$. The ε_1-approximate solution is quite far away from the true optimum. © 2019 IEEE [29]

constraint is a minimum throughput constraint and the second, the total information leakage to non-cooperative eavesdroppers.

The feasible set of (2.54) is illustrated in Figure 2.5 for parameters $\alpha_1 = \alpha_2 = 10$, $\beta_1 = 0.5$, $\beta_2 = 1$, $Q = \log_2(61)$, $L = \log_2(8.99)$. *The optimum solution* $\mathbf{p}^*$ *is* $(4.00665, 1.99335)$ *with an objective value 4.00665. Relaxing the constraints by* $\varepsilon_1 = 10^{-3}$ *results in an* ε*-approximate solution* $(0.995843, 5)$. *Clearly, this solution is quite far away from the true optimum value and should be considered as wrong. Instead, choosing* $\varepsilon_2 = 10^{-4}$ *results in the correct solution.*

Another numerical issue can be observed for $L = \log_2(9)$ *with all other parameters as before. In that case, there exists an isolated feasible point at* $(1, 5)$ *that also happens to be the optimal solution to (2.54). There are several issues with such an isolated solution. First, it is numerically challenging to compute and some global optimization algorithms fail to converge should the solution be such an isolated point. Moreover, this solution is numerically instable in the sense that it reacts very sensitive to small perturbations, e.g., due to an* ε*-relaxation. Lastly, it is usually not desired as solution to engineering problems because parameters are usually subject to small variations such as measurement errors, hence rendering such a solution most likely wrong.*

2.4.1 ε-Essential feasibility and the SIT scheme

We have identified two issues in the previous section. The first is that relaxing the constraints by some small ε, the common remedy to obtain an finite algorithm from an infinite, might actually lead to very wrong solutions. The other is that solutions in isolated feasible points are numerically unstable and undesirable from an engineer's point of view. In this section, we discuss an alternative approach that improves the traditional method in both regards.

Instead of relaxing the feasible set $\mathscr{D}$ as defined in (2.49), we tighten it by a small amount $\varepsilon > 0$, i.e.,

$$\mathscr{D}^* = \{\mathbf{x} \mid g_i^+(\mathbf{x}) - g_i^-(\mathbf{x}) \leq -\varepsilon, \;\; i = 1, 2, \ldots, m\}. \tag{2.55}$$

A point $\mathbf{x} \in \mathscr{D}^*$ is called ε-essential feasible and a point $\bar{\mathbf{x}}$ is said to be an essential (ε, η)-optimal solution of (2.16) if it satisfies

$$f(\bar{\mathbf{x}}) \geq f(\mathbf{x}) - \eta \;\; \text{for all } \mathbf{x} \in \mathscr{D}^*, \tag{2.56}$$

for some $\eta > 0$. For $\varepsilon, \eta \to 0$, this is a non-isolated feasible point that solves (2.16).

Applying this approach directly to Algorithm 2.3 does not fix the convergence issues. In fact, it will most likely have a negative impact on the convergence speed. However, when combined with the SIT scheme discussed next, it results in a numerically stable BB algorithm with finite convergence to the globally optimal solution.

Consider the optimization problems:

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g_i^+(\mathbf{x}) - g_i^-(\mathbf{x}) \leq 0, \;\; \text{for all } i = 1, 2, \ldots, m, \tag{2.57}$$

and

$$\min_{\mathbf{x}} \max_{i=1,\ldots,m} \{g_i^+(\mathbf{x}) - g_i^-(\mathbf{x})\} \quad \text{s.t.} \quad f(\mathbf{x}) \geq \gamma, \tag{2.58}$$

which are dual in the sense that if the optimal value of (2.58) is greater than zero, the optimal value of (2.57) is less than γ. This is because every point $\mathbf{x}$ feasible in (2.58) having a non-positive objective value is a feasible point in (2.57) with objective value greater than γ. We can formalize this observation in the following proposition that is adapted from [12, Proposition 7.13].

Proposition 2.10 ([29, Proposition 1]). *For every $\varepsilon > 0$, the ε-essential optimal value of (2.57) is less than γ if the optimal value of (2.58) is greater than $-\varepsilon$.*

Proposition 2.10 gives rise to a simple sequential solution method for (2.57) where γ is increased as long as (2.58) has a non-positive solution. This is known as the SIT scheme that was developed by Tuy [12,27,30,31]. The approach is formalized in Algorithm 2.4.

The core problem in Algorithm 2.4 is the feasibility check in Step 1. Leveraging Proposition 2.10, this can be implemented by solving (2.58). At a first glance, this

Algorithm 2.4: SIT algorithm [12, Section 7.5.1]. © IEEE [29]

Step 0	Initialize $\bar{\mathbf{x}}$ with the best known non-isolated feasible solution and set $\gamma = f(\bar{\mathbf{x}}) + \eta$; otherwise do not set $\bar{\mathbf{x}}$ and choose $\gamma \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{D}$.
Step 1	Check if (2.57) has a non-isolated feasible solution $\mathbf{x}$ satisfying $f(\mathbf{x}) \geq \gamma$; otherwise, establish that no such ε-essential feasible $\mathbf{x}$ exists and go to Step 3.
Step 2	Update $\bar{\mathbf{x}} \leftarrow \mathbf{x}$ and $\gamma \leftarrow f(\bar{\mathbf{x}}) + \eta$. Go to Step 1.
Step 3	Terminate: If $\bar{\mathbf{x}}$ is set, it is an essential (ε, η)-optimal solution; else Problem (2.57) is ε-essential infeasible.

approach seems to increase computational complexity significantly because in its straightforward implementation the non-convex optimization problem (2.58) would be solved very often. However, (2.58) has to be solved only partially until some feasible point with non-positive objective value is found or it can be established that no solution to (2.58) with objective value less than $-\varepsilon$ exists. This can be achieved efficiently by a BB procedure with termination criterion "$\max_{i=1,\dots,m}\{g_i^+(\mathbf{x}) - g_i^-(\mathbf{x})\} \leq 0$ or $\beta(\mathscr{M}) > -\varepsilon$" [12, Proposition 7.14]. Moreover, this BB procedure can be directly incorporated with Algorithm 2.4 where γ is updated every time a point $\mathbf{x}$ is encountered with $\max_{i=1,\dots,m}\{g_i^+(\mathbf{x}) - g_i^-(\mathbf{x})\} \leq 0$ and $f(\mathbf{x}) \geq \gamma$. It is shown in the proof of [29, Theorem 1] that updating γ during the BB procedure does not prevent convergence.

The great benefit of this method is that (2.58) is often much easier to solve than (2.57). We have seen in Section 2.3.3.3 that finding a feasible point for general DI constraints is a challenging problem that leads to an infinite BB procedure. Instead, the feasible set of (2.58) is often much easier to deal with. Indeed, if f is an increasing (or decreasing) function, Proposition 2.8 establishes that a feasible point in a box $\mathscr{M}$ is easily found. Similarly beneficial properties of f are linearity or concavity. Even if f itself is a DI function, we can easily transform (2.57) such that it has a "nice" objective by introducing an auxiliary variable and bringing it into the epigraph form.

In the next section, we apply Algorithm 2.4 to a very general resource allocation problem. Other applications of this scheme can be found in [12, Section 7.5] for difference of convex (DC) programming, and in [12, Section 11.3] for canonical monotonic optimization.

2.4.2 SIT for fractional DI problems with some convex variables

Consider the optimization problem:

$$\begin{cases} \max\limits_{(\mathbf{x},\boldsymbol{\xi})\in\mathscr{D}} & \dfrac{f^+(\mathbf{x},\boldsymbol{\xi})}{f^-(\mathbf{x},\boldsymbol{\xi})} \\ \text{s.t.} & g_i^+(\mathbf{x},\boldsymbol{\xi}) - g_i^-(\mathbf{x}) \leq 0, \quad i = 1,2,\dots,m, \end{cases} \tag{2.59}$$

with global variables $\mathbf{x}$ and non-global variables $\boldsymbol{\xi}$. We say that a variable is non-global if there exists an algorithm to solve (2.59) in $\boldsymbol{\xi}$ with fixed global variables $\mathbf{x}$ that has much lower computational complexity than in the case when (2.59) is solved for all variables $(\mathbf{x},\boldsymbol{\xi})$. The goal of this section is to develop an algorithm that solves (2.59) with global optimality and preserves the low computational complexity in the non-global variables. This is done by means of an adaptive rectangular BB algorithm in combination with the SIT scheme. For a more detailed exposition, the interested reader is referred to [29] where this algorithm was first published. This optimization framework is motivated by joint power and rate allocation problems as in (2.8), where the rates are linear variables while the powers are non-convex.

Some technical assumptions on (2.59) are necessary. As in every rectangular BB method, an initial box $\mathscr{M}_0$ is required. Because the BB procedure operates only on the global variables $\mathbf{x}$, it has to enclose the $\mathbf{x}$ dimensions of $\mathscr{D}$, i.e., $\mathscr{M}^0 \supset \text{proj}_{\mathbf{x}}\,\mathscr{D}$. The functions $\{g_i^-(\mathbf{x})\}$ are required to have a common maximizer over every box

$\mathscr{M} \subseteq \mathscr{M}_0$. A sufficient condition for this is that the functions $g_i^-(\mathbf{x})$ have the same monotonicity properties per dimension. Further, we assume the functions $f^-(\mathbf{x}, \boldsymbol{\xi})$, $g_i^+(\mathbf{x}, \boldsymbol{\xi})$, $i = 1, \ldots, m$, to be lower semicontinuous, the functions $f^+(\mathbf{x}, \boldsymbol{\xi})$, $g_i^-(\mathbf{x})$, $i = 1, \ldots, m$, to be upper semicontinuous, and, without loss of generality, $f^-(\mathbf{x}, \boldsymbol{\xi}) > 0$. The remaining assumptions can be grouped into two different cases. Some of these assumptions depend on the variable γ that was introduced in Algorithm 2.4 and stores the current best primal objective value.

Case 2.1 (DC problems). If $\mathscr{D}$ is a closed convex set and $\gamma f^-(\mathbf{x}, \boldsymbol{\xi}) - f^+(\mathbf{x}, \boldsymbol{\xi})$, $g_1^+(\mathbf{x}, \boldsymbol{\xi}), \ldots, g_m^+(\mathbf{x}, \boldsymbol{\xi})$ are jointly convex in $(\mathbf{x}, \boldsymbol{\xi})$ for all γ, problem (2.59) resembles a DC optimization problem but with fractional objective and additional non-DC variables.

Case 2.2 (separable problems). Let $\mathscr{D} = \mathscr{X} \times \Xi$ such that $\mathbf{x} \in \mathscr{X}$ and $\boldsymbol{\xi} \in \Xi$ with Ξ being a closed convex set, and let each function of $(\mathbf{x}, \boldsymbol{\xi})$ be separable in the sense that $h(\mathbf{x}, \boldsymbol{\xi}) = h_x(\mathbf{x}) + h_\xi(\boldsymbol{\xi})$. Further, let the functions $\gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi})$, $g_{1,\xi}^+(\boldsymbol{\xi}), \ldots, g_{m,\xi}^+(\boldsymbol{\xi})$ be convex in $\boldsymbol{\xi}$ for all γ, and let the functions $\gamma f_x^-(\mathbf{x}) - f_x^+(\mathbf{x})$, $g_{1,x}^+(\mathbf{x}), \ldots, g_{m,x}^+(\mathbf{x})$ have a common minimizer over $\mathscr{X} \cap \mathscr{M}$ for every box $\mathscr{M} \subseteq \mathscr{M}_0$ and all γ. Finally, let the function $\gamma f_x^-(\mathbf{x}) - f_x^+(\mathbf{x})$ be either increasing for all γ with $\mathscr{X}$ being a closed normal set in some box or decreasing for all γ with $\mathscr{X}$ being a closed conormal set in some box.

Interchanging objective and constraints in (2.59) leads to the dual problem:

$$\begin{cases} \min\limits_{(\mathbf{x},\xi)\in\mathscr{D}} & \max\limits_{i=1,2,\ldots,m} \left(g_i^+(\mathbf{x}, \boldsymbol{\xi}) - g_i^-(\mathbf{x})\right) \\ \text{s.t.} & \dfrac{f^+(\mathbf{x}, \boldsymbol{\xi})}{f^-(\mathbf{x}, \boldsymbol{\xi})} \geq \gamma, \end{cases} \tag{2.60}$$

or, equivalently,

$$\begin{cases} \min\limits_{(\mathbf{x},\xi)\in\mathscr{D}} & \max\limits_{i=1,2,\ldots,m} \left(g_i^+(\mathbf{x}, \boldsymbol{\xi}) - g_i^-(\mathbf{x})\right) \\ \text{s.t.} & \gamma f^-(\mathbf{x}, \boldsymbol{\xi}) - f^+(\mathbf{x}, \boldsymbol{\xi}) \leq 0, \end{cases} \tag{2.61}$$

since $f^-(\mathbf{x}, \boldsymbol{\xi}) > 0$ by assumption. The feasible set of (2.61) is convex if the assumptions in Case 2.1 are met. While this does not hold for Case 2.2, these assumptions still lead to a sufficiently nice feasible set [29, Proposition 2] that facilitates an efficient solution of (2.61) via BB. This requires a lower bound on the objective of (2.61).

Proposition 2.11 ([29, Proposition 3]). *Let $\bar{\mathbf{x}}^*_{\mathscr{M}}$ be a common maximizer of $\{g_i^-(\mathbf{x})\}_{i=1,\ldots,m}$ over the box $\mathscr{M}$. Then, (2.61)'s objective is lower bounded over $\mathscr{M}$ by*

$$\max_{i=1,2,\ldots,m} \left\{g_i^+(\mathbf{x}, \boldsymbol{\xi}) - g_i^-(\bar{\mathbf{x}}^*_{\mathscr{M}})\right\}. \tag{2.62}$$

This bound is tight at $\bar{\mathbf{x}}^*_{\mathscr{M}}$.

Then, a lower bound $\beta(\mathscr{M})$ of (2.61) over a box $\mathscr{M}$ is determined as the optimal value of

$$\begin{cases} \min\limits_{\mathbf{x},\xi} & \max\limits_{i=1,2,\ldots,m} \left\{ g_i^+(\mathbf{x},\boldsymbol{\xi}) - g_i^-(\bar{\mathbf{x}}^*_{\mathscr{M}}) \right\} \\ \text{s.t.} & \gamma f^-(\mathbf{x},\boldsymbol{\xi}) - f^+(\mathbf{x},\boldsymbol{\xi}) \le 0 \\ & (\mathbf{x},\boldsymbol{\xi}) \in \mathscr{D}, \ \ \mathbf{x} \in \mathscr{M}. \end{cases} \tag{2.63}$$

For Case 2.1, this is a convex optimization problem. For Case 2.2, rewrite it as

$$\begin{cases} \min\limits_{\mathbf{x},\xi} & \max\limits_{i=1,2,\ldots,m} \left\{ g_{i,\boldsymbol{\xi}}^+(\boldsymbol{\xi}) + g_{i,x}^+(\mathbf{x}) - g_i^-(\bar{\mathbf{x}}^*_{\mathscr{M}}) \right\} \\ \text{s.t.} & \gamma f_{\xi}^-(\boldsymbol{\xi}) - f_{\xi}^+(\boldsymbol{\xi}) + \gamma f_x^-(\mathbf{x}) - f_x^+(\mathbf{x}) \le 0 \\ & \boldsymbol{\xi} \in \Xi, \ \ \mathbf{x} \in \mathscr{X} \cap \mathscr{M}. \end{cases} \tag{2.64}$$

The functions $\gamma f_x^-(\mathbf{x}) - f_x^+(\mathbf{x})$, $g_{1,x}^+(\mathbf{x}), \ldots, g_{m,x}^+(\mathbf{x})$ have a common minimizer $\underline{\mathbf{x}}^*_{\mathscr{M}}$ over $\mathscr{X} \cap \mathscr{M}$ by assumption. This allows one to partially solve (2.64) and obtain the convex problem:

$$\begin{cases} \min\limits_{\xi \in \Xi} & \max\limits_{i=1,2,\ldots,m} \left\{ g_{i,\boldsymbol{\xi}}^+(\boldsymbol{\xi}) + g_{i,x}^+(\underline{\mathbf{x}}^*_{\mathscr{M}_i}) - g_i^-(\bar{\mathbf{x}}^*_{\mathscr{M}_i}) \right\} \\ \text{s.t.} & \gamma f_{\xi}^-(\boldsymbol{\xi}) - f_{\xi}^+(\boldsymbol{\xi}) + \gamma f_x^-(\underline{\mathbf{x}}^*_{\mathscr{M}_i}) - f_x^+(\underline{\mathbf{x}}^*_{\mathscr{M}_i}) \le 0. \end{cases} \tag{2.65}$$

Both bounding procedures, i.e., (2.63) for Case 2.1 and (2.65) for Case 2.2, provide the necessary points for an adaptive subdivision procedure. Recall that for the box $\mathscr{M}_k$ two points $\mathbf{x}^k, \mathbf{y}^k$ satisfying (2.23) are required, where $\mathbf{y}^k$ is often chosen as the point where the bounding procedure is tight and $\mathbf{x}^k$ is a feasible point in the box $\mathscr{M}_k$. Thus, $\mathbf{x}^k$ is chosen as the optimal solution of (2.63) for Case 2.1. For Case 2.2, the optimal solution in $\mathbf{x}$ of (2.64) is $\underline{\mathbf{x}}^*_{\mathscr{M}}$ and, hence, $\mathbf{x}^k = \underline{\mathbf{x}}^*_{\mathscr{M}}$. By virtue of Proposition 2.11, both bounds are tight at $\mathbf{y}^k = \bar{\mathbf{x}}^*_{\mathscr{M}}$.

The following proposition, which is adapted from [12, Proposition 7.14], formally connects Step 1 in Algorithm 2.4 and Proposition 2.10 into a BB procedure.

Proposition 2.12 ([12, Proposition 4]). *Let $\varepsilon > 0$ be given. Either $g(\boldsymbol{x}^k, \boldsymbol{\xi}^*) < 0$ for some k and $\boldsymbol{\xi}^*$ or $\beta(\mathscr{M}^k) > -\varepsilon$ for some k. In the former case, $(\boldsymbol{x}^k, \boldsymbol{\xi}^*)$ is a non-isolated feasible solution of (2.59) satisfying $\frac{f^+(\boldsymbol{x}^k,\boldsymbol{\xi}^*)}{f^-(\boldsymbol{x}^k,\boldsymbol{\xi}^*)} \ge \gamma$. In the latter case, no ε-essential feasible solution $(\boldsymbol{x},\boldsymbol{\xi})$ of (2.59) exists such that $\frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \ge \gamma$.*

Algorithm 2.5 integrates the SIT scheme in Algorithm 2.4 into an adaptive BB procedure modified according to the earlier-shown proposition, i.e., a BB procedure with pruning criterion $\beta(\mathscr{M}) > -\varepsilon$ and stopping criterion $\min_{\xi \in \mathscr{D}_{\mathbf{x}^k}} g(\mathbf{x}^k, \boldsymbol{\xi}) < 0$. Of course, as already discussed, the final algorithm does not terminate upon $\min_{\xi \in \mathscr{D}_{\mathbf{x}^k}} g(\mathbf{x}^k, \boldsymbol{\xi}) < 0$ but updates the incumbent γ and continues.

Algorithm 2.5: SIT algorithm for (2.59) [29, Algorithm 2]. © IEEE [29]

Step 0 Initialize $\varepsilon, \eta > 0$ and $\mathscr{M}_0 = [\mathbf{r}^0, \mathbf{s}^0]$ as in (2.68), $\mathscr{P}_1 = \{\mathscr{M}_0\}$, $\mathscr{R} = \emptyset$, and $k = 1$. Initialize $\bar{\mathbf{x}}$ with the best known non-isolated feasible solution and set γ as described in Step 4; otherwise do not set $\bar{\mathbf{x}}$ and choose $\gamma \leq \frac{f^+(\mathbf{x},\boldsymbol{\xi})}{f^-(\mathbf{x},\boldsymbol{\xi})}$ for all feasible $(\mathbf{x}, \boldsymbol{\xi})$.

Step 1 For each box $\mathscr{M} \in \mathscr{P}_k$:

- Compute $\beta(\mathscr{M})$. Set $\beta(\mathscr{M}) = \infty$ if (2.63) (or (2.65)) is infeasible.
- Add $\mathscr{M}$ to $\mathscr{R}$ if $\beta(\mathscr{M}) \leq -\varepsilon$.

Step 2 Terminate if $\mathscr{R} = \emptyset$: If $\bar{\mathbf{x}}$ is not set, then (2.59) is ε-essential infeasible; else $\bar{\mathbf{x}}$ is an essential (ε, η)-optimal solution of (2.59).

Step 3 Let $\mathscr{M}_k = \arg\min\{\beta(\mathscr{M}) \mid \mathscr{M} \in \mathscr{R}\}$. Let $\mathbf{x}^k$ be the optimal solution of (2.63) for the box $\mathscr{M}_k$ (or $\underline{\mathbf{x}}^*_{\mathscr{M}_k}$ if (2.65) is employed for bounding), and $\mathbf{y}^k = \bar{\mathbf{x}}^*_{\mathscr{M}_k}$. Solve the feasibility problem:

$$\begin{cases} \text{find} & \boldsymbol{\xi} \in \mathscr{D}_{\mathbf{x}^k} \\ \text{s.t.} & g_i^+(\mathbf{x}^k, \boldsymbol{\xi}) - g_i^-(\mathbf{x}^k) \leq 0, \quad i = 1, \ldots, m. \end{cases} \tag{2.66}$$

If (2.66) is feasible, go to Step 4; otherwise go to Step 5.

Step 4 $\mathbf{x}^k$ is a non-isolated feasible solution satisfying $\frac{f^+(\mathbf{x}^k,\boldsymbol{\xi})}{f^-(\mathbf{x}^k,\boldsymbol{\xi})} \geq \gamma$ for some $\boldsymbol{\xi} \in \mathscr{D}_{\mathbf{x}^k}$. Let $\boldsymbol{\xi}^*$ be a solution to:

$$\begin{cases} \min\limits_{\xi \in \mathscr{D}_{\mathbf{x}^k}} & \dfrac{f^+(\mathbf{x}^k, \boldsymbol{\xi})}{f^-(\mathbf{x}^k, \boldsymbol{\xi})} \\ \text{s.t.} & g_i^+(\mathbf{x}^k, \boldsymbol{\xi}) - g_i^-(\mathbf{x}^k) \leq 0, \quad i = 1, 2 \ldots, m. \end{cases} \tag{2.67}$$

If $\bar{\mathbf{x}}$ is not set or $\frac{f^+(\mathbf{x}^k,\boldsymbol{\xi}^*)}{f^-(\mathbf{x}^k,\boldsymbol{\xi}^*)} > \gamma - \eta$, set $\bar{\mathbf{x}} = \mathbf{x}^k$ and $\gamma = \frac{f^+(\mathbf{x}^k,\boldsymbol{\xi}^*)}{f^-(\mathbf{x}^k,\boldsymbol{\xi}^*)} + \eta$.

Step 5 Bisect $\mathscr{M}_k$ via $(\mathbf{v}^k, j_k)$ where $j_k \in \arg\max_j\{|y_j^k - x_j^k|\}$ and $\mathbf{v}^k = \frac{1}{2}(\mathbf{x}^k + \mathbf{y}^k)$ (cf. (2.21)). Remove $\mathscr{M}_k$ from $\mathscr{R}$. Let $\mathscr{P}_{k+1} = \{\mathscr{M}_k^-, \mathscr{M}_k^+\}$. Increment k and go to Step 1.

The algorithm is initialized in Step 0 with an initial box $\mathscr{M}_0 = [\mathbf{r}^0, \mathbf{s}^0]$ satisfying:

$$r_i^0 \leq \min_{(\mathbf{x},\boldsymbol{\xi}) \in \mathscr{D}} x_i \qquad s_i^0 \geq \max_{(\mathbf{x},\boldsymbol{\xi}) \in \mathscr{D}} x_i. \tag{2.68}$$

The set $\mathscr{P}_k$ contains new boxes to be evaluated in Step 1, γ is the current best value adjusted by the tolerance η, and $\mathscr{R}$ holds the remaining partition of not yet eliminated boxes. Step 1 computes the bounds for the new boxes in $\mathscr{P}_k$. If the bound of a box is less than $-\varepsilon$, it might contain a non-isolated feasible solution that improves γ and is added to $\mathscr{R}$ for further evaluation. The next box for branching is selected from $\mathscr{R}$ in Step 3 and the incumbent is updated in Step 4. Branching is done in Step 5 and the termination criterion is evaluated in Step 2. Convergence of this algorithm is formally established in the following.

Proposition 2.13 ([29, Theorem 1]). *Algorithm 2.5 converges in finitely many steps to the (ε, η)-optimal solution of (2.59) or establishes that no such solution exists.*

2.5 Sequential convex approximation

Although significant performance improvements have been made to global optimization methods in the past years [26,29], they are still limited by their exponential complexity. This prohibits their application to large-scale optimization problems and online resource allocation in practical systems. Instead, they are motivated by applications that are not time critical, e.g., system design and evaluation of low complexity heuristic resource allocation methods.

A powerful method to obtain computationally efficient resource allocation methods is sequential convex programming [32]. The general idea is to find KKT points of the original problem by solving a sequence of easier programs. Recall that the KKT conditions are necessary prerequisites for the solution of a continuous optimization problem. For convex programs, they are also often sufficient conditions for optimality. This, however, does not hold in general and a KKT point can at most be considered a candidate solution for non-convex problems. Indeed, a KKT point is not necessarily a local minimum and good algorithms also ensure that the objective value is monotonically increasing. It appears reasonable to expect that a solution obtained by such an ascension process is more likely to obtain a local maximum than any method solely focusing on the KKT conditions [33].

Consider the general non-convex optimization problem:

$$\begin{cases} \max\limits_{\mathbf{x}} & f_0(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \geq 0, \quad \text{for all } i = 1, 2, \ldots, m. \end{cases} \tag{2.69}$$

The following proposition is a direct consequence of [32] and provides the means to devise a low complexity method to generate candidate solutions for (2.69) [19].

Proposition 2.14. *Let the feasible set of (2.69) be compact and the functions $f_0, f_1, \ldots$ be differentiable. Consider the family of optimization problems:*

$$\max_{\mathbf{x}} \; g_{0,j}(\mathbf{x}) \quad \text{s.t.} \quad g_{i,j}(\mathbf{x}) \geq 0, \; \textit{for all } i = 1, 2, \ldots, m, \tag{2.70}$$

with solution $\mathbf{x}_j^*$ where, for all j and $i = 0, 1, \ldots, m$, $g_{i,j}$ are differentiable functions such that the feasible set is compact. Assume that, for all j and $i = 0, \ldots, m$, $g_{i,j}(\mathbf{x}_{j-1}^*) = f_i(\mathbf{x}_{j-1}^*)$ and $g_{i,j}(\mathbf{x}) \leq f_i(\mathbf{x})$ for all $\mathbf{x}$. Then, the sequence $\{f_0(\mathbf{x}_j^*)\}$ is monotonically increasing and converges to a finite limit f^*. If, in addition, $\nabla g_{i,j}(\mathbf{x}_{j-1}^*) = \nabla f_i(\mathbf{x}_{j-1}^*)$ for all j and $i = 0, \ldots, m$, then every limit point of $\{\mathbf{x}\}_j$ with objective value f^* satisfies the KKT conditions of (2.69) under suitable constraint qualifications.

Proposition 2.14 establishes that a sequence of feasible points $\{\mathbf{x}_j^*\}$ with increasing objective value $f_0(\mathbf{x}_j^*)$ can be generated by solving a sequence of approximate problems. For this tool to be of practical use, it is essential to obtain suitable approximate problems that have much lower computational complexity than (2.69).

Next, we discuss the application of Proposition 2.14 to GEE maximization following the approach in [19]. Further applications are, e.g., weighted minimum EE maximization [19], weighted sum EE maximization [34], and EE resource allocation

over multiple resource blocks [22]. It was also applied to other related system models, e.g., to full-duplex multiuser MIMO systems in [35], for BC IC in [36], or for MISO BC in [37].

Consider the generic GEE maximization problem:

$$\begin{cases} \max\limits_{\mathbf{p}} & \dfrac{\sum_{k=1}^{K} f_k^+(\mathbf{p}) - f_k^-(\mathbf{p})}{\sum_{k=1}^{K} \mu_k p_k + P_c} \\ \text{s.t.} & g_i^+(\mathbf{p}) - g_i^-(\mathbf{p}) \geq 0, \text{ for all } i = 1, 2, \ldots, m \\ & \mathbf{p} \in [\mathbf{0}, \mathbf{P}], \end{cases} \tag{2.71}$$

with $f_k^+, f_k^-, g_i^+, g_i^-$ being concave functions on $[\mathbf{0}, \mathbf{P}]$ for all i and k. Since concave functions are upper bounded by their first-order Taylor expansion [3, Section 3.1.3], each term in the objective's numerator is lower bounded as

$$f_k^+(\mathbf{p}) - f_k^-(\mathbf{p}) \geq f_k^+(\mathbf{p}) - \left[f_k^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} f_k^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T(\mathbf{p} - \mathbf{p}_j)\right]. \tag{2.72}$$

Similarly, the constraints are bounded as

$$g_i^+(\mathbf{p}) - g_i^-(\mathbf{p}) \geq g_i^+(\mathbf{p}) - \left[g_i^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} g_i^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T(\mathbf{p} - \mathbf{p}_j)\right]. \tag{2.73}$$

These underestimators are tight at $\mathbf{p} = \mathbf{p}_j$, i.e., at this point equality holds in (2.72) and (2.73). Thus, these are suitable functions to construct the approximate problems as

$$\begin{cases} \max\limits_{\mathbf{p}} & \dfrac{\sum_{k=1}^{K} f_k^+(\mathbf{p}) - \left[f_k^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} f_k^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T(\mathbf{p} - \mathbf{p}_j)\right]}{\sum_{k=1}^{K} \mu_k p_k + P_c} \\ \text{s.t.} & g_i^+(\mathbf{p}) - \left[g_i^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} g_i^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T(\mathbf{p} - \mathbf{p}_j)\right] \geq 0, \\ & \qquad\qquad\qquad\qquad \text{for all } i = 1, 2, \ldots, m \\ & \mathbf{p} \in [\mathbf{0}, \mathbf{P}]. \end{cases} \tag{2.74}$$

Let $\mathbf{p}_{j-1}^*$ be an optimal solution of (2.74) and denote by $f(\mathbf{p})$ the objective of (2.71). If the approximation point $\mathbf{p}_j = \mathbf{p}_{j-1}^*$ for all $j \geq 1$ with $\mathbf{p}_0$ some feasible point in (2.71), then the sequence $\{f(\mathbf{p}_j^*)\}$ is monotonically increasing and converges to a value $\bar{f}$ [19, Proposition 6].

Further, the gradient of the constraint approximation is

$$\begin{aligned} &\nabla_{\mathbf{p}}\left(g_i^+(\mathbf{p}) - \left[g_i^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} g_i^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T(\mathbf{p} - \mathbf{p}_j)\right]\right) \\ &= \nabla_{\mathbf{p}} g_i^+(\mathbf{p}) - (\nabla_{\mathbf{p}} g_i^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j}), \end{aligned} \tag{2.75}$$

while that of the constraints in (2.71) is $\nabla g_i^+(\mathbf{p}) - \nabla g_i^-(\mathbf{p})$. With $\mathbf{p}_j = \mathbf{p}_{j-1}^*$, these are equal at $\mathbf{p}_{j-1}^*$ and, thus, satisfy the third condition in Proposition 2.14. The gradients of the objective and objective approximation functions are

$$\begin{aligned}
&\nabla_{\mathbf{p}} \left(\frac{\sum_{k=1}^K f_k^+(\mathbf{p}) - f_k^-(\mathbf{p})}{\sum_{k=1}^K \mu_k p_k + P_c} \right) \\
&= \frac{\nabla_{\mathbf{p}} \left(\sum_{k=1}^K f_k^+(\mathbf{p}) - f_k^-(\mathbf{p}) \right)}{\sum_{k=1}^K \mu_k p_k + P_c} - \frac{\left(\sum_{k=1}^K f_k^+(\mathbf{p}) - f_k^-(\mathbf{p}) \right) \mu}{\left(\sum_{k=1}^K \mu_k p_k + P_c \right)^2},
\end{aligned} \tag{2.76}$$

and

$$\begin{aligned}
&\nabla_{\mathbf{p}} \left(\frac{\sum_{k=1}^K f_k^+(\mathbf{p}) - \left[f_k^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} f_k^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T (\mathbf{p} - \mathbf{p}_j) \right]}{\sum_{k=1}^K \mu_k p_k + P_c} \right) \\
&= \frac{\nabla_{\mathbf{p}} f_i^+(\mathbf{p}) - (\nabla_{\mathbf{p}} f_i^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})}{\sum_{k=1}^K \mu_k p_k + P_c} \\
&\quad - \frac{\left(\sum_{k=1}^K f_k^+(\mathbf{p}) - \left[f_k^-(\mathbf{p}_j) + (\nabla_{\mathbf{p}} f_k^-(\mathbf{p})|_{\mathbf{p}=\mathbf{p}_j})^T (\mathbf{p} - \mathbf{p}_j) \right] \right) \mu}{\left(\sum_{k=1}^K \mu_k p_k + P_c \right)^2},
\end{aligned} \tag{2.77}$$

respectively. Again, with $\mathbf{p}_j = \mathbf{p}_{j-1}^*$, the first terms are obviously equal at $\mathbf{p}_{j-1}^*$ and the second terms are equal at this point because the approximation is tight at $\mathbf{p}_j$. Hence, the third condition in Proposition 2.14 is satisfied and, thus, any limit point of the sequence $\{\mathbf{p}_j^*\}$ with objective value $\bar{f}$ is a KKT point under suitable constraint qualifications [19, Proposition 6].

The resulting algorithm is straightforward and stated in Algorithm 2.6. Convergence to a first-order optimal solution follows from the discussion made earlier and Proposition 2.14.

Experience shows that in the case of GEE maximization the point obtained by Algorithm 2.6 is usually a global maximizer of (2.71) [22]. However, we stress that the obtained solution is in no way guaranteed to be even a locally optimal solution and that this observation was only possible by extensive numerical experiments involving global optimization methods for verification.

Algorithm 2.6: Sequential convex approximation

Initialize $j = 0$ and choose a feasible $\mathbf{p}_0$.
repeat
  Solve (2.74) and denote its solution as $\mathbf{p}_j^*$
  $\mathbf{p}_{j+1} = \mathbf{p}_j^*$
  $j \leftarrow j + 1$
until convergence.

The successive convex approximation method discussed here belongs to the larger class of majorization–minimization algorithms reviewed in [38]. A completely different approach is to employ deep learning methods that are trained with globally optimal solutions to predict the optimal power allocation [39]. While the training phase is costly, the actual power allocation is tremendously fast and often much closer to the globally optimal solution than first-order optimal approaches [40,41].

2.6 Conclusions

We have reviewed generalized concavity and fractional programming theory and discussed their application to energy-efficient resource allocation for point-to-point links. We have motivated the need for more advanced global optimization techniques for resource allocation in interference networks and introduced the reader to BB methods. State-of-the-art monotonic optimization and monotonic fractional programming were explained in the context of BB methods. The MMP framework was presented and it was shown that this novel bounding approach outperforms monotonic fractional programming by several orders of magnitude. Since the MMP bounding approach is much more versatile than state-of-the-art monotonic optimization and often leads to much faster convergence, this approach has the potential to replace classical monotonic programming in the future.

We have also discussed potential issues with non-convex feasible sets and presented the SIT approach as remedy. In particular, most global optimization problems where a conclusive feasibility check is not available or has high computational cost can benefit from this method. In an application example, we have shown that the SIT scheme can also be used to avoid complicated decomposition methods. It is shown in [29] that the resulting optimization framework is well suited for resource allocation in non-orthogonal interference networks.

Although considerable progress has been made toward efficient global optimization, the exponential complexity in the number of variables prevents the solution of large-scale problems. For this reason, we have also introduced the popular successive convex approximation framework that is often employed to obtain stationary points of non-convex optimization problems.

2.6.1 Further reading

A classic text on generalized concavity is [42]. Fractional programming for EE maximization is treated in [6].

A broad and very general coverage of BB methods can be found in [43, Section IV]. The treatment in [12] is more modern and better suited for a first exposition. Reference [25] focuses modeling on resource allocation and the connection between BB and the SIT scheme. Good references for the mentioned Polyblock algorithm are [12] and [21].

The MMP framework is discussed in much more detail in [26] with several application examples. The source code, including a fairly general implementation

of Algorithm 2.2, is published alongside [26]. A specialized bounding method for EE maximization with even faster convergence than MMP is developed in [41]. Other applications of BB methods for resource allocation are [44] where multiple-input–single-output beamforming is considered and [45] where weighted sum rate maximization is reviewed.

The SIT scheme is developed in a series of publications [27,30,31] and is included in the text book [12]. Besides being applied to resource allocation, the fundamentals of the SIT scheme are also reviewed in [25,29].

Successive convex approximation is developed in [32]. It is shown that it often converges to the globally optimal solution for GEE maximization in [19]. A generalization to pseudo-convex approximation functions can be found in [46]. This work is extended to QoS constraints in [47]. The larger class of majorization–minimization algorithms is reviewed in [38]. An alternative approach to the efficient implementation of resource allocation algorithms is deep-learning based methods [39]. Although this approach does not have any optimality guarantee, it is shown in [41] that it can outperform successive convex approximation. While this is a promising approach to implement almost globally optimal resource allocation in real-world systems, there are still some open problems that need to be addressed first.

References

[1] Chiang M, Low SH, Calderbank AR, *et al.* Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures. Proc IEEE. 2007;95(1):255–312.

[2] Altman E, Avrachenkov K, and Garnaev A. Generalized α-Fair Resource Allocation in Wireless Networks. In: Proc. IEEE Conf. Decision, Control (CDC). Cancun, Mexico; 2008. pp. 2414–2419.

[3] Boyd S and Vandenberghe L. Convex Optimization. Cambridge, UK; 2004.

[4] Zhang H, Gladisch A, Pickavet M, *et al.* Energy Efficiency in Communications. IEEE Commun Mag. 2010;48(11):48–49.

[5] Saito Y, Kishiyama Y, Benjebbour A, *et al.* Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In: Proc. IEEE Veh. Technol. Conf. (VTC Spring). Dresden, Germany; 2013.

[6] Zappone A and Jorswieck EA. Energy Efficiency in Wireless Networks via Fractional Programming Theory. vol. 11 of FnT Commun. Inf. Theory. Boston, MA, USA; 2015.

[7] Isheden C, Chong Z, Jorswieck EA, *et al.* Framework for Link-Level Energy Efficiency Optimization with Informed Transmitter. IEEE Trans Wireless Commun. 2012;11(8):2946–2957.

[8] Chung BC, Lee K, and Cho D. Proportional Fair Energy-Efficient Resource Allocation in Energy-Harvesting-Based Wireless Networks. IEEE Syst J. 2018;12(3):2106–2116.

[9] Dinkelbach W. On Nonlinear Fractional Programming. Manage Sci. 1967; 13(7):492–498.

[10] Schaible S. Fractional Programming. II, On Dinkelbach's Algorithm. Manage Sci. 1976;22(8):868–873.
[11] Charafeddine MA, Sezgin A, Han Z, *et al.* Achievable and Crystallized Rate Regions of the Interference Channel with Interference as Noise. IEEE Trans Wireless Commun. 2012;11(3):1100–1111.
[12] Tuy H. Convex Analysis and Global Optimization. vol. 110 of Springer Optim. Appl. 2nd ed. New York; Berlin, Germany; Vienna, Austria; 2016.
[13] Freund RW and Jarre F. Solving the Sum-of-Ratios Problem by an Interior-Point Method. J Global Optim. 2001;19(1):83–102.
[14] Jagannathan R. On Some Properties of Programming Problems in Parametric form Pertaining to Fractional Programming. Manage Sci. 1966;12(7): 609–615.
[15] Almogy Y and Levin O. A Class of Fractional Programming Problems. Oper Res. 1971;19(1):57–67.
[16] Falk JE and Palocsay SW. Optimizing the Sum of Linear Fractional Functions. In: Floudas CA and Pardalos PM, editors. Recent Advances in Global Optimization. Princeton Ser. Comput. Sci. Princeton, NJ, USA; 1992. pp. 221–258.
[17] Nesterov Y and Nemirovski A. Interior-Point Polynomial Algorithms in Convex Programming. vol. 13 of SIAM Stud. Appl. Numer. Math.; 1994.
[18] Matthiesen B, Zappone A, and Jorswieck EA. Resource Allocation for Energy-Efficient 3-Way Relay Channels. IEEE Trans Wireless Commun. 2015;14(8):4454–4468.
[19] Zappone A, Björnson E, Sanguinetti L, *et al.* Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks. IEEE Trans Signal Process. 2017;65(11):2844–2859.
[20] Tuy H. Monotonic Optimization: Problems and Solution Approaches. SIAM J Optim. 2000;11(2):464–494.
[21] Zhang YJ, Qian LP, and Huang J. Monotonic Optimization in Communication and Networking Systems. vol. 7 of FnT Netw. Boston, MA, USA; 2012.
[22] Zappone A, Sanguinetti L, Bacci G, *et al.* Energy-Efficient Power Control: A Look at 5G Wireless Technologies. IEEE Trans Signal Process. 2016;64(7):1668–1683.
[23] Rudin W. Principles of Mathematical Analysis. 3rd ed. Int. Ser. Pure Appl. Math. New York, NY, USA; 1976.
[24] Luo ZQ and Zhang S. Dynamic Spectrum Management: Complexity and Duality. IEEE J Sel Areas Commun. 2008;2(1):57–73.
[25] Matthiesen B. Efficient Globally Optimal Resource Allocation in Wireless Interference Networks [Ph.D. Thesis]. Technische Universität Dresden. Dresden, Germany; 2019. Available from: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-362878.
[26] Matthiesen B, Hellings C, Jorswieck EA, and Utschick W. Mixed Monotonic Programming for Fast Global Optimization. IEEE Trans Signal Process. 2020; 68:2529–2544.

[27] Tuy H, Al-Khayyal F, and Thach PT. Monotonic Optimization: Branch and Cut Methods. In: Audet C, Hansen P, and Savard G, editors. Essays and Surveys in Global Optimization; 2005. pp. 39–78.

[28] Matthiesen B, Hellings C, and Jorswieck EA. Energy Efficiency: Rate Splitting vs. Point-to-Point Codes in Gaussian Interference Channels. In: Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC); 2019.

[29] Matthiesen B and Jorswieck EA. Efficient Global Optimal Resource Allocation in Non-Orthogonal Interference Networks. IEEE Trans Signal Process. 2019;67(21):5612–5627.

[30] Tuy H. Robust Solution of Nonconvex Global Optimization Problems. J Global Optim. 2005;32(2):307–323.

[31] Tuy H. $\mathscr{D}(\mathscr{C})$-Optimization and Robust Global Optimization. J Global Optim. 2009;47(3):485–501.

[32] Marks BR and Wright GP. A General Inner Approximation Algorithm for Nonconvex Mathematical Programs. Oper Res. 1978;26(4):681–683.

[33] Murty KG and Kabadi SN. Some NP-complete Problems in Quadratic and Nonlinear Programming. Math Program. 1987;39(2):117–129.

[34] Matthiesen B, Yang Y, and Jorswieck EA. Optimization of Weighted Individual Energy Efficiencies in Interference Networks. In: Proc. IEEE Wireless Commun. Netw. Conf. (WCNC). Barcelona, Spain; 2018.

[35] Huberman S and Le-Ngoc T. Full-Duplex MIMO Precoding for Sum-Rate Maximization with Sequential Convex Programming. IEEE Trans Veh Technol. 2015;64(11):5103–5112.

[36] Kaleva J, Tölli A, and Juntti M. Decentralized Sum Rate Maximization with QoS Constraints for Interfering Broadcast Channel via Successive Convex Approximation. IEEE Trans Signal Process. 2016;64(11): 2788–2802.

[37] Tervo O, Pennanen H, Christopoulos D, *et al.* Distributed Optimization for Coordinated Beamforming in Multicell Multigroup Multicast Systems: Power Minimization and SINR Balancing. IEEE Trans Signal Process. 2018;66(1):171–185.

[38] Sun Y, Babu P, and Palomar DP. Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning. IEEE Trans Signal Process. 2017;65(3):794–816.

[39] Zappone A, Di Renzo M, and Debbah M. Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both? IEEE Trans Commun. 2019;67(10):7331–7376.

[40] Matthiesen B, Zappone A, Jorswieck E, *et al.* Deep Learning for Real-Time Energy-Efficient Power Control in Mobile Networks. In: Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC); 2019.

[41] Matthiesen B, Zappone A, Besser K, Jorswieck EA, and Debbah M. A Globally Optimal Energy-Efficient Power Control Framework and Its Efficient Implementation in Wireless Interference Networks. IEEE Trans Signal Process. 2020;68:3887–3902.

[42] Avriel M, Diewert WE, Schaible S, *et al.* Generalized Concavity. Math. Concepts Methods Sci. Eng. New York, NY, USA; 1988.
[43] Horst R and Tuy H. Global Optimization: Deterministic Approaches. 3rd ed. New York; Berlin, Germany; Vienna, Austria; 1996.
[44] Tervo O, Tran LN, and Juntti M. Optimal Energy-Efficient Transmit Beamforming for Multi-User MISO Downlink. IEEE Trans Signal Process. 2015;63(20):5574–5588.
[45] Weeraddana PC, Codreanu M, Latva-aho M, *et al.* Weighted Sum-Rate Maximization in Wireless Networks: A Review. vol. 6 of FnT Netw. Boston, MA, USA; 2012.
[46] Yang Y and Pesavento M. A Unified Successive Pseudoconvex Approximation Framework. IEEE Trans Signal Process. 2017;65(13):3313–3328.
[47] Yang Y, Pesavento M, Chatzinotas S, *et al.* Energy Efficiency Optimization in MIMO Interference Channels: A Successive Pseudoconvex Approximation Approach. IEEE Trans Signal Process. 2019;67(15):4107–4121.

Chapter 3

Deep learning for energy-efficient beyond 5G networks

Alessio Zappone[1], Marco Di Renzo[2] and Merouane Debbah[3]

3.1 Introduction

Future wireless networks are expected to serve an exponentially increasing amount of connected devices, which poses green and sustainable growth concerns. Already 5G networks, to be rolled out in 2020, will have to provide 2,000 × higher bit/J energy efficiency compared to the previous wireless generation [1–3], and the numbers will only escalate. It is estimated that the compound annual growth rate of connected devices will rise by 55% annually, reaching 607 exabytes in 2025 and 5,016 exabytes in 2030 [4]. Moreover, besides the huge volume of traffic to support, another critical challenge to be faced by future wireless networks will be the extreme heterogeneity, with many innovative vertical services to be provided, each with its own specific requirements [5]:

- End-to-end latency of 1 ms and reliability higher than 99.999% for ultra reliable low latency communications.
- Terminal densities of 1 million of terminals per square kilometer for massive Internet of Things applications.
- Per-user data-rate larger than 50 Mb/s for mobile broadband applications.
- Terminal location accuracy of the order of 0.1 m for vehicular-to-X communications.

All of these point toward wireless networks characterized by an unprecedented level of complexity, which makes traditional design approaches not suitable anymore. The design of past and present generations of wireless networks has always relied on the use of *mathematical models* that are obtained from either theoretical considerations or

[1]Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy
[2]Laboratory of Signals and Systems (CNRS – CentraleSupelec – Univ. Paris-Sud), Université Paris-Saclay, Gif-sur-Yvette, France
[3]Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France

field measurements. Unfortunately, this approach is rapidly becoming unfeasible due to the huge complexity of future wireless networks, which makes any mathematical model either too complex for practical designs, or not accurate enough.

In order to face this **complexity crunch**, it is not enough to simply devise more performing transmission technologies, but instead the conventional approaches to system design must be radically rethought. A promising approach to this end is the use of deep learning and artificial neural networks (ANNs) [6], as it will be explained in the rest of this chapter.

3.1.1 AI-based wireless networks

The complexity crunch challenge can be tackled by making wireless networks *intelligent*, i.e., able to determine the best policy automatically, with very limited need for human intervention. This motivates to endow wireless networks with **artificial intelligence (AI)** capabilities, with the aim of making them able to determine the best policy to employ based on the experience obtained by processing previous data. Regarding this point, we should explicitly state that although data-driven approaches reduce the need of mathematical models for network design and operation, it is a major point of this chapter to show that mathematical models can still be used despite their cumbersomeness or inaccuracy, to complement and improve purely data-driven methods.

But how to embed AI into wireless networks? A framework that goes in this direction is that of **deep learning** [7–9], which is a specific machine learning technique that implements the learning process by ANNs. Although being the most popular machine learning tool, and being a consolidated reality in many fields of science, deep learning has been proposed for wireless networks only very recently. This was due to the fact that, unlike other fields of science, wireless networks have always admitted a suitable mathematical modeling, which made data-driven approaches not strictly necessary. However, as we have mentioned earlier, the increasing complexity of wireless networks makes data-driven approaches more and more appealing. In addition, there are other recent factors that facilitate the use of deep learning for wireless networks:

- The exponential increase of wireless devices results in a corresponding growth of traffic data [10–12], which can be exploited to train ANNs.
- Modern advancements in computing capacity, such as the use of graphics processing units, make it possible to execute larger and more complex algorithms much faster.

The use of deep learning in wireless communication is being supported also by industry players [13,14], as well as by regulatory bodies, such as European Telecommunications Standards Institute, which activated an Industry Specification Group named *Experiential Network Intelligence*, and International Telecommunication Union (ITU), which recently approved the ITU-T Y.3172 architectural framework for machine learning in future networks, including IMT-2020 [15].

Despite all these interests and efforts, there are some critical issues to be understood in order to realize the vision of AI-based wireless networks. Specifically:

- **Data acquisition.** Deep learning requires a large amount of training data to perform well. As mentioned earlier, this data is now available over the air, but the specific mechanisms to acquire it, store it, and process it are not clear yet. The next section puts forth the concept of smart radio environments that are based on the use of innovative *intelligent* materials, known as **meta-materials**, which have communication as well as data storage and processing abilities. Smart radio environments appear as a promising solution for truly intelligent wireless networks. In addition, another useful approach to reduce the amount of required data is the cross-fertilization between data-driven and model-based techniques. This point will be exemplified in the rest of this chapter by presenting a concrete case study.
- **ANNs deployment into communication networks.** While it appears clear that future communication networks will have to rely on AI, it is not clear where and how ANNs should be deployed into communication networks. Should the acquired data be stored at a centralized location, where a single ANN manages a large network domain, or should each network device store its own data and run a local ANN? This point is also discussed in the next section.

3.2 Integration into wireless networks: smart radio environments

As already anticipated earlier, future wireless networks will provide services beyond communications among people and objects [16]. Future wireless networks will become a distributed intelligent wireless communication, sensing, and computing platform, which, besides communications, will be capable of sensing the environment, as well as storing and processing data. Future wireless networks will be required to overcome the challenge of interconnecting the physical and digital worlds in a seamless and sustainable manner [17,18].

As already mentioned, in order to turn the vision of future networks into reality, it is not sufficient anymore to rely solely on wireless networks whose *logical* operation is software controlled and optimized by traditional approaches [19]. Instead, the *wireless environment* itself needs to be made intelligent [20], capable of optimizing itself to ensure seamless connectivity. We refer to a wireless environment with these characteristics as a *smart radio environment* [21].

Our vision is better described through Figure 3.1. Present wireless networks follow the well-known Shannon paradigm [22], according to which the environment is given and modeled in terms of transition probabilities (i.e., Pr $\{y|x\}$), or the Wiener

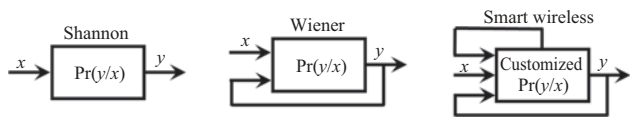


Figure 3.1 Current networks versus a smart radio environment (or smart wireless)

paradigm [23], according to which the environment is still given, but its effect can be controlled by means of feedback signals. Instead, a smart radio environment is characterized by environmental objects whose electromagnetic response can be designed in order to have a desired effect on the electromagnetic signals. Thus, the input signal and the response of the environmental objects to the radio waves can be jointly optimized for improved performance. For example, the input signal is steered toward a given environmental object, which reflects it toward the receiver by suitably optimized phase shifts. In turn, the receiver is also steered toward the incoming signal.

But how to implement such a futuristic concept? Several possible implementations of smart radio environments are currently emerging [24–35], with the use of reconfigurable intelligent surfaces (RIS) appearing as the most promising approach [36]. Intelligent surfaces, also called meta-surfaces, are thin bidimensional structures made of a special material, called meta-material, which can be reconfigured in order to act on incoming radio waves in a programmable way [37]. Thus, RIS have the potential of increasing the reliability of information transfer and processing [38] while providing a suitable distributed platform-to-perform low-energy and low-complexity sensing [34], storage [28], and analog computing [33].

3.2.1 The role of deep learning in smart radio environments

After introducing the paradigm of smart radio environments, let us discuss its connection with deep learning. This section aims at putting forth the idea that smart radio environment and deep learning are intertwined, enabling each other.

To begin with, let us discuss why smart radio environments enable the integration of AI into wireless networks. As discussed, besides the ability of improving the communication performance, meta-surfaces are also able to sense the surrounding environment and store the sensed data. This makes meta-surfaces the perfect platform for data acquisition and processing, which is an essential requirement of deep learning. In other words, meta-surfaces provide the fabric of future AI-based wireless networks. Due to the pervasive use of meta-surfaces, smart radio environments will be naturally able to acquire and harness large datasets from the signals that travel over the communication networks. In this sense, smart radio environments represent an enabler for the implementation of AI-based communication networks.

On the other hand, as already mentioned, smart radio environments offer the possibility of designing not only the transmit and receive strategies but also the transfer function of the environment. This significantly increases the amount of variables that can be optimized, which leads to an equally significant increase of the computational complexity required to perform the design. In a smart radio environment, the operation of each environmental object is an optimization variable, besides the transmitter and receiver strategies. Accurately modeling such an emerging network scenario and optimizing it in real time appear a very challenging task, which cannot be tackled by available optimization frameworks that in the best case require a polynomial complexity in the number of variables every time the resource allocation problem needs to be solved, i.e., when the propagation channel has changed. Moreover, the feedback requirements of smart radio environments are also much higher than in present

wireless networks, which further increases the complexity of the resource allocation process and leads to significant overheads. In this context, as it will be shown in detail in the rest of this chapter, the use of deep learning can prove useful to significantly simplify the resource management task. In this sense, deep learning enables smart radio environments, making their optimization feasible in terms of computational complexity. As it will be shown in Section 3.4 with reference to the specific case study of energy efficiency maximization, merging model-based approaches and deep learning tools, it is possible to obtain near-optimal designs of complex wireless networks, with a complexity that is affordable for online implementation. Considering the specific energy-efficient aspect, this will have at least two major advantages that will be exemplified in the rest of this chapter.

- Simplifying the resource management task enables to effectively maximize the energy efficiency of complex networks, even in scenarios that are considered too complex with present optimization techniques. This is anticipated to significantly increase the energy efficiency of the operating points of future wireless networks.
- Even in scenarios where near-optimal energy efficiency optimization is already possible by traditional optimization approaches, the use of deep learning is able to reduce the computational complexity, thus yielding considerable energy savings in the digital signal processor.

3.2.2 ANNs deployment into wireless networks

In order to successfully use deep learning for wireless communications, a key question is how to integrate ANNs into existing and future wireless network topologies. Otherwise stated, what is the most efficient way to store and process the data to be used by ANNs? This is a question that is specific to wireless networks, in the sense that in other application fields of deep learning, the usual approach is to have a centralized "artificial brain" that carries out the task at hand. However, this ideal solution is not practical in the wireless context, since wireless networks have specific requirements to be met. In particular, a centralized "artificial brain" that oversees the management of a whole network, dictating the actions to take to the edge-users, is problematic due to at least the following major points:

1. **Latency.** One major goal of future 5G wireless networks is to reduce the end-to-end communication latency, which, for some applications, is required to be lower than a millisecond. Thus, if this constraint is to be fulfilled, it is not possible to wait for the cloud to perform the computations and then feed back the results. Instead, the computations should be performed locally by each user equipment (UE).
2. **Privacy.** The privacy and security of wireless communication will be critical issues to turn the 5G vision of the "everything-connected world" into reality. This implies that for some vertical applications, it is not desirable to share information with the cloud, which makes cloud-based deep learning not possible.
3. **Connectivity.** One major goal of future 5G networks is to provide connectivity everywhere and every time. This makes a cloud-based AI problematic, since it makes an edge device too reliant on the cloud, even when no reliable connection

to cloud services is available. Instead, mobile terminals should possess some "local intelligence" to be able to operate in these scenarios, too.

As a result, like in human society, where there are both a collective intelligence that is a common heritage and an individual intelligence, wireless networks should possess a *cloud intelligence* that should be accessible to all nodes, as well as a *device intelligence* that belongs to each individual device. Clearly, this poses several issues that will have to be addressed in order to successfully implement AI-based wireless networks. Endowing each network node with AI capabilities will significantly impact not only the transmission technologies, but also the way the network should be controlled through feedback signals to avoid instability and malfunctioning. In a scenario in which each network node will have its own "brain," i.e., its own ANN, the interactions among the different devices should be carefully studied and mechanisms to avoid performance impairments, data inconsistencies, and system failures will have to be devised.

3.3 State-of-the-art review

The application of deep learning to the design of the physical layer of wireless communication networks has started attracting research attention only very recently, mostly in the last couple of years. A comprehensive survey/tutorial on this topic is [6] that also provides several concrete examples of wireless network design by means of ANNs.

Focusing specifically on the use of ANNs for resource allocation purposes, we cite the following contributions [29,30,39–48].

In [39,43], the idea of using ANNs for network resource management is proposed, providing an overview of potential applications of AI for network resource management in future 5G wireless networks and discussing supervised, unsupervised, and reinforcement learning. In [45], a fully connected ANN is used for sum-rate maximization in interference-limited networks. Specifically, the ANN is trained to mimic the performance of the weighted MMSE resource allocation algorithm [32] but with lower computational complexity. Instead, [6,26,47] consider the problem of energy efficiency maximization and propose to train a fully connected neural network based on the optimal power allocation rule, which is computed based on a novel branch-and-bound procedure proposed in [26], and which is amenable to offline implementation. The results indicate that the optimal performance can be approached with limited online complexity, thus enabling an online implementation. The effectiveness of this approach is demonstrated in [30,48] for power control and user-cell association in massive MIMO multicell systems. Instead, a different approach is taken in [44], where a fully connected ANN is trained to solve the sum-rate maximization problem subject to the maximum power and minimum rate constraints. In order to reduce the complexity of building the training set, the authors propose to train the ANN using directly the system sum-rate as training cost function. The results show a gain compared with previous low-complexity optimization methods.

In [40], a cloud-RAN system with caching capabilities is considered. Echo-state neural networks are used to enable base stations (BSs) to predict the content

request distribution and mobility pattern of each user, thus determining the best content to cache. An improvement of the network sum effective capacity of around 30% compared with baseline approaches based on random caching is observed. In [42], deep reinforcement learning is used to develop a power control algorithm for a cognitive radio system, in which a primary and a secondary users share the spectrum. The use of deep learning enables both the users to meet their QoS requirements despite the fact that the secondary user has no information about the primary user's transmit power. The use of deep reinforcement learning is also considered in [29] to propose a power control algorithm for weighted sum-rate maximization in interference channels subject to the maximum power constraints. In [31], online power allocation policies for a large and distributed system with energy-harvesting nodes are developed by merging deep reinforcement learning and mean field games. It is shown that the proposed method outperforms all other available online policies and suffers a limited gap compared to the use of noncausal offline policies.

Parallel to the research on the use of deep learning in communication systems, considerable work is also being made on smart radio environments. Current research efforts toward making the vision of smart radio environments true are mostly focused on implementing hardware testbeds or on realizing point-to-point experimental tests [24–35]. Few technical results are available about the performance of RIS-based wireless networks. A first contribution in this direction is [49] where alternating optimization is used to allocate the RIS phase shifts and the BS beamforming in a MISO downlink system, with the aim of maximizing the system sum-rate and energy efficiency. A similar setup is considered in [50], with the difference that the problem of minimizing the power consumption is addressed. In [51], again a MISO downlink system is considered, with the addition that the OFDM transmission scheme is considered, and the problem of sum-rate maximization is addressed. Similarly, alternating optimization methods are used in [52] to tackle again the problem of sum-rate maximization in a MISO downlink system. The BS beamformer and the RIS phase shifts are optimized, with the additional difficulty that discrete phase shifts at the RIS are assumed. In [53], an RIS is used to enhance the secrecy rate of a MISO downlink channel with multiple eavesdropper. Alternating maximization is again used to come up with a practical, yet suboptimal, method to optimize the transmit beamformer and the RIS phase shifts. In [54], the asymptotic rate of an RIS-based system is derived in the limit of the number of reflecting elements at the RIS growing large. In [27], the minimum signal-to-interference plus noise ratio (SINR) achieved by linear detection in downlink RIS-based systems is characterized considering line-of-sight between the BS and the RIS. In [55,56], it is shown that RIS can be used to implement phase-shift-keying and spatial modulation techniques.

3.4 Energy efficiency optimization by deep learning

This section presents two case studies that illustrate how deep learning can be used in conjunction with traditional mathematical approaches to perform energy-efficient wireless network design. Two types of applications will be discussed. The first

considers the case in which the resource allocation problem at hand admits an accurate mathematical formulation, which is, however, too complex to be solved with practical complexity. Instead, the second approach considers the situation in which no accurate mathematical model is available for the problem at hand, which calls for the use of approximate models.

The approach to tackle the first scenario is based on the property that ANNs are universal function approximators [57], i.e., their input–output map can be tuned to emulate any continuous map. This property can be exploited observing that any resource allocation problem can be regarded as a map from the ensemble of all network parameters of interest, denoted by $\mathbf{d} \in \mathbb{R}^N$, to the corresponding optimal resource allocation $\mathbf{x}^* \in \mathscr{S}$, with $\mathscr{S}$ denoting the set of feasible resource allocations. Formally speaking:

$$\mathscr{F}: \mathbf{d} \in \mathbb{R}^N \rightarrow \mathbf{x}^* \in \mathscr{S} \subseteq \mathbb{S}. \tag{3.1}$$

Then, an ANN can be trained to emulate the unknown map $\mathscr{F}$. This enables to optimize a desired performance function for given system parameters without explicitly having to solve the resource allocation problem by numerical optimization methods, but rather using the trained ANN to receive as input the current realization of system parameters, and reading the corresponding optimal resource allocation as the output of the ANN. This has the huge advantage that any time the system parameters change, it is not needed to solve again an optimization problem, but it is sufficient to change the input of the ANN, which will compute the output by means of a simple forward propagation.

On the other hand, this approach does not directly apply to the second scenario, since no reliable model is available. However, if only an approximate model is available, it is still possible to use it to pretrain an ANN by the same approach outlined earlier. Next, the configuration of the ANN can be refined by means of a second training phase that assumes the availability of a training set containing empirical data samples and employs tools from transfer learning theory. The details of both the approaches are explained in detail in the rest of this section.

3.4.1 Weighted sum energy efficiency maximization

One of the key requirements of future wireless networks is recognized to be a massive increase of the global energy efficiency measured in bits reliably transmitted over Joule of the energy consumed. Accordingly, let us consider the uplink of a multicell network with M BSs and K mobile users. Each BS is equipped with N antennas, whereas the mobile users have a single antenna. Let $\mathbf{h}_{k,m}$ be the $N \times 1$ channel from user k to BS m, p_k be the kth user's transmit power, $\mathbf{c}_k$ the $N \times 1$ receive vector for user k, and σ_m^2 the received noise power at BS m. Then, the SINR enjoyed by user k at its intended receiver m_k is

$$\gamma_k = \frac{p_k |\mathbf{c}_k^H \mathbf{h}_{k,m_k}|^2}{\sigma^2 + \sum_{j \neq k} p_j |\mathbf{c}_k^H \mathbf{h}_{j,m_k}|^2} = \frac{p_k d_{k,k}}{\sigma^2 + \sum_{j \neq k} p_j d_{k,j}}, \tag{3.2}$$

with $d_{k,j} = |\mathbf{c}_k^H \mathbf{h}_{j,m_k}|^2$, for all k, j.

Based on (3.2), the network weighted sum energy efficiency (WSEE) is given by

$$\text{WSEE} = \sum_{k=1}^{K} w_k \frac{B \log_2 (1 + \gamma_k)}{P_{c,k} + \mu_k p_k} \text{ (bit/J)}, \tag{3.3}$$

wherein B is the communication bandwidth, $P_{c,k}$ is the hardware static power consumption of the kth communication link, μ_k is the inverse of the power amplifier efficiency of transmitter k, and w_k is a nonnegative weight modeling the importance given to the energy efficiency of link k. It is to be stressed that $P_{c,k}$ depends on system parameters such as the number of antennas and the efficiency of the system hardware components, but it is assumed not to depend on the transmit powers, and therefore the specific model expressing $P_{c,k}$ as a function of the system hardware components is inessential as far as maximizing (3.3) as a function of the transmit powers is concerned.

Maximizing the WSEE is considered the hardest type of energy-efficient maximization problems. Indeed, the sum of fractions are non-polynomial-hard (NP-hard) in general [58] and thus cannot be tackled with polynomial complexity by any available fractional programming technique. Moreover, each numerator of the summands of the WSEE is not concave. This implies that even the simpler special case of weighted sum rate maximization (obtained setting $\mu_i = 0$ for all $i = 1, \ldots, L$) is an NP-hard problem [25]. Thus, showing that ANNs can tackle the maximization of (3.3) represents a strong motivation for its use to tackle simpler problems, too, such as the maximization of the system sum-rate, or of other energy-efficient metrics, like the system global energy efficiency. Moreover, unlike system-wide energy-efficient metrics like the global energy efficiency, the WSEE provides the possibility of prioritizing the energy efficiencies of the individual users, through the choice of the weights $w_i \geq 0$. This might be useful in cases when some users require a higher energy efficiency, e.g., because they are powered by energy-harvesting techniques.

Thus, the power control problem is stated as the maximization of the WSEE subject to power constraints, namely,

$$\max_{\{p_k\}_{k=1}^K} \ \text{WSEE}(p_1, \ldots, p_K) \tag{3.4a}$$

$$\text{s.t. } P_{\min,k} \leq p_k \leq P_{\max,k}, \forall\, k = 1, \ldots, K, \tag{3.4b}$$

with $P_{\max,k}$ and $P_{\min,k}$ being the maximum feasible and minimum acceptable transmit powers for user k. Problem (3.4) is a so-called sum-of-ratios problem, which is considered the hardest class of fractional problems. Moreover, the difficulty of (3.4) is further increased by the fact that the numerators of (3.4a) are not concave functions of $\mathbf{p} = \{p_k\}_{k=1}^K$ due to the presence of multi-user interference. As a result, an ANN is able to learn the optimal map between the system channels, and the power vector that solves (3.4) makes a very strong case in favor of the use of deep learning for energy-efficient resource allocation in wireless networks.

At present, in order to solve (3.4), only global optimization methods are available, while more practical approaches guarantee only the first-order optimality. Moreover, as already mentioned, if traditional resource allocation methods were used, Problem (3.4) would have to be solved anew whenever the channel realizations $\{h_{\ell,m_k}\}_{k,\ell}$

change. This clearly is a critical drawback that prevents the use of optimization-theoretic approach for online resource allocation, i.e., following the small-scale variations of the channel coefficients. Indeed, the optimal transmit power must be computed well before the end of the channel coherence time in order to be practically useful.

Instead, adopting an ANN-based approach, we model the power control problem as the unknown map from the coefficients $\{d_{k,\ell}\}_{k,\ell}$ and the maximum/minimum transmit powers $P_{\max}$ and $P_{\min}$, to the optimal power allocation vector p^*, namely,

$$\mathscr{F}\colon \mathbf{d} = \{d_{k,\ell}, P_{\min,k}, P_{\max,k}\}_{k,\ell} \in \mathbb{R}^{K(M+2)} \rightarrow p^* \in \mathbb{R}^K, \tag{3.5}$$

and then training an ANN so that its input–output relationship reproduces the unknown map (3.5). This leads to considering an ANN $\mathscr{N}$ with $K(M+2)$ input nodes and K output nodes, to be trained so that it outputs the optimal $K \times 1$ power vector $\mathbf{p}^*$ corresponding to a given $K(M+2) \times 1$ input of system parameters $\mathbf{d}$. Specifically, exploiting the mathematical model represented by Problem (3.4) it is possible to generate a training set by solving offline many instances of Problem (3.4), corresponding to many different realizations of the system parameters $\mathbf{d}$. At this point, the considered ANN $\mathscr{N}$ can be trained by any training algorithm to learn the optimal map in (3.5). After the training phase, the trained ANN can be used online to infer the desired resource allocation corresponding to any system configuration by simply performing a forward propagation. Thus, the proposed ANN-based resource allocation framework can be divided into two phases, as described next:

1. **Offline phase.** During this phase, the ANN is trained and configured. It should be stressed that both the generations of the training set and the implementation of the training algorithm can take place offline and only sporadically, i.e., at a much longer timescale than the rate of change of the network parameters. Thus, the complexity of this phase becomes negligible in the long term. Moreover, a recent optimization framework proposed in [26] significantly simplifies the generation of the training set for energy-efficient resource allocation problems.
2. **Online phase.** After the ANN is trained, it is used online to infer the optimal resource allocation corresponding to any realization of the system channels. This phase is repeated many times for each offline phase, namely, for any coherence block until the ANN must be trained again. In each coherence block, the current channel realizations are the input of the ANN, and the corresponding powers are computed by performing a forward propagation of the trained ANN. This requires $\sum_{\ell=1}^{L+1} N_{\ell-1}N_\ell$ real multiplications* and evaluating $\sum_{\ell=1}^{L+1} N_\ell$ activation functions, with N_ℓ denoting the number of neurons in Layer ℓ and L the number of hidden layers of the ANN.

It should be stressed how this approach is not fully data driven but rather represents one example of cross-fertilization between data-driven and model-based approaches. Indeed, in the case at hand a model of the problem to solve is given by Problem (3.4) and

*The complexity related to additions is negligible compared to that related to multiplications.

allowed us to simply generate a training set without any need of field measurements. The difficulty is that Problem (3.4) would be too hard to be solved with a complexity that is compatible with an online implementation, and here is where deep learning has proved useful.

Numerical performance analysis. This section provides numerical results to show the performance of the described method. Consider the uplink of a wireless interference network with $K = 4$ single-antenna UEs placed in a square area with edge 2 km and communicating with four access points placed at coordinates (0.5, 0.5), (0.5, 1.5), (1.5, 0.5), and (1.5, 1.5) km and equipped with $n_R = 2$ antennas each. The path-loss is modeled following [59], with carrier frequency 1.8 GHz and power decay factor equal to 4.5, while fast fading terms are modeled as realizations of zero-mean, unit-variance circularly symmetric complex Gaussian random variables. In addition, $P_{c,k} = 1\,\text{W}$ and $\mu_k = 4$ for all $k = 1, \ldots, K$, respectively, while the noise power at each receiver is $\sigma^2 = F\mathcal{N}_0 B$, with $F = 3\,\text{dB}$ the receiver noise figure, $B = 180\,\text{kHz}$ the communication bandwidth, and $\mathcal{N}_0 = -174\,\text{dBm/Hz}$ the noise spectral density. The maximum transmit powers are the same for all users, i.e., $P_{\max,1} = \cdots = P_{\max,K} = P_{\max}$, while $P_{\min,k} = 0$ for all $k = 1, \ldots, K$.

The ANN-based solution of Problem (3.4) is implemented by employing a feed-forward ANN with $L + 1$ fully connected layers, in which the $L = 5$ hidden layers have 128, 64, 32, 16, 8 neurons, respectively. After generating training set by solving Problem (3.4) for different realizations of the vector d, the realizations of the parameter vectors d and the optimal output powers in the training set have been converted to logarithmic units, which has been observed to reduce numerical problems during the execution of the training algorithm. Moreover, in order to avoid numerical problems due to the computation of the logarithm of transmit power values that are very close to zero, logarithmic values approaching $-\infty$ have been clipped at $-M$ for $M > 0$. In our experiments, $M = 20$ worked well.[†] Thus, the considered normalized training set is

$$\mathcal{S}_T = \{(\log_{10} d_n, \max\{-20, \log_{10} \tilde{p}_n^*\}) \,|\, n = 1, \ldots, N_T\},$$

where all functions are applied element wise to the vectors in the training set.

The activation functions have been set to exponential linear unit (ELU) for the first hidden layer, while the following hidden layers alternate ReLU and ELU activation functions, and the output layer uses a linear activation function. The use of a linear activation in the output layer is motivated by the consideration that it allows the ANN to produce low training error as a result of a proper configuration of the hidden layers, instead of artificially reducing the output error due to the use of cut-off levels in the activation function. In other words, a linear output activation function allows the ANN to learn whether the present configuration of weights and biases is truly leading to a small output error.

[†]Note that although using a logarithmic scale, the transmit powers are not expressed in dBW, since the logarithmic values are not multiplied by 10. Thus $-M = -20$ corresponds to $-200\,\text{dBW}$.

The ANN is implemented in Keras 2.2.4 [60] with TensorFlow 1.12.0 [61] as backend, using Glorot uniform initialization [62], the Adam training algorithm with Nesterov momentum, and the mean-squared error as the loss function. The training is obtained by solving Problem (3.4) for 102,000 independent and identically distributed realizations of UEs' positions and propagation channels, and different values of $P_{\max}$. In each scenario, the UEs are associated with the access point toward which they enjoy the strongest effective channel. A validation and a test set of 10,200 and 510,000 samples, respectively, were also generated following a similar procedure.

Considering training, validation, and test sets, 622,200 data samples were generated, which required solving the NP-hard Problem (3.4) 622,200 times. This has taken 8.4 CPU hours on Intel Haswell nodes with Xeon E5-2680 v3 CPUs running at 2.50 GHz, by employing the improved branch-and-bound method proposed in [26]. Thus, the average time required to optimally solve one instance of the WSEE maximization problem is $\bar{T} = 4.86 \times 10^{-2}$ s. On the other hand, a forward propagation of the considered ANN requires 10,912 real multiplications and 252 activation function evaluations. Thus, 11,164 elementary operations are required for a forward propagation, which, given the clock frequency of 2.50 GHz of the computer used to run our simulations, yields an average time to solve one instance of the WSEE maximization problem of $\bar{T} = 4.47\,\mu$s, i.e., four orders of magnitude smaller than with the improved branch-and-bound method.

This clearly shows how the offline generation of a suitable training set for ANN-based power control is quite affordable. Finally, all performance results reported in the sequel have been obtained by averaging over ten realizations of the network obtained by training the ANN *on the same training set* with different initialization of the underlying random number generator.[‡] The average training and validation losses for the final ANN are shown in Figure 3.2. It can be observed that both the errors quickly decrease and approach a very small value; thus showing that the adopted ANN configuration is able to properly fit the training data, without underfitting or overfitting.

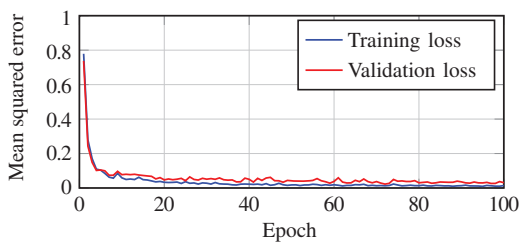


Figure 3.2 Training and validation losses versus training epoch number. It is seen that after the training phase, the ANN neither underfits nor overfits

[‡]Note that this is not equivalent to *model ensembling* [63, Sect. 7.3.3] or *bagging* [9, Sect. 7.1].

Next, we present the performance of the proposed method over the test set, comparing the proposed ANN-based method with the following benchmarks:

- **SCAos:** A first-order optimal method from [26] that leverages sequential convex approximation (SCA) methods. For each value of $P_{\max}$, the algorithm initializes the transmit power to $p_i = P_{\max}$, for all $k = 1, \ldots, K$.
- **SCA:** Again the first-order optimal method based on sequential convex approximation developed in [26], but with a double-initialization approach. Specifically, at $P_{\max} = -30$ dBW maximum power initialization is used. However, for all the values of $P_{\max} > -30$ dBW, the algorithm is run twice, first with the maximum power initialization, and then initializing the transmit powers with the optimal solution obtained for the previous $P_{\max}$ value. Then, the power allocation achieving the better WSEE value is retained.
- **Max. power:** All UEs transmit at maximum power, i.e., $p_k = P_{\max}$, for all $k = 1, \ldots, K$. This strategy is known to perform well in interference networks for low $P_{\max}$ values.
- **Best only:** Only one UE is allowed to transmit, specifically that with the best effective channel. This approach is motivated for high $P_{\max}$ values, as a naive way of nulling out multi-user interference.

The results in Figure 3.3 show that the ANN-based approach outperforms all other practical approaches. The only benchmark that performs comparably with the ANN-based approach is the SCA algorithm that employs the complex initialization rule requiring to solve the WSEE maximization problem twice and for the complete range of $P_{\max}$ values. Thus, this SCA approach is quite more complex than the ANN-based method, but, despite this, it performs slightly worse. In conclusion, we can argue

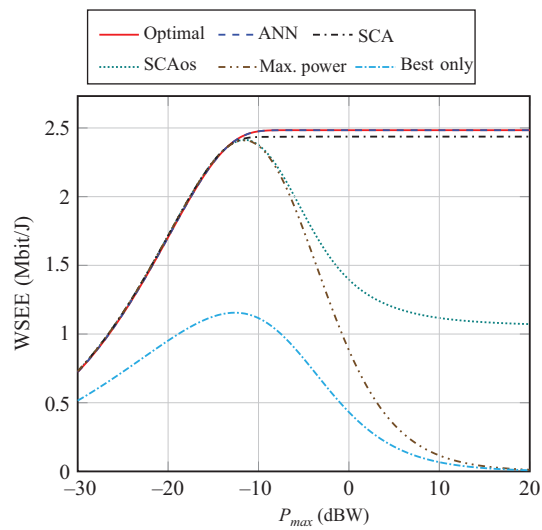


Figure 3.3 WSEE performance of the proposed ANN-based method compared to the global optimum and to several state-of-the-art algorithms

that the ANN approach strikes a much better complexity-performance trade-off than state-of-the-art approaches, and thus it enables online power allocation in wireless communication networks.

3.4.2 Energy efficiency in non-Poisson wireless networks: a deep transfer learning approach

The previous example assumes that a mathematical model of the problem at hand was available to allow us to formulate the optimization problem. The difficulty lied in the cumbersomeness of the model. Instead, this section considers the different scenario in which only approximate models are available, while sufficiently accurate models are lacking. Specifically, consider the problem of energy efficiency optimization with respect to the BS density [64], in non-Poisson cellular networks [65], which is known to be an intractable optimization problem because of the analytical complexity of the utility function to optimize. In this case, the approach used in the previous section does not directly apply. However, it can be extended to this case by merging it with the framework of deep transfer learning.

Transfer learning is a machine learning framework that provides tools to transfer the knowledge acquired when solving a given task, and use it to solve a new, but related task, without starting the learning process from scratch. This general concept can be applied in several ways, but here we follow the approach known as network-based transfer learning, which works by first optimizing the wireless network based on a mismatched, but simpler for optimization, model, and then refining this inaccurate optimization by a fully data-driven approach based on a few empirical samples. This general idea can be applied to our specific problem as the following two-step approach.

First, we assume that the nodes in the wireless network are distributed following a Poisson point process, while the *true* point process model is assumed to be the square grid model [66]. This is a simple example that is chosen in order to shed light on our proposed approach, and that is also easy to simulate and reproduce. Then, leveraging the Poisson assumption, the EE of the network can be optimized in the closed-form, as shown in [64]. This makes it quite simple to generate a large training set of optimal values for the EE as a function of any system parameters, which is then used to train an ANN, following a similar approach as in the previous section. Thus, this first step yields a tentative ANN configuration, that is mismatched due to the fact that the wireless network is not Poisson distributed.

Next, we assume that a second training set is available, which needs to contain only a few data, but based on the actual measurements. In other words, a second, smaller dataset of empirical data is needed to refine the ANN configuration. In particular, this dataset is used to perform a second training phase, in which, however, the weights and biases of the ANN are initialized to the values obtained after the first training phase, instead of using typical random initializations. In other words, the idea behind this approach is to employ the first training phase to obtain an efficient initialization point for the second training phase. Of course, in order for this to be true, it is desirable that the mismatch between the approximate model and the true network model is not too large.

The rest of this section provides more details and numerical examples to show the merits of the described approach.

Model-based optimization. From [64], the EE in Poisson cellular networks can be formulated as follows:

$$\mathrm{EE}(\lambda_{\mathrm{BS}}) = \frac{\mathrm{SE}(\lambda_{\mathrm{BS}})}{P_{\mathrm{grid}}(\lambda_{\mathrm{BS}})}, \tag{3.6}$$

where

$$\mathrm{SE}(\lambda_{\mathrm{BS}}) = B_{\mathrm{W}}\ \log_2(1+\gamma_{\mathrm{D}})\frac{\lambda_{\mathrm{BS}}L(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}{1+\Upsilon L(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})} \times Q\left(\lambda_{\mathrm{BS}}, P_{\mathrm{tx}}, \frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right), \tag{3.7}$$

$$P_{\mathrm{grid}}(\lambda_{\mathrm{BS}}) = \lambda_{\mathrm{BS}}P_{\mathrm{tx}}L\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right) + \lambda_{\mathrm{MT}}P_{\mathrm{circ}} + \lambda_{\mathrm{BS}}P_{\mathrm{idle}}\left(1 - L\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\right), \tag{3.8}$$

are the spectral efficiency and the power consumption of the cellular network, respectively.

Equations (3.7) and (3.8) depend on many parameters, whose definitions can be found in [64]. For the purpose of this discussion, it suffices to observe that λ_{BS} is the deployment density of the BSs, P_{tx} is the transmit power of the BSs, P_{circ} is the circuit power consumption of the BSs, and P_{idle} is the idle power consumption of the BSs. Throughout this section, P_{circ} and P_{idle} are assumed to be fixed, and they are further analyzed in the next section. The goal of the optimization is to determine the optimal deployment density of the BSs, λ_{BS}, given the values of the transmit power P_{tx}. In [64], it is proved that this optimization problem has a unique solution, which corresponds to the unique root of a non-linear equation. This enables the efficient computation of the optimal BSs density, for any given values of the transmit power, which allows for the simple generation of a large training set containing the optimal pairs $(P_{\mathrm{tx}}, \lambda_{\mathrm{BS}}^{(\mathrm{opt})})$, where $\lambda_{\mathrm{BS}}^{(\mathrm{opt})} = \arg\max_{\lambda_{\mathrm{BS}}}\{\mathrm{EE}(\lambda_{\mathrm{BS}})\}$. Such a training set is then used for the first training phase of an ANN whose input is P_{tx} and whose output is $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$.

Data-driven optimization. In the case in which we cannot rely on any analytical models, the EE values need to be estimated by collecting empirical samples from the cellular network, from which the optimal BS density needs to be inferred. In particular, the spectral efficiency and the power consumption can be estimated, respectively, as follows:

$$\mathrm{PSE}(\bullet) = \frac{1}{\mathrm{AreaNet}} \sum_{\mathrm{Cell}(1)\in\mathrm{Net}} \sum_{N_{\mathrm{MT}}\in\mathrm{Cell}(1)} \frac{B_{\mathrm{W}}}{N_{\mathrm{MT}}} \log_2(1+\gamma_{\mathrm{D}})\mathbf{1}(\mathrm{SIR} \geq \gamma_{\mathrm{D}}, \overline{\mathrm{SNR}} \geq \gamma_{\mathrm{A}}), \tag{3.9}$$

$$P_{\mathrm{grid}}(\bullet) = \frac{1}{\mathrm{AreaNet}} \left(\sum_{\mathrm{Cell}(0)\in\mathrm{Net}} P_{\mathrm{idle}} + \sum_{\mathrm{Cell}(1)\in\mathrm{Net}} \left(P_{\mathrm{tx}} + P_{\mathrm{circ}} \sum_{N_{\mathrm{MT}}\in\mathrm{Cell}(1)} N_{\mathrm{MT}} \right) \right). \tag{3.10}$$

The previous two formulas can be interpreted as follows. Considering the spectral efficiency to fix ideas, each mobile terminal in the cellular network determines, based on the received signal, whether it is in coverage. This is accomplished by measuring the average signal-to-noise ratio during the cell association phase and the signal-to-interference ratio during data transmission (if the first phase was successful). This condition corresponds to the term $\mathbf{1}(\mathrm{SIR} \geq \gamma_{\mathrm{D}}, \overline{\mathrm{SNR}} \geq \gamma_{\mathrm{A}})$, where $\mathbf{1}(\cdot)$ is the indicator function. Each mobile terminal can transmit one bit of information to a network controller to report whether it is in coverage or not. Based on the number of mobile terminals that are in coverage on a given cell (say N_{MT}), the BS of that cell equally allocates the available spectrum (say B_{W}) among them and transmits data with a fixed rate $(B_{\mathrm{W}}/N_{\mathrm{MT}})\log_2(1+\gamma_{\mathrm{D}})$. Moreover, exploiting the information from all the mobile terminals, it is possible to identify the BSs that serve at least one mobile terminal (say Cell (1)) and to compute the number of mobile terminals that lie in each of them for each network realization. The spectral efficiency can then be estimated by summing the rates of all active BSs and by normalizing by the area of the network under analysis. If the optimization variable is the BS density, all possible values of density need to be tested, and the value corresponding to the optimal EE needs to be recorded and used to train an ANN. Based on this simple description, we can readily understand that the amount of empirical data that is necessary to train an ANN resorting only to data-driven optimization would not be negligible, thus causing a significant overhead.

Numerical results. Figures 3.4 and 3.5 illustrate some numerical examples that analyze the performance of the described transfer learning approach. A feed-forward ANN architecture with fully connected layers and ReLU activation functions is considered. Specifically, after trying many different ANN configurations, an ANN with three hidden layers equipped with 8, 8, and 2 neurons was selected, as it was found to yield the best performance-complexity trade-off.

Figure 3.4 shows the training and validation-relative MSE versus the number of training epochs for the following approaches:

- the proposed deep transfer learning technique that employs both model-based and empirical data samples;
- the baseline approach, where only empirical data samples are used.

As for the first approach, the size of the training set is always equal to 30,000 samples, out of which x samples follow the true BS distribution (square grid model), while the remaining $(30{,}000 - x)$ follow the Poisson distribution. As for the second approach, the adopted training set contains only the x empirical samples. Thus, this comparison is fair in terms of number of empirical data samples employed and is aimed at showing that augmenting a small dataset of empirical data with a larger dataset of model-based data can provide substantial performance improvements. For both the approaches, the values $x = 300$, 600, 1,500, 2,100, and 3,000 have been considered, and, for each value of x, it is seen that the proposed deep transfer learning method performs much better than the baseline approach.

A similar consideration emerges in the testing phase, too. Figure 3.5 shows the density of BSs as a function of their transmit power, considering a test set of 8,000 new

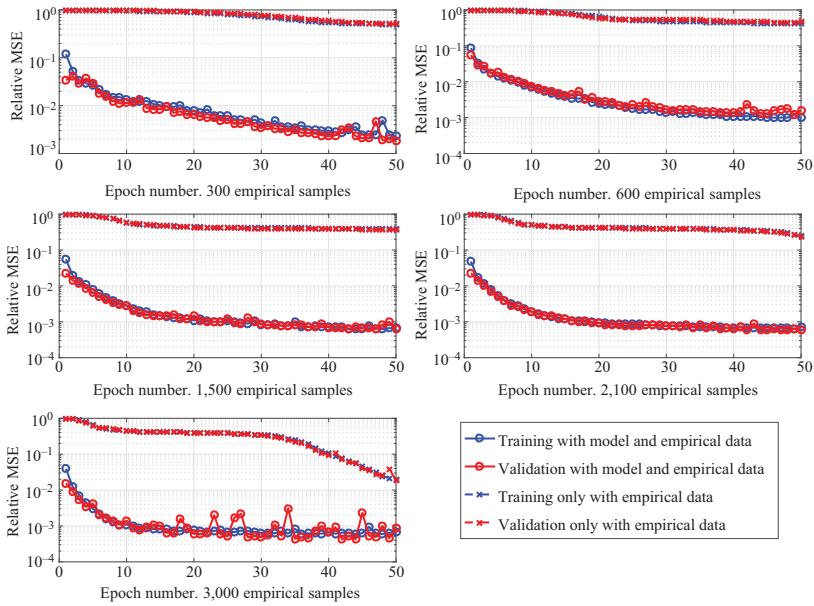


Figure 3.4 *Learning and validation relative MSE versus training epochs for $x = 300$, 600, 1,500, 2,100, and 3,000 samples. For each case, the performance with and without PPP-based samples is reported. It is seen how the use of PPP-based data significantly improves the performance*

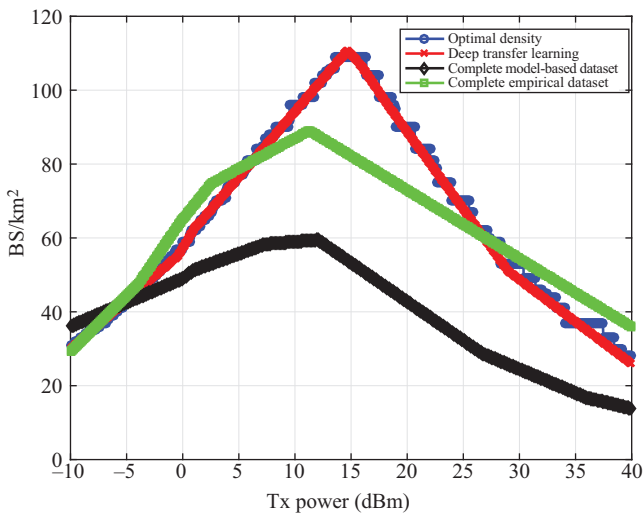


Figure 3.5 *Comparison between model-based, data-driven, and deep-transfer-learning-based optimization—optimal deployment*

transmit powers, which were independently generated from the training and validation sets. The results of the following four schemes are compared:

- the optimal density computed through exhaustive search;
- the density predicted by means of deep transfer learning, where 3,000 empirical samples are used in the second training step;
- the density obtained without transfer learning and performing the training by using only 3,000 empirical samples; and
- the density obtained without transfer learning and performing the training by using only 30,000 model-based samples.

Remarkably, it is seen that using only the 3,000 empirical samples yields quite worse performance compared to the deep transfer learning method that merges the same 3,000 empirical data samples with model-based data. This highlights how performing the model-based pretraining before employing actual measurements for system optimization can overcome the lack of enough empirical samples. Moreover, it is important to note that using only the 30,000 model-based samples does not lead to satisfactory performance, thus showing that it is necessary to employ both model-based and empirical data samples to obtain accurate performance.

3.5 Conclusions

This chapter has been concerned with the description of the use of AI based on deep learning, transfer learning, and ANNs for the design of energy-efficient future wireless networks. Two main conclusions can be drawn at the end of this chapter:

- Developing intelligent wireless networks is becoming more and more necessary as the complexity of communication networks increases. More and more demanding requirements are being enforced, which cannot be met by simply developing faster transmission technologies, but rather rethinking the whole architecture of wireless networks. In this sense, the paradigm of smart radio environments has been put forth, and it has been shown how deep learning will be an essential enabler for the optimization of smart radio environments.
- The complexity of wireless networks is exceeding our ability to derive suitable mathematical models that lend themselves to being optimized with affordable complexity. In this context, complementing the a priori knowledge given by (possibly inaccurate) mathematical models, with deep learning techniques based on ANNs, provides excellent complexity-performance trade-offs that are quite superior to any other available optimization frameworks.

Thus, this chapter demonstrates that while deep learning will be an essential tool for the design of future wireless networks, it should not replace the traditional model-based design paradigm. On the contrary, there is much to be gained by the joint use of data-driven and model-based techniques for wireless networks design.

References

[1] NGMN alliance 5G white paper. *https://www.ngmn.org/5g-white-paper/5g-white-paper.html*, 2015.

[2] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone. A survey of energy-efficient techniques for 5G networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(5):697–709, 2016.

[3] A. Zappone and E. A. Jorswieck. Energy efficiency in wireless networks via fractional programming theory. *Foundations and Trends® in Communications and Information Theory*, 11(3–4):185–396, 2014.

[4] ITU. IMT traffic estimates for the years 2020 to 2030. *Report ITU-R M.2370-0*, 2015.

[5] 5G-PPP. 5G empowering vertical industries. *Euro-5G Project Brochure*, 2016.

[6] A. Zappone, M. Di Renzo, and M. Debbah. Wireless networks design in the era of deep learning: Model-based, ai-based, or both? *IEEE Transactions on Communications*, 67(10):7331–7376, 2019.

[7] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009.

[8] L. Deng and D. Yu. Deep learning methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.

[9] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[10] S. Bi, R. Zhang, Z. Ding, and S. Cui. Wireless communications in the era of big data. *IEEE Communications Magazine*, 53(10):190–199, 2015.

[11] X. Cheng, L. Fang, L. Yang, and S. Cui. Mobile big data: The fuel for data-driven wireless. *IEEE Internet of Things Journal*, 4(5):1489–1516, 2017.

[12] A. Imran, A. Zoha, and A. Abu-Dayya. Challenges in 5G: How to empower SON with big data for enabling 5G. *IEEE Network*, 28(6):27–33, 2014.

[13] S. Kovach. What the big innovation house that powered the mobile boom is betting on next. *http://www.businessinsider.com/qualcomm-ceo-steve-mollenkopf-interview-2017-7*, 2017.

[14] R. Yu. Huawei reveals the future of mobile AI at IFA 2017. *http://www.businesswire.com/news/home/20170902005020/en/Huawei-Reveals-Future-Mobile-AI-IFA-2017*, 2017.

[15] International Telecommunication Union. ITU-T Y.3172 architectural framework for machine learning in future networks including IMT-2020. *ITU-T SG13 Plenary, https://www.itu.int/md/T17-SG13-190304-TD-PLEN/en*, 2019.

[16] *https://www.comsoc.org/ctn/what-will-6g-be*.

[17] P. Hu, P. Zhang, M. Rostami, and D. Ganesan. An integrated active-passive radio for mobile devices with asymmetric energy budgets. In *ACM SIGCOMM*, 2016.

[18] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz. Realizing wireless communication through software-defined hypersurface environments. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2018.

[19] 5GPPP vision on software networks and 5G SN WG. 2017.

[20] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz. A new wireless communication paradigm through software-controlled metasurfaces. *IEEE Communications Magazine*, 56(9):162–169, 2018.

[21] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, *et al.* Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come. *EURASIP Journal on Wireless Communications and Networking, https://arxiv.org/abs/1903.08925*, 2019.

[22] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[23] N. Wiener. *Cybernetics, or Control and Communication in the Animal and the Machine*. MIT Press, 1948.

[24] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz. Using any surface to realize a new paradigm for wireless communications. *Communications of the ACM*, 61(11):30–33, 2018.

[25] Z.-Q. Luo and S. Zhang. Dynamic spectrum management: Complexity and duality. 2(1):57–73, 2008.

[26] B. Matthiesen, A. Zappone, K. L. Besser, E. A. Jorswieck, M. Debbah. A globally optimal energy-efficient power control framework and its efficient implementation in wireless interference networks. *IEEE Transactions on Signal Processing*, 68:3887–3902, 2020.

[27] Q. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini. Asymptotic analysis of large intelligent surface assisted MIMO communication. *https://arxiv.org/pdf/1903.08127.pdf*, 2019.

[28] T. Nakanishi, T. Otani, Y. Tamayama, and M. Kitano. Storage of electromagnetic waves in a metamaterial that mimics electromagnetically induced transparency. *Physical Review B*, 87(161110), 2013.

[29] Y. S. Nasir and D. Guo. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE Journal on Selected Areas in Communications*, 37(10):2239–2250, 2019.

[30] L. Sanguinetti, A. Zappone, and M. Debbah. A deep-learning framework for energy-efficient resource allocation in massive MIMO systems. In *Asilomar Conference on Signals, Systems, and Computers*, 2018.

[31] M. K. Sharma, A. Zappone, M. Assaad, M. Debbah, and S. Vassilaras. Distributed power control for large energy harvesting networks: A multi-agent deep reinforcement learning approach. *IEEE Transactions on Cognitive Communications and Networking*, 5(4):1140–1154, 2019.

[32] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, 2011.

[33] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alu, and N. Engheta. Performing mathematical operations with metamaterials. *Science*, 343(6167):160–163, 2014.

[34] L. Spada. Metamaterials for advanced sensing platforms. *Research Journal on Optical Photonics*, 1(1), 2017.

[35] L. Subrt and P. Pechac. Controlling propagation environments using intelligent walls. In *European Conference on Antennas and Propagation*, 2012.

[36] A. Tsioliaridou, C. Liaskos, and S. Ioannidis. Towards a circular economy via intelligent metamaterials. In *IEEE International Conference on Computer-Aided Modeling Analysis and Design of Communication Links and Networks*, 2018.

[37] N. Yu, P. Genevet, M. A. Kats, *et al.* Light propagation with phase discontinuities: Generalized laws of reflection and refraction. *Science*, 334(6504): 333–337, 2011.

[38] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith. An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials. *IEEE Antennas and Propagation Magazine*, 54(2):10–35, 2012.

[39] F. D. Calabrese, L. Wang, E. Ghadimi, G. Peters, and P. Soldati. Learning radio resource management in 5G networks: Framework, opportunities and challenges. *https://arxiv.org/abs/1611.10253*, 2017.

[40] M. Chen, W. Saad, C. Yin, and M. Debbah. Echo state networks for proactive caching in cloud-based radio access networks with mobile users. *IEEE Transactions on Wireless Communications*, 16(6):3520–3535, 2017.

[41] P. De Kerret and D. Gesbert. Robust decentralized joint precoding using team deep neural network. In *IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2018.

[42] J. Fang, X. Li, W. Cheng, Z. Chen, and H. Li. Intelligent power control for spectrum sharing: A deep reinforcement learning approach. *https://arxiv.org/pdf/1712.07365.pdf*, 2018.

[43] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang. Intelligent 5G: When cellular networks meet artificial intelligence. *IEEE Wireless Communications*, 24(5):175–183, 2017.

[44] F. Liang, C. Shen, W. Yu, and F. Wu. Towards optimal power control via ensembling deep neural networks. *https://arxiv.org/abs/1807.10025*, 2018.

[45] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos. Learning to optimize: Training deep neural networks for interference management. *IEEE Transactions on Signal Processing*, 66(20):5438–5453, 2018.

[46] A. Zappone, M. Debbah, and Z. Alltman. Online energy-efficient power control in wireless networks by deep neural networks. In *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications*, 2018.

[47] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian. Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless systems optimization. *IEEE Vehicular Technology Magazine*, 14(3):60–69, 2019.

[48] A. Zappone, L. Sanguinetti, and M. Debbah. User association and load balancing for massive MIMO through deep learning. In *Asilomar Conference on Signals, Systems, and Computers*, 2018.

[49] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen. Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Transactions on Wireless Communications*, 18(8):4157–4170, 2019.

[50] Q. Wu and R. Zhang. Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design. *IEEE Transactions on Wireless Communications*, 18(11):5394–5409, 2019.

[51] Y. Yang, S. Zhang, and R. Zhang. IRS-enhanced OFDM: Power allocation and passive array optimization. *https://arxiv.org/pdf/1905.00604.pdf*, 2019.

[52] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson. Weighted sum-rate optimization for intelligent reflecting surface enhanced wireless networks. *https://arxiv.org/pdf/1905.07920.pdf*, 2019.

[53] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo. Intelligent reflecting surface: A programmable wireless environment for physical layer security. *IEEE Access*, 7:82599–82612, 2019.

[54] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi. Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects. *https://arxiv.org/pdf/1810.05667.pdf*, 2019.

[55] E. Basar. Large intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G. *https://arxiv.org/pdf/1904.06704.pdf*, 2019.

[56] W. Yan, X. Kuai, and X. Yuan. Passive beamforming and information transfer via large intelligent surface. *https://arxiv.org/pdf/1905.01491.pdf*, 2019.

[57] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6:861–867, 1993.

[58] R. W. Freund and F. Jarre. Solving the sum-of-ratios problem by an interior-point method. *Journal of Global Optimization*, 19(1):83–102, 2001.

[59] G. Calcev, D. Chizhik, B. Goransson, *et al.* A wideband spatial channel model for system-wide simulations. *IEEE Transactions on Vehicular Technology*, 56(2), 2007.

[60] F. Chollet Keras. *https://keras.io*, 2015.

[61] M. Abadi TensorFlow: Large-scale machine learning on heterogeneous systems. *https://tensorflow.org*, 2015.

[62] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS'2010*, 2010.

[63] F. Chollet. *Deep Learning With Python*. Manning, 2017.

[64] M. Di Renzo, A. Zappone, T. T. Lam, and M. Debbah. System-level modeling and optimization of the energy efficiency in cellular networks—A stochastic geometry framework. *IEEE Transactions on Wireless Communications*, 17(4):2539–2556, 2018.

[65] M. Di Renzo, S. Wang, and X. Xi. Modeling and analysis of cellular networks by using inhomogeneous Poisson point processes. *IEEE Transactions on Wireless Communications*, 17(8):5162–5182, 2018.

[66] J. G. Andrews, F. Baccelli, and R. K. Ganti. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11):3122–3134, 2011.

Chapter 4

Scheduling resources in 5G networks for energy efficiency

Cristian Rusu[1] and John Thompson[2]

The deployment of a large number of small cells is currently regarded as an important approach in the pursuit of addressing the traffic demands and sensing capabilities of the next-generation 5G networks. In this chapter, we propose an adaptive scheduling algorithm that activates/deactivates network resources to establish a trade-off between the estimation accuracy and the energy consumption of the network. The method is based on the iterative reweighted convex optimization techniques. We provide experimental results to show how the proposed method can establish, in polynomial time, a trade-off between the quality of the network estimation accuracy and its energy consumption. Furthermore, the proposed method can adapt to changes in the network's topology, say due to unexpected hardware and communication failure or when network resources completely run out of energy and power down.

4.1 Introduction

As 4G networks are reaching their mature state, research has long moved toward the design and deployment of the fifth-generation cellular network technology (5G networks) [1] that comes with big promises: high speeds (large bandwidths that lead to $\times 1{,}000$ capacity over current networks) and low latency of communications (targeting 1–4 ms to the base station).

Several factors have to lead to the development of 5G networks: (i) rapidly increasing mobile data traffic, mostly due to the significantly increased demand of high-quality video streaming; (ii) an explosion in the number of devices connected to the networks, particularly due to the growth in Internet of Things (IoT) devices; (iii) the demands of customers for better services while keeping flat-rate tariffs; and (iv) new business opportunities stemming from innovative services that exploit the high bandwidth and low-latency features of the networks. Currently, industrial companies in many nations have taken several concrete steps toward the deployment

[1]LCSL, Istituto Italiano di Tecnologia, Genova, Italy
[2]Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, UK

of a full 5G network: spectrum auctioning* and allocation,† the development of 5G devices and telecommunications infrastructures,‡,§ and carriers making available 5G subscriptions.‖

Due to a large number of mobile devices and the increased data traffic, one potential downside of 5G networks is the high energy consumption [2]. Scientists have already raised serious concerns regarding the high levels of carbon dioxide in the atmosphere¶ and have highlighted concerns about several information and communications technologies, among them are modern (deep) machine learning [3] and millimeter-wave (mmWave)/5G communications [4]. Researchers and companies are making significant efforts to improve several aspects of 5G networks to tame the energy consumption of the overall communication system. This effort bears the name of *green communications*. While, for environmental reasons, energy consumption reduction is a goal in itself, there are several other factors to consider, for example, high electricity bills might turn customers off reducing the network utilization; extending the life of battery-operated devices increases their usability and reducing the amount of radiation absorbed by the human body, as measured by the specific absorption ratio, will have significant benefits on the health of the users.

Research that breaks down the energy needs of next-generation mobile networks has been underway for quite some time now [4]. While mobile device manufacturing and operation of the data centers that collect and manage the data collected from the network make up for approximately 50% of the energy consumption, the rest is almost entirely taken by network operational costs: mobile device operation, radio access network, or operator activities.

While it is beyond the scope of this chapter to provide a detailed, exhaustive review of the current state of the art, we detail next some of the main green communications research directions.

To achieve large data transfer bandwidths, massive multiple-input–multiple-output (MIMO) and mmWave [5] systems are the wireless communication technology at the heart of 5G. The large number of antennas deployed on each device in such

*Archive for the "5G Auctions" category. The European 5G Observatory [accessed 2019 Aug 16]. Available from: https://5gobservatory.eu/category/5g-auctions/.

†Spectrum for 5G networks: licensing developments worldwide. Global mobile Suppliers Association (GSA) [accessed 2019 Mar 26]. Available from: https://gsacom.com/paper/5g-spectrum-licensing-mar-2029/.

‡Me F., San Marino primo Stato 5G d'Europa, accesa l'antenna. Corriere Comunicazioni (in Italian) [accessed 2019 Jul 11]. Available from: https://www.corrierecomunicazioni.it/digital-economy/san-marino-primo-stato-5g-accessa-la-prima-antenna.

§Thune J., Streamlining the rapid evolution and modernization of leading-edge infrastructure necessary to enhance small cell deployment act (S. 3157) also known as the Streamline Small Cell Deployment Act. US Congress; [accessed 2019 Jun 30]. Available from: https://www.congress.gov/bill/116th-congress/senate-bill/1699.

‖Joon-ho H., From the first day of 5G, 40,000 subscribers ... The Asia Business Daily [accessed 2019 Apr 6]. Available from: http://view.asiae.co.kr/news/view.htm?idxno=2019040610062165080.

¶The Climate Group on behalf of the Global e-Sustainability Initiative (GeSI). SMART 2020: Enabling the Low Carbon Economy in the Information Age. Global e-Sustainability Initiative (GeSI) [accessed 2019 Jun 15]. Available from: http://www.gesi.org.

systems and the high carrier frequencies has spawned several ideas at the radio frequency chain level to reduce the energy consumption, among these: perform less baseband processing [6], use simple hardware such as switches [7] or use low-resolution ADCs and DACs [8].

Regarding the spectrum itself, dynamic allocation is necessary to ensure continuous coverage at all times. To this end, spectrum sharing technologies [9] have been developed, which allow secondary users to access underutilized (or unutilized) parts of the spectrum that have been originally allocated to different users/tasks. This technology is one of the few that simultaneously improves both energy and spectrum efficiency—others show a trade-off between communication efficiency and energy consumption.

The number of devices connected to networks is expected to increase drastically especially because of the development of IoT [10]. The infrastructure that supports this technology consists mainly of wireless sensor networks (WSNs) [11] where optimal power control is essential to making the network competitive from an energy consumption perspective. In this context, the management of sensors and sensor networks is essential to the IoT, and its energy management can lead to significant gains in energy efficiency (EE) [12].

One way to deal with a large number of devices connected to mobile 5G networks is to allow device-to-device communications [13], i.e., direct link formation is allowed between two mobile users without the involvement of the base stations or any other core network elements. When devices are in the range of each other, direct communications are allowed leading to high data rates that also simultaneously improve EE and latency [14]. Such situations do require the network to provide mandatory power management and interference mitigation among the devices [15] .

To deal with the large concurrent number of connections in 5G networks, ultra-dense networks (UDNs) [16] have also been developed. It has been shown that a large number of picocells and femtocells (i.e., small and very small, respectively) lead to improved spectral and EE. These cells have low-cost access points (APs) and exhibit reduced transmission power: 20 dBm for femtocells as compared to 46 dBm for a macrocell [17].

As the research community has identified the need for efficient 5G communications early, several excellent works provide an overall literature review of the main ideas and research efforts on this topic, among these the reader is encouraged to check [2,18,19]. Researchers have identified several general areas of research and proposed optimization (reduction) of the power consumption: (i) make use of renewable energy using energy harvesting techniques; (ii) optimized network planning that will check for optimal base station distribution and activation; and (iii) optimized allocation of resources such as time and frequency which lead to natural trade-offs between the EE and other metrics that relate to the communications or sensing performance of the network.

A particular interesting energy reduction technique in 5G system is Wi-Fi offloading [20], i.e., the ability to transfer some of the mobile traffic from a base station to nearby Wi-Fi APs or other active base stations. When used in conjunction with base station activation/deactivation, the energy benefits can be considerable, as entire

radio resources can be moved offline [21]. In this case, management tools that decide activation/deactivation policies need to be deployed.

Scheduling a set of mobile network resources, in time, has been addressed in various scenarios. In the context of 5G networks, scheduling usually refers to the allocation of users according to some quality or energy criterion: allocate user data packets in the frequency domain for each time interval to increase the quality of service (QoS) given partially unknown network conditions and highly dynamic traffic scenarios. This makes resource allocation in the 5G networks extremely complex (space, time, and frequency allocation have to be considered).

The work in [22] proposed a scheduling algorithm based on ideas from reinforcement learning to minimize the packet delay and drop rates for given QoS requirements. Then, scheduling in [23] is performed by prioritizing public safety data and users who are mandatory to intervene during special alerts in active disaster areas. Several other constrained scenarios are also considered: a throughput constrained mid-haul element in the network [24], an overall cross-layer optimization for scheduling and resource allocation (SRA) at both the physical (PHY) and data link control layers to investigate and compare the performance between orthogonal frequency-division multiplexing and quadrature amplitude modulation-based filter bank multicarrier (OQAM/FBMC) [25], a dynamic programming algorithm, which supports fairness, for effective cross-layer downlink SRA considering the channel and queue state [26], dynamic cell bandwidth allocation is proposed to enhance quality of experience (QoE) using inter-cell scheduling of users [27], a user scheduling method that exploits the properties of massive MIMO 5G systems of the BS [28], and finally scheduling under severe time constraints [29]. An overview of the agile multiuser scheduling functionality in 5G is given in [30], while different scheduling algorithms that consider the QoS requirements of each user, as well as the individual channel conditions, are evaluated and compared in [31]. QoS and QoE are related, as QoE measures the overall level of customer satisfaction, while QoS refers to hardware and software characteristics and performance, i.e., coverage, data rates, and delay.

Scheduling of hardware/radio/sensor resources has been studied to a larger extent in the signal processing (for communications) literature. From this line of work, two important examples are hardware resource scheduling in UDNs [32] (especially BS scheduling [33,34]) and sensor measurement scheduling in distributed measurement scenarios [35]. 5G networks will be greatly enhanced by IoT sensor networks. Indeed, many services associated with 5G technologies would not be possible without the massive deployment of (i) localization sensors; (ii) environment monitorization sensors; and (iii) data flow sensors at the base stations and other important network nodes. The GSMA foundation estimated** that a regular 5G cell will be able to cope with the data flow from more than 250,000 sensors without affecting the regular traffic performance of the network. This amount of sensors needs to be managed not only to ensure correct functionality but also to optimize for energy consumption and improve

**The Australian, 5G sensor revolution is coming [accessed 2019 Sep 20]. Available from: https://www.theaustralian.com.au/life/5g-sensor-revolution-is-coming/news-story/20b1991b0220b72c61fda4fc55269bff.

robustness to hardware failures that will inevitably happen when so many resources are deployed in the field.

The sensor placement/scheduling/management problems have been extensively studied in the past. A general approach is to use greedy methods based on a minimum eigenspace approach [36] or with submodularity-based performance guarantees [37] that provide results within $(1 - e^{-1})$ of the optimal solution. Another popular greedy sensor selection method, FrameSense [38], initially activates all the sensors and then removes one at each step based on a "worst-out" principle to optimize its submodular objective function. Convex optimization techniques have also been proven useful for experimental design [39, Chapter 7.5] with ℓ_1 norm minimization [40]. Additional scenarios where further limitations are added to the network sensing problem include energy budget constraints [41] and ways to maximize the lifetime per unit cost in WSNs [42], regularization terms that discourage the selection of the same sensors over some time [43,44], and scheduling [45] over the network. In the same spirit, recent work introduces a greedy sampling set selection algorithm for graph signal processing (GSP) applications [46].

In line with the previous work on optimized network planning, in this chapter we provide a high-level view and solution to the energy management problem for 5G networks. We are concerned with the management and scheduling of 5G network resources that exhibit some level of redundancy such that a trade-off between quality of communications and EE can be established. We outline a simple example of the network scenario we consider in Figure 4.1. We start the chapter by describing some of the most important and widely used EE metrics. Then, based on ideas from convex optimization, we describe the basic mathematical concepts and then propose an algorithm that deals with scheduling network resources in time such that a measure of network efficiency is optimized under some operational constraints.

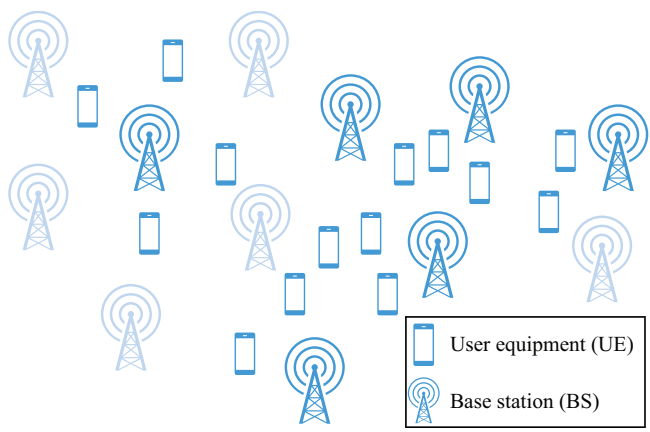


Figure 4.1 An outline of the proposed 5G network scenario. A UDN with multiple base stations that serve multiple users but note that only a subset of the base stations are active (those in solid blue) as they are able to successfully cover all currently active users in the network

We describe in detail the proposed method and then proceed to various experimental results under various assumptions that we can accommodate due to the convex optimization framework we are using.

The work presented here is based on the previously proposed algorithms for sensor scheduling published in [47–49].

4.2 Preliminaries

We start by describing basic metrics for green performance evaluation and a primer on convex optimization. Then with these prerequisites, in this section, we describe the proposed method for resource scheduling in 5G networks.

4.2.1 Energy efficiency metrics and objectives

In any wireless communications system, we can define clear performance metrics: number of users served, throughput, downtime, etc. In general, the most used performance metric is the spectral efficiency, a measure of the maximum data rates achievable without error (the channel capacity). When it comes to metrics that measure EE, we again find several that are appropriate in different scenarios. A fairly detailed enumeration is given in [50], among the most used, we have the following:

1. Energy efficiency (EE): measured in bits/s/W, it is defined as the total capacity over the total power consumed:

$$\text{EE} = \frac{\text{Data rate (bits/s)}}{\text{Power consumption (W)}}; \tag{4.1}$$

2. Area energy efficiency (AEE): is an EE measure applied in the case of cellular heterogeneous network cells that take into account the area of the cell under consideration:

$$\text{AEE} = \frac{\text{EE}}{\text{Cell area (km}^2\text{)}}. \tag{4.2}$$

While these two indicators provide an overall performance measure, the next two refer to per user (assuming users connected in the downlink) performance.

3. Outage probability (OP): is defined as the probability that a user, say the *i*th in the downlink, will have a signal-to-noise ratio (SNR) below a prescribed, performance guaranteeing target value, $\text{SNR}_{\text{target}}$:

$$\text{OP} = \mathbb{P}(\text{SNR}_i \leq \text{SNR}_{\text{target}}); \tag{4.3}$$

4. Energy harvest ratio (EHR): is a dimensionless unit that measures the fraction of time, out of a total standard time slot/unit T_{total}, during which energy harvesting is done T_{EH}:

$$\text{EHR} = \frac{T_{\text{EH}}}{T_{\text{total}}}; \tag{4.4}$$

In this setting, an important drawback is to note that many simple devices are not able to perform data decoding and harvest information simultaneously [51].

5. Average sum rate (ASR): measured in bits/s, is the sum of all the data rates (R_i for the ith user) of all the n users connected to the network:

$$\text{ASR} = \sum_{i=1}^{n} R_i. \tag{4.5}$$

While this metric measures the data rate directly, it is connected to the power consumption as the achieved rates depend on the power allocation strategy in the cellular network. We would like to contrast here the simple equal power allocation (EPA) strategy against the water filling algorithm (WFA) which performs power allocation to maximize capacity in inter-symbol interference and frequency-selective fading channels. Other power allocation strategies, taking into account other performance indicators, can also be applied here.

These indicators can be used separately or in some combination to assess the performance level of the network. As EE is usually inversely proportional to achieved data rate (due to the simple fact that higher SNR can always be achieved by increasing the consumed power), most of the time we will observe a trade-off between the QoS provided by the network and its energy consumption. Furthermore, the use of renewable energy sources and battery power backups may put more constraints on base station operation and lead to the necessity of optimizing the active/sleep cycle of base stations [52].

As the network designers, we can define some desirable characteristics and goals for its operation:

1. Maximize network life: as the network contains several devices that are battery powered or that are known to have different wear and tear properties, one goal might be to operate the network in such a way that no (or few) devices break down completely over a fixed period of time;
2. Network sensing/coverage and connectivity: as the network needs to serve connecting users and sense environment parameters, an important goal is to ensure that a predefined geographical area has coverage and that users trying to connect to the network from this area can be serviced promptly;
3. Robustness: coverage and user connectivity have to be ensured even in the case of failure of some of the devices in the network.

While we consider that the network is fixed and static, i.e., no major restructuring takes place in a certain time period, we assume some level of redundancy in the network (several devices are capable of providing the same services to the same users in the same geographical area) and that the resources of the network are controllable and can be switched to active and deactivated modes. The redundancy assumption is essential. Without some level of redundancy in the network, all resources need to activate all the time and no planning or scheduling is possible. This assumption is realistic as traffic demand and network resources are usually concentrated in a relatively small portion of the network (20% of sites account for 50% of traffic [53]).

Luckily, UDNs exhibit a very high density of radio resources in a given area (dense, small cells) to serve approximately the same number of users as a classic network but where each user has a significant increase in traffic. This means that either the number of base stations is on the same order of magnitude with the number of users, or the distance between any user and a base station is only a few meters. Moreover, this improved (higher) density is the property that will allow us to schedule resources in time. Furthermore, the high density of base stations has one significant drawback: there is strong interference between all the base stations, which degrades performance. As such, mechanisms to set some of the base stations in sleep mode are essential for energy savings.

In this chapter, we focus on the three main characteristics listed previously. We will design an algorithm that can schedule (turn on/off or set to active/sleep mode) devices that are in the network such that (i) we ensure that the life of each device is extended as much as possible (e.g., to a particular, predefined number of time slots) and (ii) the network is robust to hardware failures and it is able to reorganize (reschedule) its structure such that the goal of point (i) is still ensured to the best current capabilities of the network.

To incorporate several real-world constraints in the design of our scheduling algorithm, we base the development on the convex optimization platform. We will define the objective function we consider and the constraints imposed upon the network. This will allow us to make design adjustments with little effort to the scheduling algorithm itself.

4.2.2 A primer on convex optimization

Convex optimization [39] is a subbranch of mathematical optimization concerned with the study of optimization problems with convex objective functions and convex constraints. Such a problem is expressed as

$$\begin{aligned} &\underset{\mathbf{z}}{\text{minimize}} \quad f(\mathbf{z}) \\ &\text{subject to} \quad g(\mathbf{z}) = \mathbf{0}_{p\times 1} \text{ and } h(\mathbf{z}) \leq \mathbf{0}_{r\times 1}, \end{aligned} \tag{4.6}$$

where $f(\mathbf{z})$ is convex and is called the objective function, $g(\mathbf{z})$ encodes the p affine constraints, $h(\mathbf{z})$ encodes the r convex constraints, and $\mathbf{z} \in \mathbb{R}^m$ is the optimization variable. If the constraint set admits at least one point, i.e., it is not empty, then the problem is called feasible. The convexity assumptions guarantee that every local minima of the problem is actually a global minimum and therefore the problem is solved exactly. Furthermore, if there are multiple solutions their set is itself convex and if the problem is strictly convex, then the problem has at most a single optimal point. There are several other reasons that account for the popularity of convex optimization methods:

1. Many real-world problems can be formulated as convex optimization problems (linear programming, least squares, semidefinite programming, geometric programming, etc.);
2. Many more real-world problems can be approximated by (or relaxed to) convex optimization problems;

3. It is trivial to model additional real-world constraints by augmenting the optimization problem with further constraints (as long as they are convex/affine);
4. There are several specialized ready-to-use software packages (including, CVX[††] and CVXPY[‡‡]) for different programming languages that solve convex optimization problems in polynomial time (mostly by interior-point methods [54]).

In this chapter, our goal is to define convex optimization problems that can be solved efficiently. We will use extensively optimization with ℓ_p norms, which are convex. We are interested mainly in the ℓ_1 and ℓ_∞ norms, defined as

$$\|\mathbf{z}\|_1 = \sum_{i=1}^{n} |z_i| \quad \text{and} \quad \|\mathbf{z}\|_\infty = \max_i |z_i|, \tag{4.7}$$

where the first is known to promote sparsity [55] in the solution and the latter one promotes equal entries (in absolute values) in the solution [56]. These two penalties are useful in our case, and we will use them jointly: the ℓ_1 will serve as a selection mechanism to model on/off states, while the ℓ_∞ will ensure that no network resource is overused (by minimizing selection of the maximally used resources) and we can strike a balance between the performance of the work and its energy consumption.

General scheduling algorithms, which give solutions to time-dependent job allocation tasks, have been extensively studied in the past. Unfortunately, these problems are in general non-convex and in most situations are NP-hard to solve to optimality (these are nonlinear mixed-integer optimization problems [57] but where the objective functions are convex or concave). This is because activating and deactivating resources are modeled most of the time as a binary decision variable, i.e., $z_i \in \{0, 1\}$, which is discontinuous and therefore non-convex (to ensure optimality, a combinatorially large number of solutions should be checked exhaustively). Because of this, several approaches (or heuristics) have been proposed in the literature to reach approximate solutions to the scheduling plans: greedy methods (i.e., scheduling resources one-at-a-time), and search algorithms (including, genetic algorithms, simulated annealing, particle swarm optimization, tabu search and, inspired by more recent research in reinforcement learning, dynamical programming, and approximations of it). In this chapter, our approach will be to relax the binary constraints to convex interval constraints (i.e., $z_i \in [0, 1]$) and then use an iterative process to push the entries to either limiting value.

4.2.3 Sensors and their measurements

With the basic mathematical formulation described, it is also important to clarify what we mean when we refer to "a sensor" in the context of this work. Broadly speaking, we consider two scenarios.

[††]M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx, 2013 [accessed 20 September 2019].
[‡‡]S. Diamond and S. Boyd. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. https://www.cvxpy.org, 2016 [accessed 20 September 2019].

First, in the spirit of the work from [58], the sensors represent base stations and the measurements they take are of signals from mobile devices in the network. Our objective is to minimize the number of activated base stations in the network. The measurements thus represent the propagation channel from the mobile devices to all the base stations in the network. Second, the sensors have a more abstract function and are measuring properties of the environment and network [59], e.g., data flows through the network, energy consumption in the overall network, and environmental parameters. In this situation, the measurements represent the connections between the quantities of interest and where they can be measured.

In this chapter, we model sensors as devices capable of linear measurements, communication with some neighboring sensors, and possessing a known energy profile (or characteristics).

4.3 The proposed scheduling algorithm

In this section, we introduce the different system modeling choices that we assume throughout the chapter and then outline the proposed solution for adaptive resource scheduling over time. We build up the formulation starting from resource selection in a single time instance, extending to scheduling resources over multiple time instances and finally describing adaptive scheduling while allowing network topology changes (including due to unexpected hardware failure). We conclude the section by discussing the computational properties of the proposed method.

4.3.1 The mathematical model: the measurements

We now introduce the mathematical notation and set up the objective function that models the sensor's accuracy and that we consider in this chapter.

We assume that the network can perform a total number of m different measurements and is used to estimate a parameter vector $\mathbf{x} \in \mathbb{R}^n$ that can change over T time instances such that at time t we use k_t linear measurements (such that $n \leq k_t \leq m$) as

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{n}_t, \text{ where } \mathbf{A}_t \in \mathbb{R}^{k_t \times n} \text{ and } t = 1, \ldots, T. \tag{4.8}$$

We have denoted by $\mathbf{n}_t \in \mathbb{R}^{k_t \times 1}$ zero-mean i.i.d. Gaussian noise with variance $\sigma^2 \mathbf{I}$, $\mathbf{y}_t \in \mathbb{R}^{k_t \times 1}$ is the measurement vector at time t, and $\mathbf{x}_t$ is the realization of $\mathbf{x}$. In the simplest situation, $\mathbf{x}_t$ can be constant over all time instances T. We do assume that the whole sensor network can be modeled as a large linear measurement operator $\mathbf{A} \in \mathbb{R}^{m \times n}$, where each row $\mathbf{a}_i^T$ models the measurement a single sensor can perform. The matrix $\mathbf{A}_t$ models the measurements that a subset of the sensor network performs at time t. We also assume that we always take $k_t \geq n$ measurements such that estimation of $\mathbf{x}$ leads to a well-posed inverse problem. With these measurements, we consider the classic least squares estimator $\hat{\mathbf{x}}_t$, given by applying the pseudo-inverse of the sensing matrix to the measurements as

$$\hat{\mathbf{x}}_t = \mathbf{A}_t^{\dagger} \mathbf{y}_t = (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T \mathbf{y}_t. \tag{4.9}$$

The quality of the estimation at time t can be measured in several ways. In this chapter, we choose the mean-squared error (MSE) (or A-optimality quantifier):

$$\mathrm{MSE}(\mathbf{A}_t) = \mathrm{tr}((\mathbf{A}_t^T\mathbf{A}_t)^{-1}) = \sum_{i=1}^{n} \frac{1}{\lambda_i(\mathbf{A}_t^T\mathbf{A}_t)}. \tag{4.10}$$

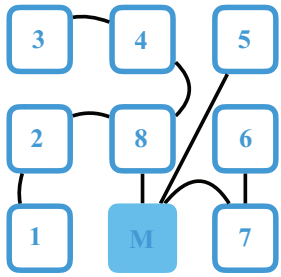
Here we have denoted with λ_i the ith eigenvalue of $\mathbf{A}_t^T\mathbf{A}_t$ that is always real valued, and the trace that is the sum of the eigenvalues. Without loss of generality, we do assume the ordering $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. There are two other classic performance indicators, the worst case error variance (E-optimality) and the volume of the confidence ellipsoid (D-optimality), which for clarity of exposition we choose not to describe in detail here. We just note that the framework we propose easily extends to these two indicators as well. From an algorithmic perspective, all these indicators are very interesting as A and E optimality are convex (and therefore can be minimized), while D-optimality is concave (and therefore can be maximized).

In general, with the assumption of white Gaussian noise, to achieve low error indicators, we focus on the eigenvalues of the symmetric positive semidefinite $\mathbf{A}_t^T\mathbf{A}_t \in \mathbb{R}^{n\times n}$ that should obey two properties: (i) the smallest eigenvalue $\lambda_n(\mathbf{A}_t^T\mathbf{A}_t)$ should be as high as possible and (ii) all the eigenvalues should be approximately equal $\lambda_i(\mathbf{A}_t^T\mathbf{A}_t) \approx \lambda_j(\mathbf{A}_t^T\mathbf{A}_t),\ \forall\ i \neq j$, i.e., the measurement matrix $\mathbf{A}_t$ behaves approximately as a tight frame with large Frobenius norm SNR. All the three optimality criteria enumerated previously are functions of the eigenvalues of $\mathbf{A}_t^T\mathbf{A}_t$ and capture, in different ways, these two desirable properties.

4.3.2 The mathematical model: the network

An innovative concept in 5G networks is the idea of integrated access and backhaul links in mmWave systems. This leads to scenarios where some base stations are directly connected to the core network, while others need to relay their data via wireless links to other base stations [60]. In this chapter, we assume a similar structure imposed on our sensor network and explore how it affects the management of the resources.

Therefore, the sensors in the network are not isolated nor fully connected, but communication between them is governed by a graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices (the sensors) such that $|\mathcal{V}| = m$ and $\mathcal{E}$ is the set of edges (the allowed connections between the sensors) which we assume is not dense, i.e., $|\mathcal{E}| = O(m)$, which determines possible links and their connection to a centralized (master) node where the estimation (4.9) takes place following the collection of the measurements. For clarity, let us consider a concrete example in Figure 4.2. The closest sensors to the master node are 5, 7, and 8, while the furthest are 1 and 3 which need 3 hops to get to the master node. In our context, the graph structure has two consequences: (i) it shows that even sensors that do not perform measurements will still consume energy in order to forward measurements taken by other sensors to the master node and (ii) it highlights bottlenecks in the network, single nodes that collect measurements from many nodes and whose failure would lead to the loss of a large number of sensors.



$$\mathbf{C} = \begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & * & 0 \\ * & * & * & * & 0 & 0 & 0 & * \end{bmatrix}$$

Figure 4.2 An example of sensor network with only $m = 8$ sensors and a master node. Left: the topology of the undirected network and right: the cost matrix representation that we use for the network on the left

We have denoted by $\mathbf{C} \in \mathbb{R}^{m \times m}$ the sensing and communication costs of all the sensors. The entry $c_{ij} \geq 0$ expresses the cost incurred by the ith sensor to convey data from the jth sensor to the master node. When the ith sensor has a direct link to the master node, the ith row of $\mathbf{C}$ has only one nonzero entry, namely c_{ii} that captures simultaneously the cost of sensing and communications. The nonzero values c_{ij}, shown here as "*," are interpreted as the cost of sensing (when $i = j$) or the cost of transmitting data from the ith to the jth sensor (when $i > j$). The summation along the ith line of $\mathbf{C}$ denotes the total cost of the sensor network to transmit data collected by the ith sensor to the master node. The nonzero values "*" can just denote the existence of a connection or can convey an actual positive cost in energy consumption, if a value is known. If we are just interested in the number of activations, then the matrix $\mathbf{C}$ is just binary, while otherwise, we need to be able to estimate the energy consumption. In fact, the diagonal entries c_{ii} can model the cost of performing the measurement at sensor i and in general, these will obey $c_{ii} \gg c_{ij},\ i \neq j$, i.e., cost of sensing data (which might include acquisition, filtering, storing, compression, etc.) is in most cases significantly larger than the cost of just communication and forwarding information. The authors in [48] have linked the sensing cost to the quality (SNR) of the measurement, i.e., the ℓ_2 norm of the ith row like $c_{ii} \propto \|\mathbf{a}_i^T\|_2^2$ meaning that precise measurements are most likely more expensive from an energy consumption perspective. The same paper proposes to model the communication cost as a function of the distance between the sensors that communicate, i.e., $c_{ij} \propto d(i,j)$ where $d(i,j)$ denotes the Euclidean distance between sensors i and j. If no such estimations are available or possible, then the cost matrix $\mathbf{C}$ will be modeled similarly to the selection variable $\mathbf{z}$, nonzero entries will be taken as "1," i.e., the sensor is used but we are not exactly sure about the cost incurred.

To model the sensor network, we use ideas from the GSP literature and experimentally we will generate these networks using standard GSP toolboxes.[§§] From

[§§]P. Nathanael *et al.*, GSPBOX: A toolbox for signal processing on graphs. Available online: https://epfl-lts2.github.io/gspbox-html/.

this perspective, the cost matrix $\mathbf{C}$ can be viewed as a partial (potentially weighted) adjacency matrix. To avoid overly complicated network scenarios and optimal routing issues, we consider that a single (possible multi-hop) predefined path was computed from each sensor to the master node and this is the only communication path. If no graph structure is imposed then simply $\mathbf{C} = \mathbf{I}_{m\times m}$.

4.3.3 Scheduling for a single time instance

Let us now consider the optimization problem we would like to solve considering that there is only one time instance, i.e., $T = 1$. Let us denote by $\mathscr{I}_1$ the set of sensors that perform a measurement at this time. Then we have that:

$$\mathbf{A}_1^T\mathbf{A}_1 = \sum_{i\in\mathscr{I}} \mathbf{a}_i\mathbf{a}_i^T = \mathbf{A}^T \operatorname{diag}(\mathbf{z}_1)\mathbf{A} \in \mathbb{R}^{n\times n}. \tag{4.11}$$

We have now denoted the following: with $\mathbf{z}_1 \in \{0,1\}^m$ a binary variable that has values one only for indices that belong to $\mathscr{I}$ and diag is a function that takes as input a vector and returns a square matrix whose diagonal is the input vector. Ideally, we would like to solve exactly the following non-convex optimization problem:

$$\begin{aligned} \underset{\mathbf{z}_1}{\text{minimize}} \quad & \|\mathbf{z}_1\|_1 \\ \text{subject to} \quad & \text{MSE}(\mathbf{A}_1) \leq \gamma_1 \ \text{ and } \ \mathbf{z}_1 \in \{0,1\}^{m\times 1}, \end{aligned} \tag{4.12}$$

where γ_1 is given by the designer and is a maximum level of error (in the MSE sense) that is allowed. It is important to understand what the limits of this parameters are. If we are using the entire sensor network to measure the parameters, i.e., $\mathbf{A}_t = \mathbf{A}$, then on average this yields the lowest possible MSE, which is

$$\gamma_{\min} = \operatorname{tr}((\mathbf{A}^T\mathbf{A})^{-1}), \tag{4.13}$$

and therefore it is only natural that $\gamma_1 = \delta_1\gamma_{\min}$ with $\delta_1 \geq 1$. Note that if $\delta_1 = 1$, the only feasible solution in (4.12) is to select all the sensors. Now, combining (4.10) with (4.11), we have that:

$$\text{MSE}(\mathbf{A}_t) = \operatorname{tr}((\mathbf{A}_t^T\mathbf{A}_t)^{-1}) = \operatorname{tr}((\mathbf{A}_t \operatorname{diag}(\mathbf{z}_1)\mathbf{A})^{-1}), \tag{4.14}$$

which is a convex quantity (as it is the composition between a convex function, the trace inverse, and a linear function (4.11) in the unknown $\mathbf{z}_1$). Given this formulation, we can now state the following optimization problem that is the convex relaxation of (4.12), which is

$$\begin{aligned} \underset{\mathbf{z}_1}{\text{minimize}} \quad & \mathbf{1}_{m\times 1}^T\mathbf{z}_1 \\ \text{subject to} \quad & \operatorname{tr}((\mathbf{A}_t \operatorname{diag}(\mathbf{z}_1)\mathbf{A})^{-1}) \leq \gamma_1 \ \text{ and } \ \mathbf{z}_1 \in [0,1]^{m\times 1}. \end{aligned} \tag{4.15}$$

Here we have used that since the entries of $\mathbf{z}_1$ are all positive, the ℓ_1 norm is just the sum of the entries, while the discrete constraints are relaxed to the interval $[0,1]$. Of course, in this case, we are not guaranteed to have a binary solution, as $\mathbf{z}_1$ will have in

general entries between [0, 1] and not just at the extreme points. To fix this problem, we will use iteratively reweighted ℓ_1 optimization [61] by following the next steps:

1. Solve problem (4.15) as is.
2. Initialize the weights vector $\mathbf{w}_1 = \mathbf{z}_1$.
3. If $\mathbf{z}_1$ is not binary yet, update it by solving the new optimization problem

$$\begin{aligned} \underset{\mathbf{z}_1}{\text{minimize}} \quad & \mathbf{w}_1^T \mathbf{z}_1 \\ \text{subject to} \quad & \text{tr}((\mathbf{A}_t \, \text{diag}(\mathbf{z}_1)\mathbf{A})^{-1}) \leq \gamma_1 \;\text{ and }\; \mathbf{z}_1 \in [0,1]^{m\times 1}, \end{aligned} \tag{4.16}$$

   i.e., we are weighting the summation in order to "push to" zero entries that are low and "lift to" one those that are high in $\mathbf{z}_1$. Go to step 2 and repeat the process until the solution $\mathbf{z}_1$ does not change anymore.
4. If the solution $\mathbf{z}_1$ is still not binary, then select the largest entry in $\mathbf{z}_1$, which is not one and set it to one. With this constraint now fixed, go to step 2 and repeat the process until $\mathbf{z}_1$ is binary.

The idea of the approach is to let the iterative process set to zero/one as many entries as possible and only intervene with a hard constraint when it gets stuck permanently. In such a case, we have to set to one and not zero to guarantee that the MSE performance indicator is below the threshold, otherwise, we risk the possibility of an infeasible solution. Also, note that in this formulation, we cannot control directly the number of sensors that are activated. Of course, indirectly, lowering γ_1 will lead to more activations of sensors in general.

We finally note that the scheduling strategy we develop can be applied also when considering some other objective functions to optimize for the network operation. As long as that new objective function is either convex or concave, the proposed method extends directly and trivially with minimal modifications. The requirement to have an appropriate objective function is there to ensure that the same numerical solver can be used.

This completes our discussion on the sensor scheduling problem with a single time step, i.e., sensor selection. We now move to the case where the sensor network operates over multiple time instances and we want to take into account two additional factors: balancing the resources of the network and taking into account the possibility of sensor failure.

4.3.4 Scheduling for multiple time instances

In the previous section, we have seen a method for approximating the optimal set of sensors from a network that we activate to establish a minimum level of accuracy when estimating some parameters. In many real-world situations, a sensor network does not just take a snapshot of some unknown quantities and then becomes inactive, but it continuously measures and monitors their progress over time. Given the approximate solution that we got in the previous section, the obvious solution would be to use those sensors all the time. This would, of course, raise the question of why bother have the large network in the first place? What purpose would it serve? From this perspective, the problem that we previously looked at might be practically a *sensor*

placement problem: we have several candidate sensor types and sensor positions and the goal is to select the best combinations such that the real physical network will perform measurements with high accuracy.

In this section, we extend the goal previously set. Yes, we want at each time instance a minimum level of accuracy but we also have now two additional constraints: a balance between the network devices utilization and the accuracy (MSE is below a given threshold) by *scheduling resources* and take into account, adaptively, sensors breakdown by *rescheduling resources*.

We deal with the same m sensors that need now to be scheduled over T time instances. We define the binary scheduling table:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_T \end{bmatrix} \in \{0,1\}^{m \times T}, \tag{4.17}$$

where the scheduler at time t is denoted $\mathbf{z}_t \in \{0,1\}^{m \times 1}$ (these are the columns of the table $\mathbf{Z}$). It is convenient to access a particular entry in the scheduling table and therefore z_{ij} denotes whether the ith sensor is active during the jth time instance.

Based on the optimization problems we have seen before, we now propose to solve the following convex problem:

$$\begin{aligned} \underset{\mathbf{Z}}{\text{minimize}} \quad & \sum_{t=1}^{T} \mathbf{w}_t^T \mathbf{z}_t + \lambda \max\left(\mathbf{W}\mathbf{C} \sum_{t=1}^{T} \mathbf{z}_t \right) \\ \text{subject to} \quad & \operatorname{tr}((\mathbf{A}^T \operatorname{diag}(\mathbf{z}_t)\mathbf{A})^{-1}) \leq \delta_t \gamma_{\min} \quad \text{for } t = 1, \ldots, T \\ & \mathbf{Z} \in [0,1]^{m \times T} \quad \text{and} \quad \sum_{t=1}^{T} \mathbf{z}_t \geq \mathbf{1}_{m \times 1}. \end{aligned} \tag{4.18}$$

Let us analyze each element of this problem separately. First, note that this is a convex optimization problem. Of course, it is a relaxation of the problem we would actually like to solve exactly (i.e., with the binary scheduling table in (4.17)). That is why, we again have used the idea of ℓ_1 penalization and weights (the first term of the objective function) to try to obtain a binary solution $\mathbf{Z}$. The second term of the objective function discourages the activation of the same sensor over multiple time instances, in the spirit of the ℓ_∞ norm: we penalize the maximum (highest) entry. The ℓ_∞ combines: the activation frequency of each sensor (the summation over the $\mathbf{z}_t$), the topology of the sensor network ($\mathbf{C}$), and any prior information about the power profile of the sensors ($\mathbf{W}$). The matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$ is diagonal and acts as a weight that describes our motivation to use one sensor rather than another. For example, if $w_{ii} = 0$ we are modeling a situation where we do not care if the ith sensor is used in all time instances (maybe because we know it is highly robust or that it has an unlimited power supply at its disposal or that it has been deployed in a sensitive area and therefore its measurements are highly valuable). Otherwise, let us say that, for example, $w_{22} = 2$, while all other $w_{ii} = 1$, for $i \neq 2$, would discourage the activation of the second sensor twice as much as any other, on average. Then, we have the requirement that each time instance performs the estimation with a minimum degree of accuracy at each time step. Finally, the relaxed constraint of sub-unitarity and that

we want all sensors in our network to be activated at least once in the lifetime of the network.

Overall, this new problem is a regularized optimization problem and the regularization parameter $\lambda \in \mathbb{R}_+$ needs to be chosen appropriately. The parameter λ controls the trade-off between accuracy (on one extreme, select all sensors all the time) and activation of the sensors (on the other extreme, do not activate anything to save power). Choosing λ is done by running a grid search to decide the desired trade-off (a Pareto curve). A trade-off necessarily appears because for the trivial value of $\lambda = 0$, the sensor network would activate at each step all the sensors and therefore produce the most accurate results. Conversely, for $\lambda \to \infty$ no sensors would be activated in any of the time slots, i.e., maximum EE with no measurements from the network. It is up to the user to choose the desired balance and solve the tension between sensing accurately and sensing efficiently.

4.3.5 Adaptive scheduling for multiple time instances

We now consider the final element in the setup of our formulation. We assume that in the T time instances the sensor network can change, and in particular we assume that some sensors might fail or are constrained by limited energy availability or battery life. Initially, we schedule our sensors using the optimization problem in (4.18), but then we assume that at time t_f the sensors with indices in the set $\mathscr{F}$ are offline. The optimization problem we consider now is

$$\begin{aligned}
\underset{\mathbf{z}_{t_f+1},\ldots,\mathbf{z}_T}{\text{minimize}} \quad & \sum_{t=t_f+1}^{T} \mathbf{w}_t^T \mathbf{z}_t + \lambda \max\left(\mathbf{WC} \sum_{t=1}^{T} \mathbf{z}_t\right) \\
\text{subject to} \quad & \mathbf{z}_t \in [0,1]^{m\times T} \quad \text{and} \quad \operatorname{tr}((\mathbf{A}^T \operatorname{diag}(\mathbf{z}_t)\mathbf{A})^{-1}) \leq \delta_t \gamma_{\min} \\
& z_{it} = 0 \ \text{ for } \ i \in \mathscr{F} \text{ and } t = t_f + 1, \ldots, T \\
& \left(\mathbf{C} \sum_{t=t_f+1}^{T} \mathbf{z}_t\right)_{\mathscr{F}} = \mathbf{0}_{|\mathscr{F}|\times 1} \\
& \left(\sum_{t=1}^{T} \mathbf{z}_t\right)_{\overline{\mathscr{F}}} \geq \mathbf{1}_{|\overline{\mathscr{F}}|\times 1}.
\end{aligned} \tag{4.19}$$

Note that the scheduling takes place only on variables from time $t_f + 1$ with the additional constraints that any sensor index in $\mathscr{F}$ can no longer be activated, while the ℓ_∞ constraint still takes into account the history of the sensors' activations. We have denoted $\overline{\mathscr{F}} = \{1, \ldots, m\} \backslash \mathscr{F}$, i.e., the complement of $\mathscr{F}$, and $|\overline{\mathscr{F}}|$ denotes the size of the set. We keep the previous historic activations fixed in $\mathbf{z}_1, \ldots, \mathbf{z}_{t_f}$ as we still have to keep track of the most used sensors. The second constraint $z_{it} = 0$ from the current time step makes sure that the sensors in $\mathscr{F}$ will not be activated in the future. Still, they can be part of the path from some other sensor (that is not in $\mathscr{F}$) to the master node. This is again not allowed, and therefore the third constraint guarantees that sensors in $\mathscr{F}$ do not receive any information to forward toward the master node.

Algorithm 4.1: Adaptive sensor scheduling by ℓ_1/ℓ_∞ minimization
Input: The sensing matrix of the network with m sensors $\mathbf{A} \in \mathbb{R}^{m\times n}$, the total number of time instances T, the maximum allowed error $\delta > 1$, the regularization parameter $\lambda > 0$, and the matrix of communication costs $\mathbf{C} \in \mathbb{R}^{m\times m}$.
Output: The scheduling table $\mathbf{Z} \in \{0,1\}^{m\times T}$ for sensor activations at each time step.

Initialize variables: set initial weights $\mathbf{w}_t = \mathbf{1}_{m\times 1}$ and initial all-zero solution $\mathbf{z}_t = \mathbf{0}_{m\times 1}$ for $t = 1,\ldots,T$, i.e., $\mathbf{Z} = \mathbf{0}_{m\times T}$; initialize sets $\mathcal{N} = \emptyset$ indexing sensors that are not selected and $\mathcal{K} = \emptyset$ indexing sensors that are selected; and establish the best MSE performance $\gamma_{\min}$ by (4.13).
Iterative procedure:
1. Set $\mathbf{Z}^{(\text{prev})} \leftarrow \mathbf{Z}$.
2. Update weights according to $w_{ij} = (z_{ij}^{(\text{prev})} + \varepsilon)^{-1}$.
3. Solve (4.18) (or (4.19) if in adaptive mode) with the additional linear equality constraints $z_{ij} = 1,\ \forall\,(i,j) \in \mathcal{K}$, and $z_{ij} = 0,\ \forall\,(i,j) \in \mathcal{N}$, to get the current $\mathbf{Z}$.
4. Update the sets $\mathcal{N} = \{(i,j) \mid z_{ij} \le \varepsilon\}$ and $\mathcal{K} = \{(i,j) \mid z_{ij} \ge 1-\varepsilon\}$.
5. If the iterative process (4.18) (or (4.19) if in adaptive mode) has converged (or 20 iterations have been completed since the last convergence) and the solution is not binary, i.e., $\|\mathbf{Z} - \mathbf{Z}^{(\text{prev})}\|_F^2 \le \varepsilon$ and $|\mathcal{N}| + |\mathcal{K}| = mT$, then set $\mathcal{K} \leftarrow \mathcal{K} \cup \{\arg\max_{(i,j)} z_{ij},\ (i,j) \in \mathcal{K}\}$ and update $\mathbf{Z}$ such that $z_{ij} = 1,\ \forall\,(i,j) \in \mathcal{K}$.
6. If the solution is binary, i.e., $|\mathcal{N}| + |\mathcal{K}| = mT$, then continue otherwise go to step 1 of the iterative process.
Adaptive procedure: if t_f is set then update $\gamma_{\min}$, set $\mathcal{K} = \emptyset$, set $\mathcal{N} = \{(i,t),\ \forall\, i \text{ and } t = t_f + 1,\ldots,T\}$, the solution is cleared for all future time slots by setting $\mathbf{z}_t = \mathbf{0}_{m\times 1}$ for $t = t_f + 1,\ldots,T$, and we go to step 1 of the iterative process to compute the new activations given the new network.

In the adaptive mode, a large number of sensors might be removed from the network and therefore the sensing ability might change significantly. To take this into account we update $\gamma_{\min}$ to reflect the best new performance achievable by the new sensor network. In most situations, sensor failure does lead to more activations of the sensors that are left in the network to keep a similar level of sensing accuracy.

4.3.6 The proposed algorithm

Based on the convex optimization problems described before, we now propose an algorithm for the adaptive sensor scheduling problem. The full procedure is depicted in Algorithm 4.1. We note that Algorithm 4.1 is an improved variant of the algorithm described in [48]. The improvement consists of the capacity of the new algorithm to cope with adaptive situations: the network is not static, but its topology can change during the scheduled T time instances and the algorithm has to take this situation into account and reschedule network resources in real time.

The proposed method keeps track of two sets $\mathcal{K}$ and $\mathcal{N}$, storing the indices from the solution $\mathbf{Z}$ for the activation and deactivation of sensors, respectively. Our

algorithm terminates only after every sensor indexed is either in $\mathscr{K}$ or $\mathscr{N}$, i.e., we have reached a binary solution. Every convex relaxation approach to binary optimization problems needs a similar rounding procedure as we describe here. The goal of the proposed approach is to apply a thresholding operation only when the iterative procedure converges to nonbinary solutions, and even then only a single variable (the highest entry in $\mathbf{Z}$ which is not one) is forcibly thresholded. The same threshold is applied if the iterative reweighted ℓ_1 does not converge (in our case this means that convergence is not achieved in 20 iterations—which happens rarely but does happen).

Given its structure, the result of the proposed method is that most variables are set to zero/one by the optimization procedure. Classically, when using convex relaxations of binary problems we solve a single instance of the relaxed problem followed by a rounding procedure. This approach works very well for a single time instance, i.e., when there are only m variables, but the performance deteriorates quickly in the sense that many entries will be nonbinary when mT is large. If many entries are nonbinary and we round the result to the nearest binary solution we will end up with a very large number of sensor activations. This is in contradiction with our goal of reducing energy consumption by activating sensors only when absolutely necessary. Therefore, our approach is different from many other strategies that usually threshold only once, at the end of the procedure.

Algorithm 4.1 is numerically efficient. It enjoys polynomial complexity and can be implemented using any off-the-shelf convex optimization library in either MATLAB® or python (in this chapter we have used CVX). Furthermore, as the algorithm progresses, the optimization problems (4.18) and (4.19) get progressively smaller. Starting from the mT variables, as the sets $\mathscr{K}$ and $\mathscr{N}$, grow they remove variables with equality constraints by ensuring that indices allocated to these sets will never flip—once an entry (i,j) has been allocated to either $\mathscr{K}$ or $\mathscr{N}$ it will remain there for the remainder of the algorithm. As such, the first steps of the iterative procedure in Algorithm 4.1 are relatively slower than the final steps, which will deal with problems that have a number of variables much smaller than mT.

The only fixed parameter of Algorithm 4.1 is set to $\varepsilon = 0.001$ (the thresholding parameter) while λ is found using a grid search.

Due to the trace inverse constraints, the proposed optimization problem is a semidefinite program (SDP) with binary constraints used iteratively and therefore its analysis in terms of the optimality of the solution is very difficult in general (without any assumptions on the measurement matrix $\mathbf{A}$). In our case, the theoretical analysis is even harder due to the addition of the sets $\mathscr{K}$ and $\mathscr{N}$ that store all indices that were decided to be binary. As such, we extensively validate our approach experimentally. Furthermore, the thresholding operation complicates a possible analysis, but as we will see in the next experimental section it does provide very good performance.

4.4 Experimental results

In this section, we provide experimental numerical simulations to show the performance of the proposed method to schedule a sensor network over time while also

balancing its power consumption, communication effort, and allowing for unexpected sensor failures.

We model the sensor measurements $\mathbf{A}$ in two possible ways: random measurements drawn from i.i.d. standard Gaussian distributions or random tight frames [48]. In the latter case, all measurement matrices used are m-tight, i.e., we scale the measurement matrix with the number of sensors. They are constructed after projecting random Gaussian matrices on the set of tight matrices (numerically this is done by taking the polar factor of a random Gaussian matrix via the singular value decomposition: $\mathbf{A} \leftarrow \sqrt{m}\mathbf{U}\mathbf{V}^T$ starting from the random matrix $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$).

Communications constraints are modeled with graph structures, as depicted in Figure 4.3. These structures serve to quantify the communication costs and are essential in the adaptive setting where sensor failure can lead to removal from the network of many other sensors that are no longer able to communicate the collected data. For the graph generation, we use the GSP Toolbox from which we use the community graph generation feature. The toolbox also provides coordinates for the sensors and we control explicitly the connectivity between the groups of sensors by adding/removing communications links. We assume dense connectivity in the same group of sensors and sparse connectivity between them.

Based on these assumptions, we next consider scheduling sensor networks in 5G systems with energy and communications constraints. We separate the simulations between two cases: static network and adaptive (when we allow reconfiguration of the network, say due to sensor failure or communications node failure).

4.4.1 Scheduling without sensor failures

To demonstrate the versatility of Algorithm 4.1 to scheduling problems, we show how to deal with energy and communication constraints when scheduling a sensor network over multiple, fixed, time instances.

First, we show the implicit energy constraint approach, i.e., we assume no explicit information about the energy profiles of the sensors, and our goal is to operate the sensor network over T time instances such that we diversify the activations and do not

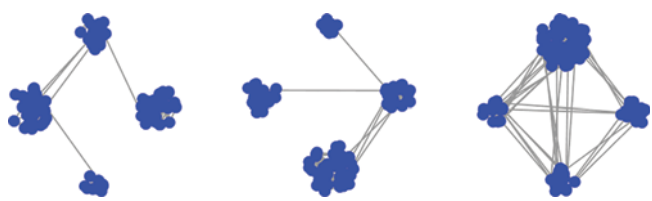


Figure 4.3 Three examples of network architectures we consider with different levels of connectivity between four groups of sensors (from left to right with increased number of communications links). The graphs are generated using the GSP Toolbox and model the cost matrix $\boldsymbol{C}$, starting from the adjacency matrices of the graphs

activate the same sensors continuously at each time instance. The ℓ_1/ℓ_∞ style optimization problem balances between the estimation accuracy of the sensor networks and making sure that the sensing is distributed more evenly between the network sensors. Results are shown in Figure 4.4. We show results for a low $m = 100$ to ease of exposition, as we show activations per sensor. The measurement matrix is taken to be standard Gaussian, and the performance indicator is fixed to $\gamma_t = 0.5$ for all $t = 1, \ldots, T$, i.e., 50% measurement accuracy on average as compared to the full network. We fix the number of time instances to $T = 10$.

Simulation results with the regularization parameter $\lambda = 100$ are shown in Figure 4.4 (top right). The plot clearly shows a more balanced activation of the sensors, as opposed to the results in Figure 4.4 (top left) that are obtained with no regularization, i.e., $\lambda = 0$. Given higher regularization parameter values λ the sensor scheduling is "moved" to rarely select the same sensors again (unless absolutely necessary to achieve the prescribed MSE). For example, in Figure 4.4 (top right) most sensors are selected six times with a maximum of eight times as compared to Figure 4.4 (top left) where several sensors are selected in all ten-time instances, while others are never selected. This comes at the cost of activating, overall, a larger number of sensors over the ten-time instances.

The purpose of Figure 4.4 (top left) is to show, for reference, a scheduling table generated by the proposed algorithm without the ℓ_∞ regularization (the max regularization). The almost flat solution depicted in Figure 4.4 (top right) is typical of solutions to convex optimization problems whose objective functions involve ℓ_∞ regularization (for details see [39, Chapter 6]). For our purposes, these solutions are ideal. The nearly uniform activation of the sensors over time ensures that the sensing

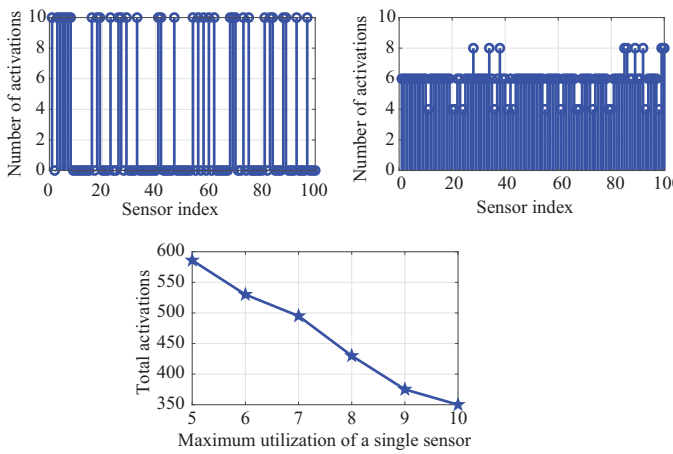


Figure 4.4 Number of activations for a sensor network with $m = 100$ elements: no ℓ_∞ regularization (top left), with ℓ_∞ regularization and parameter $\lambda = 100$ (top right) and activations of sensors as a function of the varying regularization parameter λ (bottom)

workload is distributed across the network and thus ensures balanced power consumption together with increased robustness and fault tolerance in the case of any particular sensor failure. Figure 4.4 (bottom) shows one of the intuitive side effects of Algorithm 4.1: we can reduce the frequency with which one particular sensor is activated but at the cost of activating other (usually more or even many more) sensors from the network such that the imposed estimation accuracy is fulfilled. Figure 4.4 (bottom) plots the maximum sensor activation max ($\sum_{t=1}^{T} \mathbf{z}_t$) against the total sensor activations $\sum_{i=1}^{m} \sum_{t=1}^{T} z_{it}$. We achieve different values for these activations by varying the parameter λ, in this case, six values between 0 and 100.

In Figure 4.5, we also highlight the use the weight matrix $\mathbf{W}$ that serves to add preferences or a-priori information about the network into the planning of the sensors. The experiments are as follows: in Figure 4.5 (top left) we have that $w_{ii} = 1,\ i = 1, \ldots, 25$ and $w_{ii} = 2,\ i = 26, \ldots, 100$ and therefore the first sensors are used far more often than the later; in Figure 4.5 (top right), we have that $w_{ii} = 7,\ i = 1, \ldots, 50$ and $w_{ii} = 1,\ i = 51, \ldots, 100$ and therefore the first half of the sensors is heavily penalized for any activation (there is one activation as per the constraint to use each sensor at least once); in Figure 4.5 (bottom), we have that $w_{ii} = 1,\ i = 1, \ldots, 75$ and $w_{ii} = 0,\ i = 76, \ldots, 100$ and therefore the last quarter of sensors are used in all time instances as there is no penalty in overusing these devices (possibly highly robust sensors, placed in sensitive areas and enjoying a virtual unlimited power supply).

We now superimpose the graph structure on the measurement network. We assume a network whose topology is similar to the examples in Figure 4.3: $m = 100$

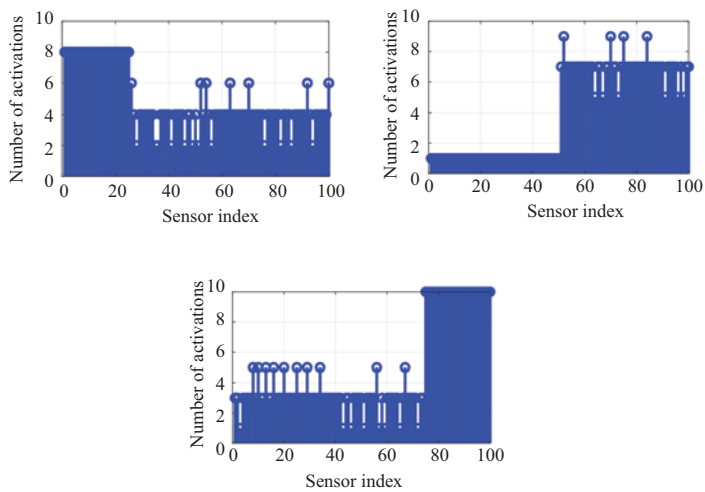


Figure 4.5 *Number of activations for a sensor network with $m = 100$ elements and various weight matrices that encourage/discourage the usage of particular sensors*

sensor broadly grouped in four, unbalanced, groups. Regarding the sensing and communications cost, we assume that entries of the matrix $\mathbf{C}$ are characterized as

$$c_{ii} = O(\|\mathbf{a}_i^T\|_2^2),\ i = 1, \ldots, m, \tag{4.20}$$

for the sensing cost incurred at the sensor and

$$c_{ij} = O(\|\mathbf{p}_i - \mathbf{p}_j\|_2^2),\ i,j = 1, \ldots, m,\ i > j, \tag{4.21}$$

as the cost of communication between sensors i and j. For communication between the groups of sensors, we assume a land connection and therefore a cost of $O(1)$. We have denoted with $\mathbf{p}_i$ the 2D position of sensor i.

In Figure 4.6, we show the achieved MSE levels as a function of the energy consumption of the whole sensor network. A clear trade-off between the two becomes apparent in the spirit of a no-free-lunch result. As expected, to achieve the high estimation accuracy (the lowest levels of MSE) we need (almost) all the sensors activated (almost) all of the time. Therefore, the highest accuracy of the network can only be achieved with substantial energy consumption. Fortunately, giving up some accuracy in the MSE has a substantial positive impact on energy consumption, especially at the limit of the best accuracy. Depending on the currently available energy supplies, Figure 4.6 shows what levels of MSE estimation accuracy are possible with the sensor network. In the figure, we distinguish between the Gaussian and tight measurement models (left and right plots). Notice that the tight random measurement model behaves much better in two ways: lower levels of the MSE are reached for the same energy measure and the MSE scales much better (decrease is steepest) with the energy consumption.

4.4.2 Scheduling with sensor failures

In this section, we study the behavior of the proposed algorithm in situations where the network topology changes. We mostly consider situations where elements of the network break and it falls on the other network elements to perform more measurements to keep the prescribed level of performance. This might arise, for example, in a scenario where the sensors have limited energy availability over the period of operation and run out of available energy due to being activated too often. For simplicity, we assume that the change times are at each quarter of the total number of time instances. We consider a network with $m = 500$ sensors and $T = 100$ time instances,

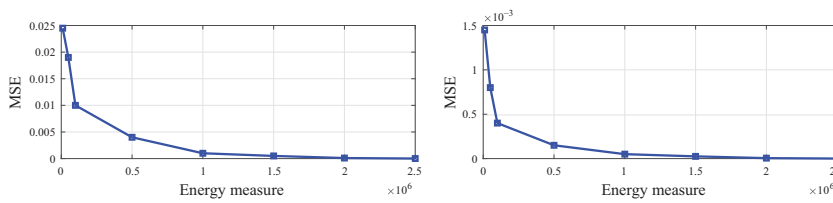


Figure 4.6 Energy consumption for a sensor network with $m = 100$ elements under a Gaussian (left) and tight (right) measurement models

while the measurement model is the random Gaussian one. The nominal measurement performance level is kept at 50% of the capacity of the whole network.

Results are shown in Figure 4.7 and we break them down in four steps: Figure 4.7 (top left) depicts the scheduling scheme before any sensor failures, this can be done off-line before the network is activated, and if no failures appear, the scheduling will persist for the rest of the time slots; Figure 4.7 (top right) shows the network activations when 100 sensors become unavailable (e.g., due to the failure of a communication hub) and clearly, on average, the utilization of the remaining sensors increases; Figure 4.7 (bottom left) shows a further deterioration of the network: an additional 200 sensors fail, while the previously failed 100 sensors are still offline, and the operating network sensors are pushed to almost permanent activations; finally, Figure 4.7 (bottom right) shows what happens when all network sensors are made available again and the previously offline sensors are now activated significantly more than the others (especially the 50 sensors that were offline for half of the operating lifetime of the network).

In these experimental results, we observe the same characteristics as with previous results: failure of network resources leads to higher strain on the active elements, for which balancing utilization becomes that bit more important.

In this setting, the running time of the proposed algorithm is extremely important. While the initial scheduling table can be computed offline, real-time changes to the network configuration happen during the operation of the network and therefore rescheduling needs to be done during one time slot. Regarding the running time, although it enjoys polynomial complexity, Algorithm 4.1 is slower in general than some of the state-of-the-art greedy methods from the literature, which also have polynomial asymptotic complexity but exhibit lower constant factors, i.e., they are $2\times$ or $3\times$ faster. Still, the numerical complexity of Algorithm 1 is polynomial in the

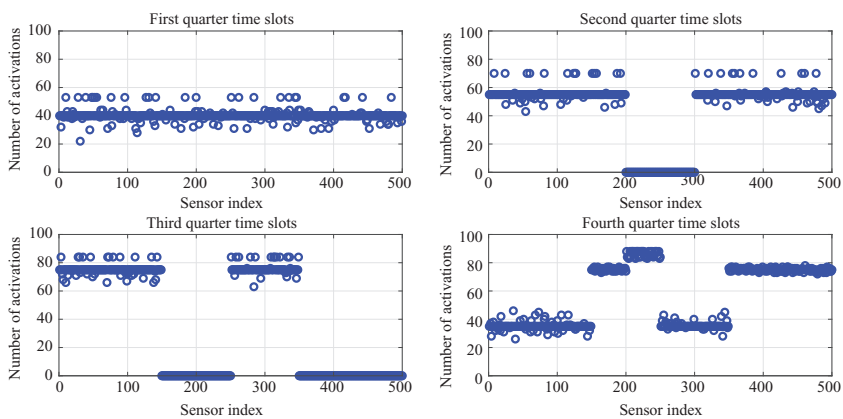


Figure 4.7 Adaptive sensor scheduling for a network of $m = 500$ sensors over $T = 100$ time instances where the network configuration changes three times during its lifetime

number of variables and constraints and simulations from this chapter take only a few minutes to complete on a modern computing Intel i7 system with sufficient memory resources. Therefore, the proposed method is well suited for highly dynamical sensor network scheduling. Furthermore, a specialized convex solver, which uses efficient first-order optimization methods, could be considered to deal with the requirements (both memory footprint and running time) of extra-large problems: over $m = 10{,}000$ sensors scheduled in more than $T = 1{,}000$ time instances.

We highlight again that one of the advantages of our convex optimization-based approach is that they allow easy generalizations and extensions. With few modifications, Algorithm 4.1 allows the addition of multiple extra constraints as long as they can also be coded with convex constraints, like: operating the sensor network over a variable number of multiple time instances, without any repetition of sensor selection, with some preference for a few sensors and not for others, etc. From this perspective, our setting is ideal for modeling, with minimum effort, real-world applications where unexpected constraints show up.

4.5 Conclusions

Large networks play a central role in the development of IoT and 5G technologies. In this chapter, we have proposed a new algorithm for the scheduling of resources in 5G networks and it aims for two goals: increase the EE of the network and its robustness to hardware failures. We provide simulation results that show how the proposed algorithm strikes a balance between the accuracy of the network's measurements/operations and the frequency of network's resource utilization or an absolute measure of energy when this is available. The algorithm can provide adaptive, real-time rescheduling of the network resources such that the estimation accuracy is kept at similar levels even when unexpected events such as hardware and communication failures take place.

References

[1] Andrews JG, Buzzi S, Choi W, *et al.* What will 5G be? IEEE Journal on Selected Areas in Communications. 2014;32:1065–1082.

[2] Gandotra P, Jha RK, and Jain S. Green communication in next generation cellular networks: A survey. IEEE Access. 2017;5:11727–11758.

[3] Strubell E, Ganesh A, and McCallum A. Energy and policy considerations for deep learning in NLP. In: the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy; 2019.

[4] Fehske A, Fettweis G, Malmodin J, *et al.* The global foot-print of mobile communications: The ecological and economic perspective. IEEE Communications Magazine. 2011;49:55–62.

[5] Marzetta T, Larsson EG, Yang H, *et al.* Fundamentals of Massive MIMO. Cambridge University Press; 2016.

[6] Rusu C, Mendez-Rial R, González-Prelcic N, *et al.* Low complexity hybrid precoding strategies for millimeter wave communication systems. IEEE Transactions on Wireless Communications. 2016;15:8380–8393.

[7] Mendez-Rial R, Rusu C, González-Prelcic N, *et al.* Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches? IEEE Access. 2016;4:247–267.

[8] Zhang J, Dai L, Li X, *et al.* On low-resolution ADCs in practical 5G millimeter-wave massive MIMO Systems. IEEE Communications Magazine. 2018;56:205–211.

[9] Gao H, Ejaz W, and Jo M. Cooperative wireless energy harvesting and spectrum sharing in 5G networks. IEEE Access. 2016;4:3647–3658.

[10] Li S, Xu LD, and Zhao S. 5G Internet of Things: A survey. Journal of Industrial Information Integration. 2018;10:1–9.

[11] Lynggaard P and Skouby KE. Deploying 5G-technologies in smart city and smart home wireless sensor networks with interferences. Wireless Personal Communications. 2015;81:1399–1413.

[12] He S, Chen J, Jiang F, *et al.* Energy provisioning in wireless rechargeable sensor networks. IEEE Transactions on Mobile Computing. 2013;12:1931–1942.

[13] Gandotra P, Jha RK, and Jain S. A survey on device-to-device (D2D) communication: Architecture and security issues. Journal of Network and Computer Applications. 2016;78:9–29.

[14] He A, Wang L, Chen Y, *et al.* Spectral and energy efficiency of uplink D2D underlaid massive MIMO cellular networks. IEEE Transactions on Communications. 2017;65:3780–3793.

[15] Safdar GA, Ur-Rehman M, Muhammad M, *et al.* Interference mitigation in D2D communication underlaying LTE-A network. IEEE Access. 2016;4:7967–7987.

[16] Chen S, Qin F, Hu B, *et al.* User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions. IEEE Wireless Communications. 2016;23(2):78–85.

[17] Park H and Hwang T. Energy-efficient power control of cognitive femto users for 5G communications. IEEE Journal on Selected Areas in Communications. 2016;34:772–785.

[18] I C, Rowell C, Han S, *et al.* Toward green and soft: A 5G perspective. IEEE Communications Magazine. 2014;52:66–73.

[19] Ge X and Zhang W. 5G Green Mobile Communication Networks. Springer Singapore Press; 2019.

[20] Kang X, Chia Y, Sun S, *et al.* Mobile data offloading through a third-party WiFi access point: An operator's perspective. IEEE Transactions on Wireless Communications. 2014;13(10):5340–5351.

[21] Luong P, Nguyen TM, Le LB, *et al.* Energy-efficient WiFi offloading and network management in heterogeneous wireless networks. IEEE Access. 2016;4:10210–10227.

[22] Comşa IS, Zhang S, Aydin M, *et al.* Towards 5G: A reinforcement learning-based scheduling solution for data traffic management. IEEE Transactions

on Network and Service Management. 2018;15:1661–1675. Available from: https://uwe-repository.worktribe.com/output/855212.

[23] Gomez K, Goratti L, Granelli F, *et al.* A comparative study of scheduling disciplines in 5G systems for emergency communications. In: 1st International Conference on 5G for Ubiquitous Connectivity; 2014. pp. 40–45..

[24] Sinha A, Andrews M, and Ananth P. Scheduling Algorithms for 5G Networks With Mid-haul Capacity Constraints; 2019.

[25] Femenias G, Riera-Palou F, Mestre X, *et al.* Downlink scheduling and resource allocation for 5G MIMO-multicarrier: OFDM vs FBMC/OQAM. IEEE Access. 2017;5:13770–13786.

[26] Vora A and Kang KD. Effective 5G wireless downlink scheduling and resource allocation in cyber-physical systems. Technologies. 2018;6:105.

[27] Gueguen C, Ezzaouia M, and Yassin M. Inter-cellular scheduler for 5G wireless networks. Physical Communication. 2016;18:113–124. Special Issue on Radio Access Network Architectures and Resource Management for 5G. Available from: http://www.sciencedirect.com/science/article/pii/S187449071500052X.

[28] Caire G. In: Wong VWS, Schober R, Ng DWK, *et al.*, editors. Massive MIMO Scheduling Protocols. Cambridge University Press; 2017. pp. 335–357.

[29] Huang Y, Li S, Hou YT, *et al.* GPF: A GPU-based design to achieve ∼100 μs scheduling for 5G NR. In: MobiCom; 2018.

[30] Pedersen K, Pocovi G, Steiner J, *et al.* Agile 5G scheduler for improved E2E performance and flexibility for different network implementations. IEEE Communications Magazine. 2018;56(3):210–217.

[31] Müller CF, Galaviz G, Andrade ÁG, *et al.* In: Sanchez MA, Aguilar L, Castañón-Puga M, *et al.*, editors. Evaluation of Scheduling Algorithms for 5G Mobile Systems. Cham: Springer International Publishing; 2018. pp. 213–233.

[32] Yu W, Xu H, Zhang H, *et al.* Ultra-dense networks: Survey of state of the art and future directions. In: 2016 25th International Conference on Computer Communication and Networks (ICCCN); 2016. pp. 1–10.

[33] Viering I, Peltomäki M, Tirkkonen O, *et al.* A distributed power saving algorithm for cellular networks. In: Spyropoulos T, and Hummel KA, editors. Self-Organizing Systems. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. pp. 232–237.

[34] Cho M, Ban T, Jung BC, *et al.* A distributed scheduling with interference-aware power control for ultra-dense networks. In: 2015 IEEE International Conference on Communications (ICC); 2015. pp. 1661–1666.

[35] Han D, Wu J, Zhang H, *et al.* Optimal sensor scheduling for multiple linear dynamical systems. Automatica. 2017;75:260–270. Available from: http://www.sciencedirect.com/science/article/pii/S0005109816303508.

[36] Jiang C, Soh YC, and Li H. Sensor placement by maximal projection on minimum eigenspace for linear inverse problems. IEEE Transactions on Signal Processing. 2016;64(21):5595–5610.

[37] Shamaiah M, Banerjee S, and Vikalo H. Greedy sensor selection: Leveraging submodularity. In: 49th IEEE Conference on Decision and Control (CDC); 2010. pp. 2572–2577.

[38] Ranieri J, Chebira A, and Vetterli M. Near-optimal sensor placement for linear inverse problems. IEEE Transactions on Signal Processing. 2014;62(5): 1135–1146.

[39] Boyd S and Vandenberghe L. Convex Optimization. New York, USA: Cambridge University Press; 2004.

[40] Chepuri SP and Leus G. Sparsity-promoting sensor selection for non-linear measurement models. IEEE Transactions on Signal Processing. 2015;63(3): 684–698.

[41] Mo Y, Ambrosino R, and Sinopoli B. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. Automatica. 2011;47(7):1330–1338.

[42] Chen Y, Chuah CN, and Zhao Q. Sensor placement for maximizing lifetime per unit cost in wireless sensor networks. In: MILCOM 2005 – 2005 IEEE Military Communications Conference; 2005. pp. 1097–1102, Vol. 2.

[43] Liu S, Vempaty A, Fardad M, *et al.* Energy-aware sensor selection in field reconstruction. IEEE Signal Processing Letters. 2014;21(12): 1476–1480.

[44] Liu S, Chen F, Vempaty A, *et al.* Sparsity-promoting sensor management for estimation: An energy balance point of view. In: 2015 18th International Conference on Information Fusion (Fusion); 2015. pp. 231–238.

[45] Tzoumas V, Jadbabaie A, and Pappas GJ. Near-optimal sensor scheduling for batch state estimation: Complexity, algorithms, and limits. In: 2016 IEEE 55th Conference on Decision and Control (CDC); 2016. pp. 2695–2702.

[46] Chamon LFO and Ribeiro A. Near-optimality of greedy set selection in the sampling of graph signals. In: 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2016. pp. 1265–1269.

[47] Rusu C and Thompson J. Node sampling by partitioning on graphs via convex optimization. In: Sensor Signal Processing for Defence Conference (SSPD); 2017. pp. 1–5.

[48] Rusu C, Thompson J, and Robertson NM. Sensor scheduling with time, energy, and communication constraints. IEEE Transactions on Signal Processing. 2018;66:528–539.

[49] Rusu C, Thompson J, and Robertson NM. Balanced sensor management across multiple time instances via l-1/l-infinity norm minimization. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2017. pp. 3311–3315.

[50] Chen T, Kim H, and Yang Y. Energy efficiency metrics for green wireless communications. In: International Conference on Wireless Communications & Signal Processing (WCSP). Suzhou, China; 2010. pp. 1–6.

[51] Ulukus S, Yener A, Erkip E, *et al.* Energy harvesting wireless communications: A review of recent advances. IEEE Journal on Selected Areas in Communications. 2015;33(3):360–381.

[52] Gong J, Thompson JS, Zhou S, *et al.* Base station sleeping and resource allocation in renewable energy powered cellular networks. IEEE Transactions on Communications. 2014;62(11):3801–3813.

[53] Frenger P, Friberg C, Jading Y, *et al.* Radio network energy performance: Shifting focus from power to precision. Ericsson Review. 2014;91:10–17.
[54] Potra FA and Wright SJ. Interior-point methods. Journal of Computational and Applied Mathematics. 2000;124(1):281–302.
[55] Candes EJ and Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory. 2006;52(12):5406–5425.
[56] Studer C, Goldstein T, Yin W, *et al.* Democratic Representations. arXiv e-prints. 2014 Jan; pp. arXiv:1401.3420.
[57] Lee J and Leyffer S. Mixed integer nonlinear programming. The IMA Volumes in Mathematics and Its Applications. Springer-US; 2011.
[58] Cao P, Liu W, Thompson JS, *et al.* Semidynamic green resource management in downlink heterogeneous networks by group sparse power control. IEEE Journal on Selected Areas in Communications. 2016;34(5):1250–1266.
[59] Coates A, Hero III AO, Nowak R, *et al.* Internet tomography. IEEE Signal Processing Magazine. 2002;19(3):47–65.
[60] Weiler RJ, Peter M, Keusgen W, *et al.* Enabling 5G backhaul and access with millimeter-waves. In: 2014 European Conference on Networks and Communications (EuCNC); 2014. pp. 1–5.
[61] Candès EJ, Wakin MB, and Boyd SP. Enhancing sparsity by reweighted l_1 minimization. Journal of Fourier Analysis and Applications. 2008;14(5): 877–905.

Part II

Renewable energy and energy harvesting

This page intentionally left blank

Chapter 5

Renewable energy-enabled wireless networks

Michela Meo[1] and Daniela Renga[1]

The introduction of renewable energy sources (RESs) as power supply for communication systems and, for wireless cellular networks in particular, is becoming more and more attractive for a number of reasons. First, the need to reduce network operation costs through energy saving. Second, the interest in bringing cellular communications to areas of the world where the power grid is not developed and/or reliable, or in emergency situations, generates a great interest in off-grid base stations (BSs) that are energy self-sufficient. Finally, the introduction of RES as power supply is a promising way to start responding to the timely issue of Information and Telecommunication Technology (ICT) sustainability. In this chapter, we discuss the technological challenges associated with the introduction of RES-based power supply for wireless networks. Sources like photovoltaic (PV) panels and small wind turbines are the most suited ones for powering cellular access networks, due to their limited size and relatively ease of deployment. However, these sources are intermittent and generate variable amounts of energy not always easy to predict. Network operations require mechanisms and algorithms for deciding the optimal configuration that depends also on consumption and energy availability. Optimality of network operation is not simply performance maximization but becomes also consumption reduction, cost minimization, and emission reduction, through the optimal usage of the locally produced energy. In addition, considerations on the power supply dimensioning will also be presented in this chapter.

5.1 Introduction

The staggering increase of mobile traffic observed in the recent years is currently leading mobile network operators (MNOs) to deploy denser and denser mobile access networks. This trend is bound to further grow at remarkable pace in the next future. According to Cisco forecast [1], by 2023 there will be nearly 5.3 billion Internet users worldwide, accounting for 66% of the global population, 29.3 billion of networked devices and connections are expected by the same year, with an average of 3.6 connected devices per capita. The number of mobile users worldwide will raise to 5.7

[1]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Turin, Italy

billion and global mobile data traffic will reach more than 77 exabytes per month. The Middle East and Africa will have the strongest mobile data traffic growth in the period 2017–22, with a 56% Compound Annual Growth Rate. Furthermore, the enhancements introduced by 5G technology in mobile networks in terms of higher bandwidth and ultralow latency will enable the deployment of massive Internet of Things (IoT) applications in the smart city framework. Several services entailing critical communications with strict delay constraints will be facilitated in 5G scenarios, like in the case of smart mobility and autonomous vehicles, machine-to-machine communications for factory automation, smart metering, environmental monitoring, smart grid (SG) management, just to cite some examples. By 2023, IoT devices will account for 50% of all networked devices, and over 10% of devices and connections will have 5G capability, leading to a disproportionate effect in the mobile traffic growth, since 5G speeds are expected to be 13 times higher than the average mobile connection [1].

The consequent substantial raise in the cellular traffic entails the need to deploy properly dimensioned cellular networks to make Internet access available everywhere and provide the high bandwidth capacity required for the increasing number of mobile users and for the introduction of applications that result more and more demanding in terms of bandwidth requirements. Considering that the access segment is responsible of up to 80% of the total network consumption [2], it appears evident how the energy demand to operate cellular networks is rapidly growing and MNOs are facing huge operational costs due to power supply [3]. In addition, with the considerable mobile data traffic growth that is expected in the next years in emerging countries, in particular in the Middle East and Africa [4], where the electric grid may not be reliable or even totally absent, MNOs face new issues related to the power supply of mobile access networks that require that the networks are made more independent from the electric grid.

Besides cost, these data hint to sustainability issues. Up to 2% of the total carbon emissions are currently accounted for by the ICT sector, and this percentage is forecast to increase [5]. Considering that the communication function contributes for 19% to the ICT energy consumption [6] and, as already mentioned, radio access networks (RANs) are responsible of 80% of mobile networks consumption [2], improving mobile network sustainability clearly represents an additional objective to be achieved in mobile networks.

In this context, the introduction of RESs as power supply of RANs appears an attracting mean to jointly achieve a number of goals: making the cellular network more independent from the power grid, decreasing the energy costs, and reducing the carbon footprint of mobile communications. Relevant research efforts are devoted to this topic, as it appears from the several studies in the literature focusing on the use of alternative power supply for green cellular networks [7–9]. RESs can be effectively exploited to power BSs jointly with the power grid in hybrid networks to reduce the on-grid energy consumption and the cost for the electric energy supply [10–12]. The potentiality of renewable energy (RE) is also investigated in Cloud-RAN scenarios, where the baseband processing unit of BS is decoupled from the remote radio head to provide more flexibility and address ever-increasing diverse communication needs in the 5G framework [13,14]. Furthermore, with the spreading of *fog computing* as

a way to move computing and storage capability closer to the network edge, RESs are expected to be adopted to power fog nodes, in order to make them more energy sustainable [15].

One of the most suitable solutions in an urban environment, due to the limited environmental impact, is the deployment of PV panels and battery units. While the most suited, still the available space for the installation of a PV-based generation system is generally rather limited due to physical constraints and to possible authorization constraints. It is necessary to trade off the system dimensioning requirements to significantly reduce costs and the carbon footprint, on the one hand, and the feasibility constraints that limit the maximum available surface that can be occupied by the RE generation system, on the other hand.

However, a wider penetration of RE to power BSs is expected in the next future. In particular, for solar energy, more efficient modules are bound to be brought into the market that will improve the feasibility of RE systems, especially in urban environments. Although currently available commercial modules, built with traditional technologies based on crystalline silicon, show an efficiency that never exceeds 20%, emerging technologies (like CPV—Concentrating Photovoltaics) should allow to increase the efficiency to more than 30% [16–18]. This will positively affect the area required to deploy a PV system, since currently almost 5 m^2 are needed per each kWp of PV panel capacity [19]. In addition, the shift from the traditional electric power grid to the SG paradigm has brought several benefits to the electricity management. Several functionalities supported by the SG rely on an efficient exchange of huge monitoring data volumes that will be further enhanced by the 5G technology. In particular, the SG framework allows an active interaction between producer and consumers, leading to a better match between energy demand and supply, and enables power grid monitoring for predictive maintenance, enhanced energy management, and better integration of RESs. However, new cyber vulnerabilities arise in an Internet-connected grid that may be subject to power outages caused by cyber attacks. In this context, RESs represent a promising solution to make the mobile network more independent from the SG and more robust to power outages to guarantee the continuity of service.

In this chapter, we discuss the challenges associated with the introduction of RESs, in particular PV panels, to power mobile access networks. To ease the discussion, we introduce three scenarios that we consider representative of the main families of application contexts and we refer to the scenarios when we present the relevant challenges and some of the solutions. In detail, this chapter is organized as follows. In Section 5.2, we first present an overview of the main challenges raised by the utilization of RE to power mobile networks. Section 5.3 details the green network scenarios that will be investigated in this chapter, whereas Section 5.4 highlights the critical issues to be addressed with specific reference to the different presented scenarios of renewable powered networks, describing possible solutions to achieve MNO goals. Section 5.5 details how the previously presented issues are tackled in real case studies, detailing the actual implementation and application of the proposed solutions and providing some performance analysis. Conclusions are drawn in Section 5.6.

5.2 Renewable energy to pursue mobile operator goals

As previously mentioned, RES-based power supply for the BSs represents a valid solution to decrease the operational expenditures (OPEXs) of MNOs, lower the carbon footprint, and make the access network energy self-sufficient with respect to the power grid. However, some critical issues may raise when introducing RES in mobile networks and, in particular, the main ones are due to the intermittent nature of RE production, as well as the intraday and inter-day variability. In this section, we investigate these variabilities.

5.2.1 Renewable energy production variability

In what follows, we consider as an example the case of the city of Turin, in Italy, that is representative of a large number of temperate areas. The daily energy production profile of a PV panel of 1 kWp production* is reported in Figure 5.1 for 3 months of different seasons. First, observe the different production levels in different months; in summer, the typical production is three times the one of December. This element is already quite critical indicating that a PV panel dimensioning for summer times is clearly inadequate in winter, and, conversely, a dimensioning defined based on winter production leads to extra production in summer times.

Second, it can be noted that different days of the same month have, in their turn, quite different production levels, again indicating the difficulty in defining a proper dimensioning of the PV panels. Moreover, within the same day the profile obviously follows the typical day/night pattern with peak productions around 1 PM and starting production time that is seasonal dependent. To investigate the latter factor, observe the duration of daylight for 3 winter months in Figure 5.2(a) and the corresponding daily production in Figure 5.2(b). The daylight slowly but significantly changes in the period and a high daily variability overlaps with these slow changes. The inter-day and intraday variability is highly affected by medium and short-term variations of the weather conditions (e.g., sunny versus rainy days, fast cloud movements within the same hour in a partially cloudy day, …).

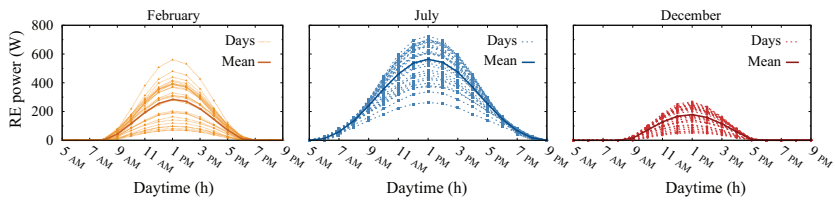


Figure 5.1 Daily profiles of RE power production (W) per 1 kWp in different sample months in Turin (results obtained with PVWatts) [20]

*The *kilowatt peak (kWp)* is the unit of measure of the nominal power of a PV panel, as measured under Standard Test Conditions (STCs). STC include a light intensity of 1,000 W/m^2, a solar spectrum of airmass 1.5 and a module temperature of 25°C. A PV panel with 1 kWp of nameplate capacity is able to produce up to 1 kWh of solar energy per hour when operating at its maximum capacity under STC.

5.2.2 The problem of uncoupled traffic demand and solar energy production

While seasonal variations imply a difficulty in properly dimensioning the power supply system, the inter-day and intraday variability calls for dynamic power management strategies. The strategies need to take into account both energy production and energy demand, keeping in mind that their patterns may not result coupled and the energy management can therefore get even more complicated. Indeed, regarding the solar energy production, its intermittence is strictly linked to the night–day alternating and to the season succession, and the level of solar generation is highly influenced by the location latitude and by the short-term and long-term weather condition variabilities. The mobile traffic patterns vary a lot over time mainly depending on the area type (whether rural or urban regions, residential or business areas or locations where special events occur periodically …), on the period of the week, and on the time of the year (national holidays or vacation periods …). Hence, the energy demand pattern is not always coupled with the RE production profiles. Figure 5.3 shows the traffic profiles for a typical weekday (continuous line) and weekend day (dashed line), both in a business (BA) and a residential (RA) area (red circles and blue squares, respectively). The traffic patterns are derived from real data provided by one of the main Italian mobile operators. In the BA, the traffic peaks are observed during the central hours of the day during the week, while in the evening and during the night almost no traffic is registered, like it is observed during the whole weekend. Conversely, in the RA, the traffic demand gradually rises in the morning hours to achieve the highest

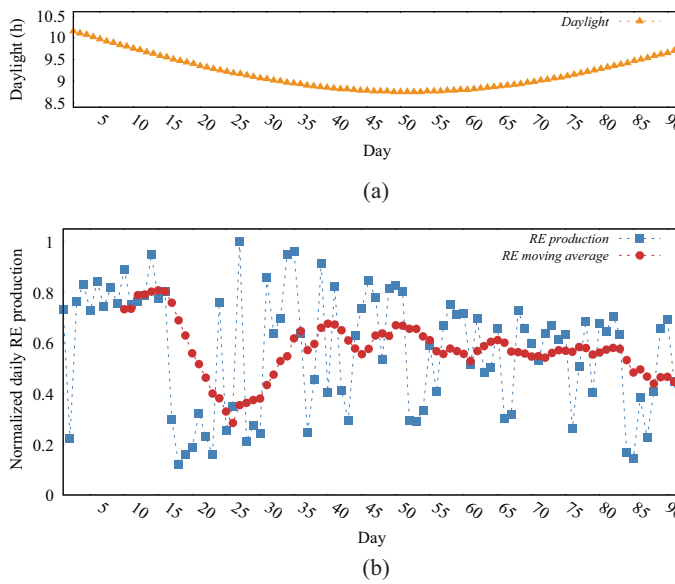


Figure 5.2 (a) Daylight time duration across the 3 months and (b) normalized daily RE production with its moving average [20]

levels in the evening and night hours. Clearly, only in the BA, the weekday energy demand appears somehow coupled with the RE production, whereas in the RA, the energy demand results higher during period in which no RE can be produced. In this latter case, the energy management becomes more critical.

The RE production intermittence and the traffic variability imply the need for some kind of storage, to harvest for future usage any extra amount of energy that is produced but cannot be immediately used. In addition, proper energy management strategies are required to efficiently use the generated RE, limiting the RE wastage and making the RE system dimensioning feasible in terms of capital expenditures (CAPEX) and area occupancy.

5.2.3 Traffic load and BS energy consumption

Finally, concerning the relation between traffic and energy demands, it is worth to be mentioned that the BS energy consumption does not grow proportionally with the traffic load. Figure 5.4 depicts the power consumption profiles for a Long-Term Evolution (LTE) macro BS in a BA and RA (red and blue curves, respectively), during the week-day (continuous lines) and the week-end (dashed lines), based on the same traffic data reported in Figure 5.3 and on the EARTH consumption model [3]. Despite huge traffic variations over time, a consistent fixed amount of power is consumed, even when the traffic is negligible, leading to a power wastage for long periods in which the traffic is low. Considering a micro LTE BS, the little traffic proportionality of the BS power consumption becomes even more evident, with a fixed consumption accounting for up to 80% of the maximum energy demand of the BS. This little load proportionality of the BS consumption represents a critical issue that must be addressed in order to make the mobile network more energy efficient. New generations of BSs exhibit a consumption more proportional to the actual traffic load. However, the penetration of these new access network devices is far from being widespread in mobile systems. Hence, proper solutions need to be deployed and implemented to make the current access network infrastructures capable of dynamically adapting their energy consumption to the actual traffic variations that are observed over time.

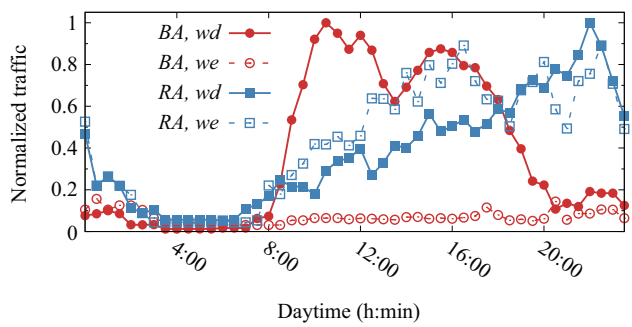


Figure 5.3 Week-day (wd) and week-end (we) traffic loads in a business (BA) and residential (RA) area [21]

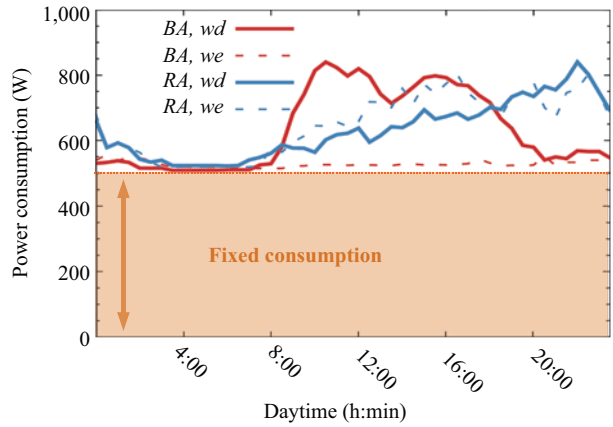


Figure 5.4 Power consumption of an LTE macro base station during a week-day (wd) and a week-end (we), in a business (BA) and residential (RA) area [22]

5.3 Scenarios

We now present the three different scenarios that will be investigated to understand the manifold roles played by the use of RE to power RANs. These scenarios represent the main situations in which the use of RE to power RANs is particularly interesting.

5.3.1 On-grid BSs in an urban environment and reliable power grid

With the exponential increase of the mobile traffic expected in the next years and the need for ubiquitous high-speed mobile access, smoothly supporting users mobility, MNOs are enforced to deploy more and more densified mobile access networks.

The typical considered scenario is shown in Figure 5.5. It consists of a portion of a densified access network, composed by a macro BS that provides baseline coverage, and a set of additional micro BSs that guarantee additional capacity during peak periods of traffic demand. The considered access network is deployed in an urban environment where the electric grid is available and fully reliable. The BSs are powered by a hybrid system, since they can draw energy from the electric grid but they are also equipped with a set of PV panels to locally produce RE. Furthermore, a set of battery units stores, for future usage, any amount of produced energy that is not immediately used. Although the PV panels and storage units may be distributed or centralized, a central control unit is in charge of energy management. Any BS can exploit the energy produced by any PV panel and stored in any battery unit, so that a wider flexibility and a higher efficiency can be guaranteed in terms of RE usage. The central controller also operates as radio resource manager to properly perform operations such as activating or deactivating the BSs and redistributing load among BSs.

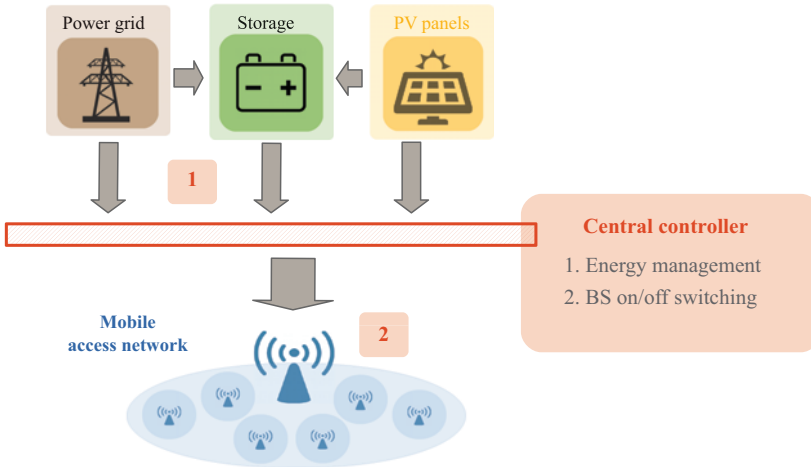


Figure 5.5 Scenario 1: urban environment and reliable power grid

5.3.2 Off-grid or on-grid BSs with unreliable power grid

Emerging countries represent a huge market where MNOs are willing to invest to deploy adequate mobile access networks to satisfy the traffic demand that is currently growing exponentially. Nevertheless, the electric power infrastructure is currently insufficient for the demand and its expansion cannot keep the pace with its staggering increase. The power grid experiences frequent power outages that make the electric network unreliable.

In addition, more and more often MNOs need to provide Internet access in remote regions, where the electric grid is completely unavailable. The use of diesel generators to make up for the lack of electric supply, besides not being a green solution, leads to huge OPEX, due to the high cost for fuel transportation and storage. In both cases, the implementation of renewable powered BSs can effectively address the need for energy self-sufficiency. However, a careful dimensioning of the RE generation system is essential, since the requirements to make the mobile network self-sustainable are clearly more difficult to meet. A crucial role is played by a properly size energy storage, whose capacity should allow to accommodate an energy amount suitable for compensating the RE production intermittence. A full level of energy self-sufficiency may be a hard target to be achieved by using RESs only. A minimum level of quality of service (QoS) may therefore be defined, envisioning the possibility of allowing some coverage interruptions or temporary suspensions of some types of services, i.e., data service.

As shown in Figure 5.6, as a case study, we consider a single off-grid BS providing coverage on a remote region that is powered only by a set of PV panels and is equipped with a set of battery units.

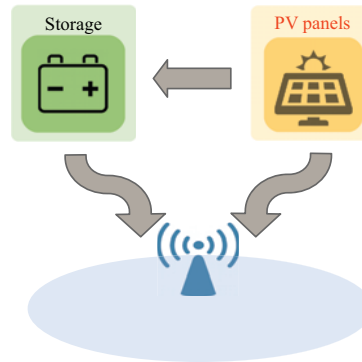


Figure 5.6 Scenario 2: off-grid BS in emerging country or emergency situation

A similar scenario, with similar challenges but slightly different constraints, occurs in emergency situations in which power supply becomes unavailable or unreliable due to the effect of the critical event.

5.3.3 Green mobile networks in the smart grid

With the shift toward the SG paradigm, the presence of several distributed energy producers is replacing the traditional approach that envisioned a single energy supplier serving many distributed consumers. The risk of a mismatch between the user energy demand and the electric power supply grows and this calls for techniques to improve or induce a better balance between energy consumption and provisioning.

SG operators (SGOs) try to induce users to shift their consumption from peak to low-demand periods, so as to flatten the energy demand pattern peaks and to dynamically adapt the energy user demand to the actual available supply. To this extent, SGOs are extensively deploying demand respond (DR) strategies. Customers participating in DR programs are periodically requested by the SGO to increase or decrease their consumption, on the basis of the current supply availability. Whenever the users satisfy the requests from the SG, they receive some incentives, either in terms of an additional monetary rewards or energy price reductions.

In this framework, MNOs can play a significant role. With the increasing mobile network densification expected in the next years, there may be a huge margin for MNOs to modulate their energy consumption from the electric grid to satisfy the SGO requests, thus obtaining a considerable reduction of the electric bill and, ultimately, contributing to improve the reliability, stability, and power quality of the SG. The introduction of RESs can help one to reduce the power drawn from the grid when the SGO requests its users to decrease their consumption. In addition, by coupling RE generators with some battery units, the BSs can be powered whenever the SGO requests to decrease consumption, even if no RE is produced in that moment, and extra amounts of energy can be drawn from the grid, in case an increase of the consumption

is requested. Finally, the participation in a DR program allows any extra produced energy that is neither used nor stored to be sold back to the SGO instead of being wasted, thus mitigating the dimensioning discrepancy issue between cold and warm seasons.

The third scenario that we consider is a portion of an on-grid renewable powered mobile access network similar to the one described in the first scenario. In this case, as depicted in Figure 5.7, the SGO implements a DR policy to enforce changes of the energy consumption pattern of its customers. The MNO that participates in such a program, applies, through a central control unit, an energy management strategy to properly interact with the SG and accomplish its requests. As in the first scenario, the central control unit also operates as radio resource manager to activate or deactivate the BSs and redistribute the load among BSs. Furthermore, the controller is also in charge of shifting some traffic toward close-by access points (APs), in case some Wi-Fi offloading (WO) techniques are implemented to temporarily reduce the mobile network load.

5.4 Challenges, critical issues, and possible solutions

We now present some among the main challenges that can be encountered when RESs are employed to power mobile networks, with reference to the scenarios already introduced in Section 5.3. Furthermore, we describe possible solutions that can be adopted by MNOs to tackle the detailed issues. The performance and effectiveness of the proposed solutions in achieving the MNO goals in the various scenarios will be investigated in the next section.

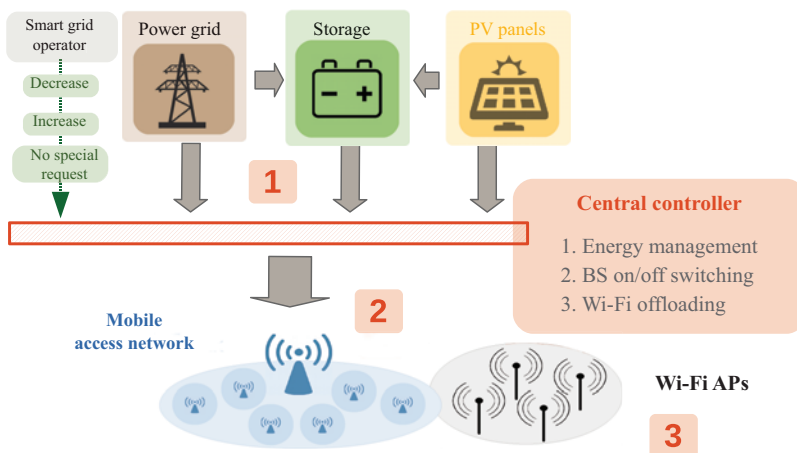


Figure 5.7 Scenario 3: demand response framework

5.4.1 PV system dimensioning

The proper sizing of PV panels and storage in a renewable powered mobile network represents a critical task under several aspects. First, feasibility issues impose strict constraints in the RE generation system deployment. The efficiency of PV modules currently available on the marketplace falls typically below 20%, requiring a surface occupancy of about 5 m^2 per kWp of capacity. Hence, in an urban environment, such as the one of the scenarios 1 and 3 previously presented, physical constraints rather than high CAPEX represent a remarkable limitation to the free deployment of RE generation power.

In addition, operational cost can be highly affected by the need to periodically replace the battery units. Whereas the lifetime of a PV panel from installation can usually last up to 25 years, the lead-acid batteries, which are the type of storage most commonly employed in RE generation systems, need to be replaced after a few years of operation. The battery lifetime depends on the number of charging/discharging cycles undergone during operation, with the need for battery replacement after an average of about 500 cycles under a maximum depth of discharge of 70% [23]. For fixed PV panel capacity, a small set of lead-acid battery undergoes a higher number of charging/discharging cycles with respect to a large battery unit operating in the same period. This means that a battery with larger capacity, despite showing higher CAPEX, will likely last longer than a smaller battery, thus leading to lower OPEX due to less frequent replacements required.

A proper planning and deployment of RE-powered mobile networks strongly depends on the presence and reliability of the electric power grid. On the one hand, the main focus of RES usage in on-grid scenarios (scenarios 1 and 3) is reducing operational costs, with physical constraints limiting the maximum expansion of PV panels. On the other hand, in the off-grid case (scenario 2), the efforts aim at achieving the highest possible level of self-sustainability, trading it off with QoS constraints. In particular, QoS constraints should be relaxed to allow a feasible deployment in terms of CAPEX, still limiting the QoS impairment to reasonable levels.

The latitude and local weather conditions of the RE generation system installation strongly affect the correct sizing of PV panels and batteries. Indeed, high seasonal, inter-day and intraday variability of RE production makes it harder to determine a proper system dimension that is suitable to satisfy the mobile network demand across the whole year. This aspect clearly constitutes a remarkable and more critical obstacle in sizing the RE powering system for off-grid BSs, considering the required trade-off between the RE overproduction during the warm season and the RE system underdimensioning during the cold season. Conversely, in on-grid scenarios, the extra need for energy that is not satisfied by the local RE production can be covered by the brown energy, i.e., the energy drawn from the electric grid, that is assumed to be derived from not RESs. In addition, in a DR framework, in the case of an RE generation system that results overdimensioned during the warm season, extra amounts of RE can be sold back to the SGO instead of being wasted, providing additional revenues.

The complexity of the dimensioning procedure requires the deployment of simulation and stochastic models that take into account all the aforementioned critical

aspects. In the literature, some studies can be found investigating the performance of renewable powered mobile networks under variable combinations of PV panels and battery sizing, either via simulation or stochastic models [21,24]. Few papers explicitly analyze the system sizing problem and thoroughly examine the impact of variable system dimensions on the battery depletion probability, system costs, and network performance, by means of simulations, stochastic models, and optimization approaches [19,25–30].

5.4.2 System operation and management

An overview of the system operation and management challenges in solar-powered mobile networks is now presented.

- **Energy management.** In a renewable powered mobile network, a proper energy management must be envisioned to efficiently use the available energy resources. In our investigation, a simple energy management policy is assumed, based on which the solar energy is first used to power the BS, whereas only the extra amounts of RE that are not immediately used for powering the BS are harvested into the storage to be used later on to power the BS in the case of RE unavailability. This behavior follows a principle that is similar to the harvest-use-store paradigm that is adopted in wireless networks [31,32]. Following this principle, charging/discharging losses can be minimized with respect to the case in which the battery is first charged and then the energy is drawn from the storage to satisfy the BS demand. Indeed, charging/discharging losses are not negligible at all, amounting up to 25% [33]. When no RE is currently being generated or if its production is not sufficient to satisfy the BS demand, energy can be drawn from the storage. Only if the battery is discharged, the required energy is taken from the power grid, if available. In our system, we assume that a single centralized PV panel installation produces RE that can be distributed to operate the various BSs. However, scaling up to wider scenarios that include larger portions of mobile access networks, distributed PV panel installations should be envisioned. Even in this case, RE could be shared among BSs by means of energy sharing techniques based on energy transmission mechanisms; in this way, it is possible to take better advantages of locally produced RE and further limit the energy procurement from the SG [34]. A proper energy management should guarantee that the battery charge level never falls below a minimum threshold, corresponding to a maximum depth of discharge of 70%, that allows to prevent charging efficiency impairment [35] and to limit battery aging [36] and capacity loss [37].
- **Radio resource management.** The benefits of the RE use in mobile networks can be enhanced by the synergic effects provided by the implementation of Resource on Demand (RoD) techniques. As it can be observed from Figure 5.4, the typical LTE BS power consumption is very little traffic proportional and even during periods of negligible traffic, up to 80% of the maximum energy consumption is consumed, with huge energy wastage and useless cost for MNOs. Furthermore, busy hour Internet traffic is growing more rapidly than average Internet traffic [4]. Hence, considering that mobile networks are typically dimensioned based on peak

demand, the over-provisioning issue and the related energy wastage are bound to become even more evident. RoD strategies are widely adopted to make the mobile network consumption more proportional to the actual traffic demand that may vary a lot over time. In particular, BS sleep mode approaches can exploit the energy consumption decrease by means of gradual deactivation of radio components, to enter progressively lower power levels. Simple on/off switching policies are based on this principle and can effectively reduce the power consumption, although MNOs can take advantage of energy saving approaches based on new multi-sleep-mode power models featured by next-generation BSs. Several studies are devoted to the application of these techniques to new-generation mobile networks, also considering renewable powered BSs [28,38–40].

- **Wi-Fi offloading.** WO techniques can also be introduced in the system. WO consists in transferring a portion of the mobile traffic from the BSs to some nearby Wi-Fi APs. This technique is commonly adopted to relieve mobile access networks from a part of their traffic load during peak hours in heterogeneous networks. WO alone may not lead to remarkable reduction in the consumption, due to the limited load proportionality of energy consumption. However, when applied in a scenario where the BSs can also be switched on and off, WO allows one to further decrease the number of active BSs, hence to further reduce the system consumption. Policies are required for managing the timely activation/deactivation of radio resources and to shift traffic from switched off BSs to active BSs or to neighboring APs.
- **Traffic prediction.** Finally, in renewable powered network scenarios, especially when RoD strategies are envisioned, the prediction of RE production, traffic, and energy prices becomes crucial to deploy energy management algorithms capable of following and adapting to the system dynamics, with the purpose of optimizing the reduction of brown energy as well as the financial benefits, still maintaining an acceptable QoS. To this extent, predictive approaches based on machine learning (ML) techniques represent a promising solution to further improve the overall system performance. Several different types of management strategies can be implemented that integrate the prediction of various system parameters. For instance, RoD policies can take decisions based on the short-term or medium-term prediction of the traffic demand, the RE production, the battery charge level.

5.4.3 Interaction with the smart grid

In an SG context, the use of RES to power the mobile network can be effectively controlled by properly designed energy management strategies to adapt the operational decisions taken on the network to follow the dynamic variations of the energy price or to dynamically satisfy the periodical requests from the SG in a DR framework, as shown by few studies from the literature confirming the potentiality of similar scenarios [41–43]. For example, when the request from the SGO is of decreasing the consumption, the mobile network can use the locally produced RE or the energy previously stored in the battery. In case the SGO requests its customers to increase

the consumption, some extra amount of energy can be drawn from the grid and stored into the battery for future usage. The interaction between MNOs and SGO can be further improved by exploiting the combined application of RoD approaches and WO techniques [44,45]. When RoD and/or WO strategies are applied, the operational management process must integrate proper decision algorithms allowing to timely activate/deactivate BSs and to move traffic among BSs or close-by APs, so that these strategies can effectively be exploited to enhance the accomplishment of SG requests when it requires to decrease the energy demand.

In addition, in a smart city environment, new business models are emerging that may further boost the penetration of RE in mobile networks, contemplating the appearance of multiservice providers and opening the way to brand new scenarios. In a similar context, MNOs could exploit the cooperation with Household Consumers (HCs) to enhance the interaction with the SG and reduce the energy bill. The MNO performs some kind of agreement with HCs that participate to the same DR program and that are willing to contribute to the achievement of the MNO goals, for instance by modulating their own energy demand to the advantage of the mobile network. In return, the HCs willing to join the cooperation program receive some benefits from the MNO in terms of discounts on the pricing of communication or free additional benefits (e.g., higher speed Internet connection) when stipulating the contract for the Internet access provisioning.

The specific scenario is presented in Figure 5.8. HCs accept to help the MNO in accomplishing SG requests when a reduction of the energy consumption is requested, in case neither RE is currently available nor any BSs can be switched off due to high traffic load.

Another example of a future possible scenario could be the cooperation between MNOs with electrical vehicles that could operate both as moving mobile antennas and backup batteries for the cellular network.

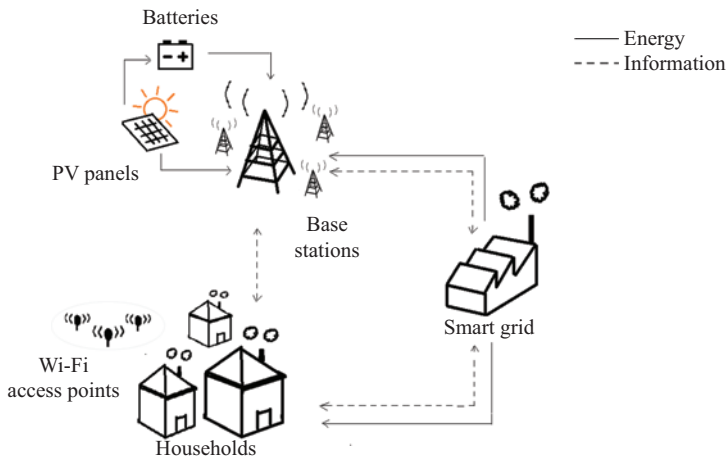


Figure 5.8 Household user cooperation scenario

Finally, a very important challenge is cybersecurity of the integrated SG and communication network infrastructures. With the diffusion of the SG concept as a new paradigm in power networks, new vulnerabilities arise in the electric system in urban scenarios, making it more prone to new types of cyber attacks that may significantly and abruptly impair the power grid reliability.

5.5 Some case studies

We now show more in detail the actual performance and effectiveness of some of the solutions presented in previous section, considering the different scenarios illustrated in Section 5.3.

5.5.1 Photovoltaic system dimensioning

Starting from the objective of reducing costs, Figure 5.9 shows the total capital and OPEX for a renewable powered BS in the city of Turin, assuming various levels of independence from the grid that result inversely proportional to the parameter *PT*. This parameter denotes the *percentage of time* in which the RE production is sufficient to satisfy the BS demand, without the need to draw energy from the grid. As comparison, the graph reports the case of an urban environment (pink line), corresponding to scenarios 1 and 3, where the BS is only powered by the grid (corresponding to scenario 1), and the case of an off-grid BS (green line, with circles), where the grid is not available at all and the power supply is only provided by a diesel generator (scenario 2). Whereas a full self-sustainability can be obtained at the price of high CAPEX, a hybrid power supply consents to remarkably reduce the initial investment, still obtaining benefits in terms of OPEX. In an urban scenario, a hybrid system

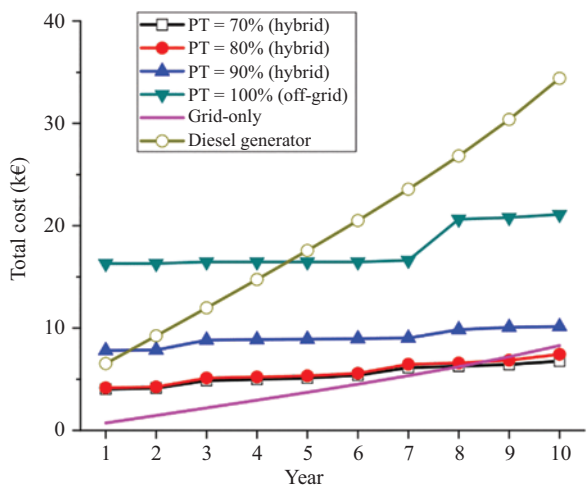


Figure 5.9 Total cost (CAPEX + OPEX) for different values of the parameter PT [46]

dimensioned to allow a $PT = 80\%$ costs about 75% less than a pure solar system, becoming even cheaper than the grid-only case after 8 years of operation. The battery units need to be replaced more frequently than the off-grid case, as shown by the small steps in the cost curves. In an emerging country scenario, the initial investment is necessarily high. However, the total cost becomes more convenient compared to a diesel generator in less than 5 years of operation.

These results indicate that, from an economical point of view, RE-based power supply is convenient in several scenarios, and the convenience increases if it is possible to rely on some backup power system and full self-sustainability is not required.

The solar radiation is highly location dependent, hence the latitude and the weather conditions of the location where the renewable powered mobile network is installed show a considerable impact on the proper system dimensioning [46]. Figure 5.10 reports the simulated battery charge level of a 12 kWh storage unit in two different urban scenarios, Turin and Aswan, over a year, assuming a PV panel capacity that guarantees a $PT = 70\%$ and a single BS. The selected locations differ for the solar radiation profiles, with the radiation in Turin featuring a much higher intraday and over season variability. The required PV panel size results 13% larger in Turin, where the average RE production is lower and shows a higher variability. The red portion over 100% represents the loss of produced RE that cannot be stored, whereas the red zone below zero corresponds to the energy drawn from the grid to operate the BS when the battery is depleted. In Aswan, although deeper battery charge/discharge cycles can be observed with respect to Turin, we note a relatively more stable depth of discharge. Furthermore in Turin, due to the remarkable solar radiation variability from day to day and over seasons, the system tends to result significantly overdimensioned in Summer, leading to consistent energy waste if no energy sell-back is possible, whereas in the rest of the year some energy must be taken from the grid in any season. Conversely, in Aswan, the lower radiation variability makes the dimensioning of the RE system better matched to the BS power demand during the whole year, with limited wastage, and the need for brown energy is mainly observed in Winter.

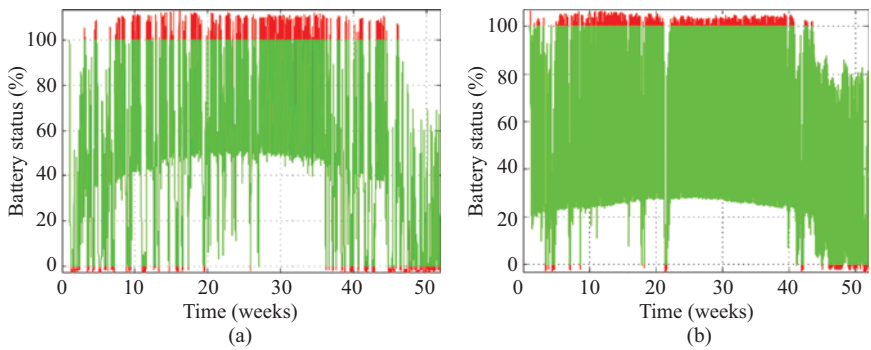


Figure 5.10 Charge level of a 12 kWh battery, assuming a PV panel capacity of 4.3 kWp in Turin (a) and 3.8 kWp in the city of Aswan (b), respectively, with PT = 70% [46]

A first indication that we can derive from these results is that the dimensioning and the dynamic behavior of the PV power system strongly depend on the meteorological conditions; for example, quite different patterns of battery usage are observed in locations with different degrees of variability of energy production.

Besides geographical location, also the type of traffic pattern affects the system dimensioning. Figure 5.11(a) reports the amount of energy bought from the grid in a year, E_G, for different PV panel sizes and battery units (S_P and S_B, respectively), considering a portion of a mobile access network in two areas characterized by a residential (RA) and a business (BA) traffic. As the panel size increases, the energy from the grid reduces, especially for values of panel sizes <45 kWp. For a fix panel size, a larger battery allows to further decrease the energy bought from the grid, both in RA and BA. Moreover, also the resource management influences the dimensioning. Figure 5.11(b) shows that under the application of a RoD strategy, a target threshold of maximum $E_G = 50$ kWh can be achieved with panels as small as 27 kWp, more than 30% smaller than without any RoD. Furthermore, the benefit of the RoD application is more evident in the RA, since the peaks in the traffic patterns result coupled with the peaks observed in the RE production profiles. In a business area, most of the traffic is generated during working hours; hence, the renewable power can be immediately used to run the BS.

In a scenario where the grid is not available, the dimensioning of the RE system becomes very critical, considering the feasibility constraints and QoS issues raised by the possible impairment of the continuity of service. Furthermore, the location-dependent RE production variability may have a remarkable impact on the proper system sizing. Figure 5.12(a) shows the battery occupancy, B_O, obtained from

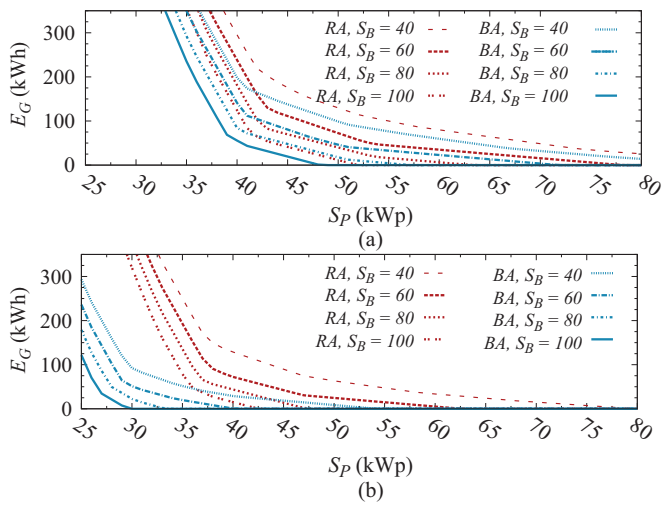


Figure 5.11 Yearly grid energy, E_G, versus S_P in residential and business area (RA and BA), for different values of the energy storage capacity [25]: (a) no RoD and (b) under RoD

a stochastic model for a single BS powered by RE, for various levels of average renewable power generation, denoted as RE_P. RE_P is represented as a random variable with uniform distribution, considering different values of its variance. As it can be observed from the figure, the impact of the variance is very high, especially for intermediate levels of RE generation. For fix mean RE production, a low variance allows to achieve higher mean values of battery charge; hence, smaller panels can be used in this case. However, when the RE production is not sufficient for satisfying the BS demand ($RE_P \leq 250$ W), the battery tends to be depleted more easily if the variance is higher. Figure 5.12(b) highlights the impact of the variance (v_R) on the amount of energy that should be provided by an alternative energy source, like a diesel generator, denoted G, assuming different values of RE production, RE_P. As the variance increases, the value of G becomes higher. The ascent results steeper for low values of mean RE production, whereas it decreases up to becoming negligible as the RE production level increases.

To investigate more in detail, the effect of the variance, we consider five different day types based on the average energy that is produced during a day and we evaluate the battery charging and discharging process during these different day types. Figure 5.13 depicts the results for the average hourly battery charge of a solar-powered BS in the various day types assuming a storage capacity of 25 kWh. During the nighttime, the BS drains the energy stored in the battery during the peak production hours. At the end of the day, the balance can be positive (for good weather days, e.g., type 5) or negative (for bad weather days, e.g., type 1). Observe how different the charge levels are in different day types. Clearly, by increasing the panel capacity, larger amounts of energy are stored in the battery. The most critical days are those with bad weather (type 1) and it is necessary to double the PV panel size (from 20 to 40 kWp) to observe a relevant effect on the battery charge level.

An important contribution to the feasibility of these systems, to the reduction of the PV system size and cost, is provided by the shift toward new technologies for the BSs. The consumption of next-generation BSs shows lower average power levels and it results to be more load proportional. Figure 5.14 compares the battery charge level and the probability of battery depletion of a renewable powered BS under

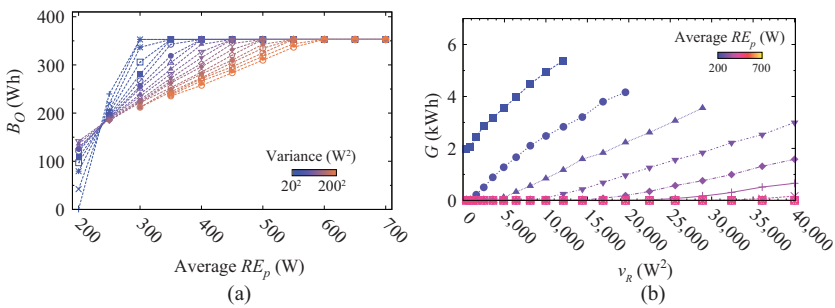


Figure 5.12 (a) Battery occupancy, B_O, and (b) energy demand from a diesel generator, G, under uniform distribution (UD) of RE_p [20]

two BS consumption models: the EARTH model (denoted as *EarthM* in the figure) [3], corresponding to the currently most widespread BS technology, and the next-generation BS consumption model (2020M) [47]. Both the residential and the business areas are considered.

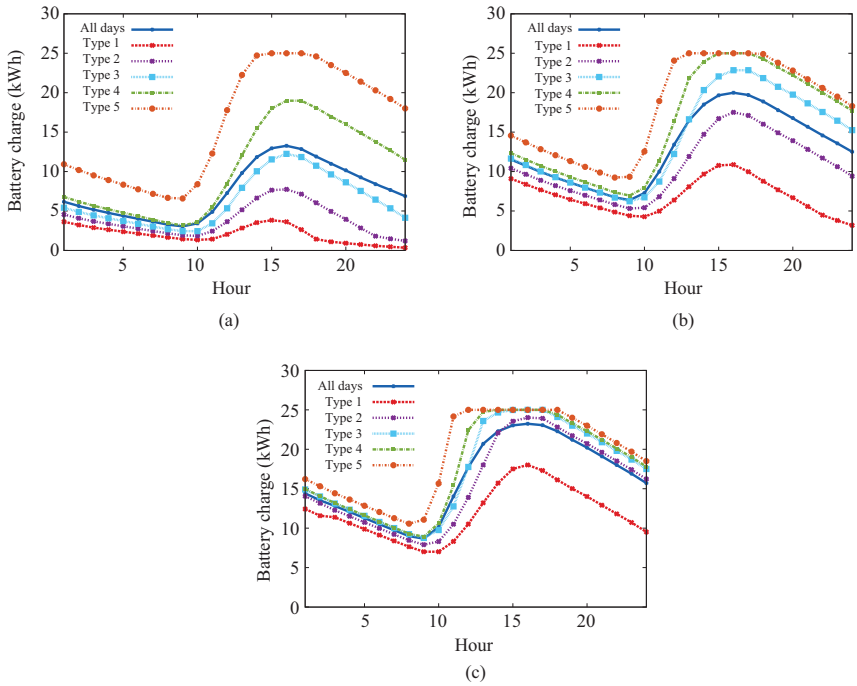


Figure 5.13 Average hourly battery charge for a macro BS versus time, with five day types, PV panel sizes (a) 20, (b) 30, and (c) 40 kWp, battery capacity 25 kWh [22]

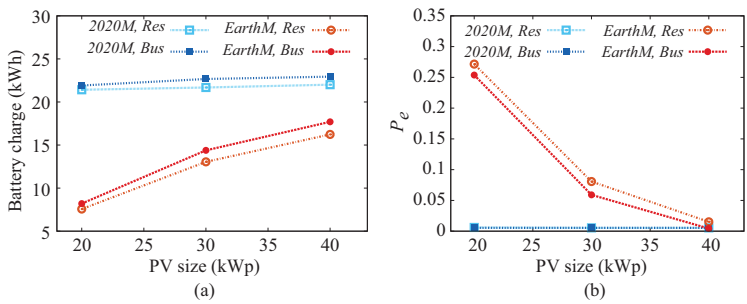


Figure 5.14 (a) Average battery charge level and (b) empty battery probability (P_e) versus PV panel size, for 2020 and Earth models (2020M and Earth M, respectively), in residential and business areas [22]

The new BSs (2020 model) yield to better performance than the traditional ones, represented through the EARTH model, guaranteeing a battery charge level that is more than three times higher even with a panel as small as 20 kWp. For the new BSs, the panel size required to achieve an almost zero probability of empty battery (20 kWp) is half the capacity needed for the traditional model. This confirms that the new-generation BSs make the use of RE less expensive, leading to the possible increase of the diffusion of renewable powered BSs, in both those regions where the grid is not available or unreliable, as well as those with a reliable power infrastructure. In addition, more efficient PV modules will be brought into the market in the coming years, further reinforcing this trend.

5.5.2 System operation and management

In previous section, we have observed that power system dimensioning is a difficult task. Hence, a proper joint management of the RE usage and of the mobile network operation can be a fundamental help to improve the system feasibility and to reduce cost. Consider the use of RoD strategies. Figure 5.15 shows the values of the mean yearly cost (CAPEX + OPEX) averaged over a period of 25 years for a cluster of solar-powered BSs, also including the cost for battery replacement, together with the requested size of the PV panel, S_P, for different target levels of energy self-sufficiency, denoted as $f_{G_{\max}}$: the bottom plots refer to the cases in which RoD is adopted, while top plots correspond to cases in which no RoD is used. For example, the case $f_{G_{\max}} = 0$ corresponds to a design target that requires that no energy should be bought from the grid; the case $f_{G_{\max}} = 1\%$ allows that up to 1% of the times energy can be purchased

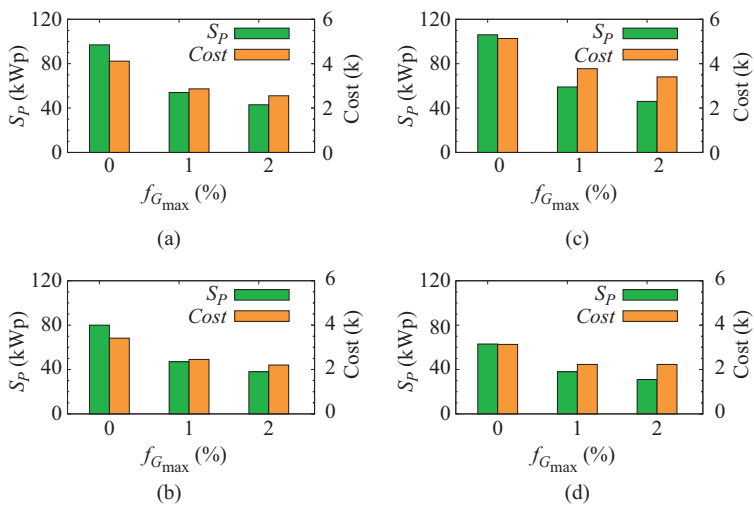


Figure 5.15 *S_P and c with no RoD strategy and under RoD strategy, in RA and BA, with different constraint on f_G [25]: (a) RA, $S_B = 40$, no RoD, (b) RA, $S_B = 40$, under RoD, (c) BA, $S_B = 30$, no RoD, and (d) BA, $S_B = 30$, under RoD*

from the grid. Note that, despite a similar energy demand in the RA and in the BA when all the BSs are always kept active, RoD in the BA allows to save twice (26%) the energy saved in the RA (11%). Assuming a system totally independent from the grid, as in scenario 2, the RoD strategy reduces cost by 17% in RA, against a 39% decrease obtained in BA, due to a larger decrease in S_P. By slightly relaxing the constraint on self-sustainability up to 1%, as for scenarios 1 and 3, RoD never decreases cost by more than 14% in RA, against a 41% reduction in BA, due to the smaller required S_P. Again, this is partly explained by the coupling between the radiation patterns and the traffic profiles in the BA. Considering that a surface of about 5 m^2 is required per kWp, a similar reduction of the panel size allows to more easily address the strict physical constraints that prevent the installation of large PV systems occupying wide areas in urban environment. Conversely, it is not very effective to further lessen the constraint on $f_{G_{\max}}$, both in terms of PV panel size and cost.

In order to operate the network dynamically, for example, using RoD strategies, accurate predictions of the traffic are required. To this extent, ML techniques represent an attracting solution, since they can be applied to obtain reliable real-time forecasts of the user demand. In particular, several papers are available in the literature investigating the potentiality of *neural networks* for load prediction in various contexts, from power networks [49–52], to heat systems [53], to Cloud Data Centers [54]. To investigate the effectiveness of similar approaches, we report in Figure 5.16 the energy saving and the traffic loss observed under application of RoD coupled with various traffic prediction algorithms. Eight portions of a RAN are considered, each featuring a different traffic profile. Traffic data are derived from real traces provided by an Italian MNO from a wide area, including the city of Milan, for a duration of 2 months in 2015. The eight areas were selected for being quite representative of the various zones that coexist in an urban environment: (1) the train station area, featuring intense activity especially at the beginning and at the end of the working day (train station); (2) a district hosting big events, fairs, and exhibitions, which might last for a

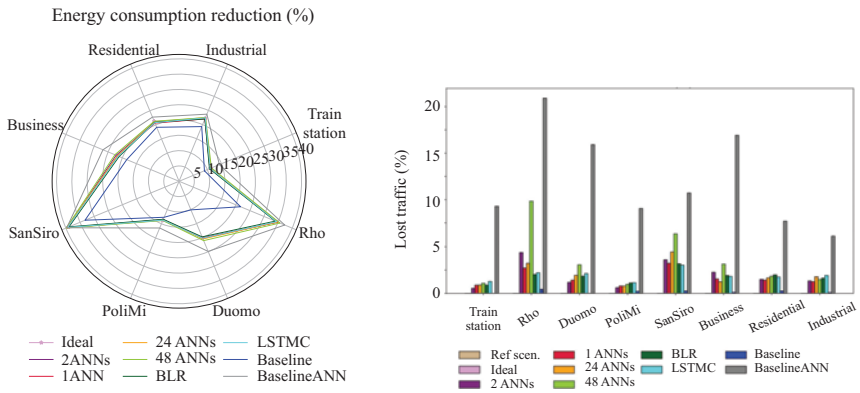


Figure 5.16 Comparison of the effectiveness of prediction techniques: (a) energy consumption reduction, (b) percentage of lost traffic [48]

few days (Rho Fiera); (3) a turistic area, with high activity levels during several hours of the day (Duomo area); (4) the university area, hosting a huge campus (PoliMi); (5) the soccer stadium area, where the activity is quite bursty and variable, depending on the held events (SanSiro); (6) a business neighborhood, with traffic peaks in the central hours of the day (business); (7) a residential area, where the traffic demand raises in the evening hours (residential); and (8) an industrial zone, with business traffic profile (industrial).

The RoD decisions are taken based on the RE production forecast and on the traffic predictions. RE production is forecast by applying a linear regression-based method on real traces that are obtained from the tool PVWatts. Regarding the traffic prediction, various ML approaches are evaluated. The training phase is performed on historical data over a period of 47 days, whereas the remaining 14 days of the dataset are used for the run-time phase. The ML techniques proposed for traffic prediction include the following:

1. A method based on Block Linear Regression (BLR) [55], in which the predictor reflects the daily and weekly periodicity of mobile traffic, is formulated using the linear regression. A single model is constructed to forecast the traffic of the entire BS cluster.
2. Four different methods based on artificial neural networks (ANNs) [56]. These methods employ as input five different sample values of the traffic selected among those registered in the previous hour and in the previous 2 days at the same daytime. For each BS, a single ANN can be used, without distinguishing the daytime (1 ANN), or 24 ANNs are used, one per each hour of the day (24 ANNs). Furthermore, 2 ANNs can be separately trained for each BS, 1 for the week-day traffic pattern and the other for the week-end pattern (2 ANNs). Finally, 48 ANNs can be separately trained, 1 per each of the 24 h of the week-day traffic pattern and the remaining 24 for each of the 24 h of the week-end pattern.
3. A long short-term memory cell (LSTMC) algorithm that exploits a recurrent neural network trained for each BS [57].
4. A *Baseline algorithm* (Baseline), a simple method against which the performance of the tested ML approaches is compared. It is based on the computation of the average traffic demand, either for week-days or for week-ends, for each hour using the traffic samples belonging to the corresponding training dataset and it is used as the predictor [48].
5. A *Baseline with ANN* (BaselineANN) algorithm that is similar to the previous one, but an ANN is adopted to perform the traffic prediction [48]. This algorithm takes as inputs the difference between the estimated traffic value (according to Baseline method) and the actual values of traffic demand at five different time instants, selected among the previous hour and in the previous 2 days, at the same time.

The energy saving reported in Figure 5.16(a) is computed with respect to the reference scenario, in which no action is taken to reduce energy consumption, i.e., RoD is not used. In Figure 5.16(b), the ideal case assumes a perfect knowledge of future traffic demand. Results show that up to 40% of energy can be saved in all the areas, with an energy consumption level that is very close to the ideal case with almost any

of the algorithms, which turn out to be rather similar in terms of effectiveness. Energy saving is obtained at the cost of QoS deterioration. Except for one of the algorithms (BaselineANN), the other ML approaches show a similar QoS deterioration in each zone, usually lower than 5% besides the traffic area. BLR, LSTMC, and n-ANNs algorithms can hence be considered equivalent. Even if one of them may provide a more energy-consuming configuration at some point during the run-time with respect to another method, the same algorithm will provide a less energy-consuming one during other periods, so that the differences among ML techniques are marginal.

It is interesting to note that a good algorithm does not necessarily make an accurate traffic prediction at any time, but it predicts the correct traffic demand right when the traffic level is around the threshold according to which the decision of switching on or off a BS is taken. Figure 5.17 reports the traffic predictions under the different algorithms, compared with the actual traffic profile (black line with stars). Some algorithms are not able to predict the traffic peaks or the periods of low traffic. However, this does not significantly affect the performance of the ML techniques that provide similar energy savings, since the prediction tends to be accurate around those values close to the switching threshold, rather than at any traffic level. This is the case, for example, of the baseline: its predictions around the peak traffic are pretty bad, but they are good for values of the traffic corresponding to RoD decisions.

These results confirm that RoD strategies are feasible, but several challenges, including correct estimation of the network condition at any time, need to be tackled.

5.5.3 Interaction with the smart grid

The possibility of interacting with the SG opens up new opportunities for MNOs in terms of financial benefits. Proper energy management strategies need to be implemented to achieve the desired economic goals. Moreover, considering that the mobile

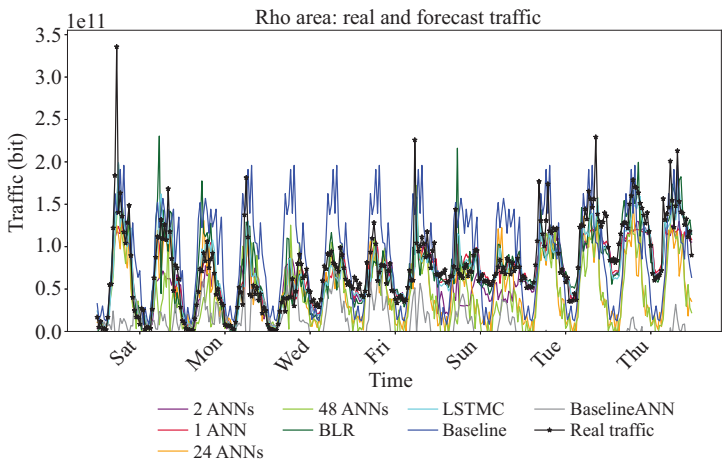


Figure 5.17 Forecast of the traffic demand under various algorithms [48]

traffic demand cannot easily be reduced, the introduction of RoD and WO approaches provides wider margin to accomplish the requests of the SG when it requests its customers to reduce the consumption. Figure 5.18 shows the cost saving that can be obtained from the combination of RE usage and resource management strategies. Realistic patterns for the SG request occurrences are considered, and the energy management algorithm described in Section 5.4.3 is implemented.

In the case in which no RE is locally produced (orange bars), the application of WO and RoD, either alone or combined, allows to achieve up to 15.2% cost saving. However, WO alone results more effective than RoD alone, providing almost double cost saving. In addition, two cases are presented in which an RE generator is installed, either with intermediate (denoted as I: $S_{PV} = 5$ kWp, $B = 5$—blue bars) or large (denoted as L: $S_{PV} = 10$ kWp, $B = 15$—green bars) size. A local RE supply leads to significantly higher cost savings. The energy bill can be completely nullified even with intermediate RE system sizing, whereas positive revenues provided by the SG to the MNO are obtained increasing the system size.

It is interesting to note that, unlike the case when no RE is produced locally, the benefits provided by the application of WO in the presence of RE are lower with

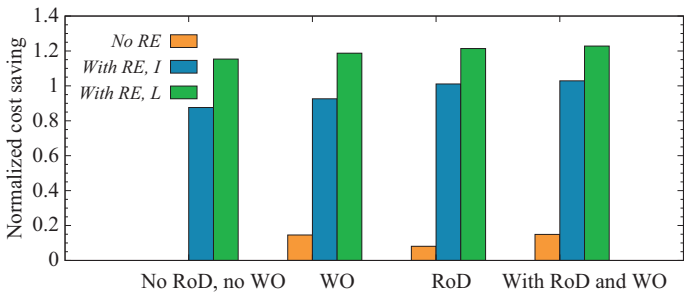


Figure 5.18 Cost saving under different RE system size [44]

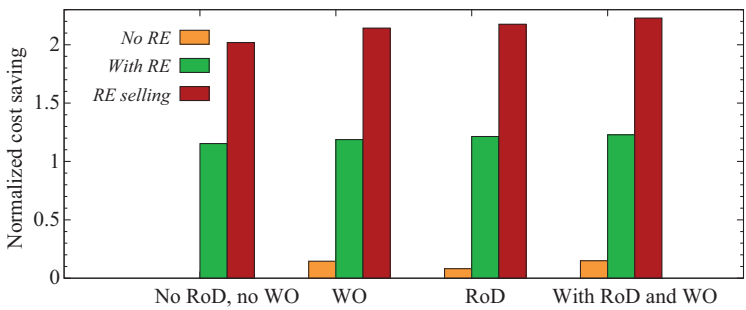


Figure 5.19 Cost saving under three scenarios: (1) No RE—no RE is produced; (2) With RE—under RE production ($S_{PV} = 10$ kWp, $B = 15$); (3) RE selling—assuming that the extra amounts of produced RE are sold back to the grid [44]

respect to the case of RoD alone. In particular, positive revenues are increased by up to 40% under RoD alone, against 27% when only WO is applied. Combining RoD and WO allows to increase revenues by up to 53% when WO and RoD are combined. Further financial benefits can be obtained by envisioning the possibility of selling back to the SGO extra amounts of RE that are not immediately used and, in case they cannot be harvested in the storage, are usually wasted. Figure 5.19 compares the cost saving obtained in a similar case (red bars) with the cases in which no RE is locally produced (orange bars) or, if it is produced, it cannot be sold back to the SG (green bars). The price per each energy unit that is sold to the grid is assumed to be half the price due for each energy unit bought by the MNO from the grid. Under any combination of resource management strategy, huge cost reduction can be obtained, amounting to up to more than twice the energy bill registered in baseline conditions in case no RE is present and no WO or RoD strategies are applied. In relative terms, WO alone or RoD alone is equally effective in raising the cost saving, since they both provide savings that are about 1.8-fold higher than those assured when no energy is sold back to the grid. When WO is applied in conjunction with RoD, up to 223% cost saving can be achieved. These revenues may contribute to compensate the higher CAPEX faced for the initial installation of an RE generation system.

It should be observed that a good energy management strategy does not operate by reducing the total grid consumption, but by timely increasing or decreasing the grid consumption exactly when required by the SG. Figure 5.20 compares the amount of energy bought from the grid, denoted E_G, and the operational cost under various combinations of resource management strategies, either with or without a local RE system. From the graphs, it appears evident the key role of a timely reaction to the SG grid requests. Despite E_G is reduced by at most 48.3% when RE is locally produced, the operational cost is not only nullified but even becomes negative, thanks to the joint action of the WO strategy and the RE generation, along with storage, in improving the interaction with the SG, providing huge rewards that compensate for energy cost and penalties.

Finally, we present a case of cooperation between MNO and household customers to show the potentiality of new business models that may arise in a smart city scenario. Figure 5.21 depicts the yearly energy demand from the grid and the yearly energy cost assuming that household customers may be available to postpone part of their load

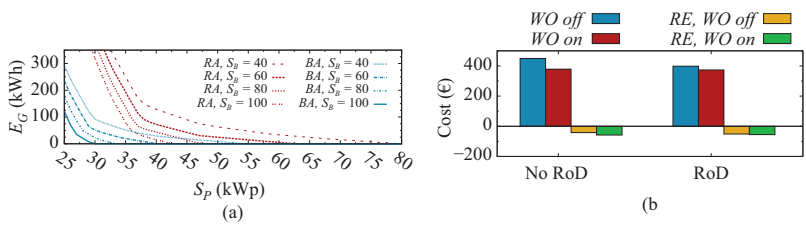


Figure 5.20 (a) Energy drawn from the grid, E_G, and (b) OPEX cost when no RE is produced and under RE production ($S_{PV} = 10$ kWp, $B = 10$) [25,44]

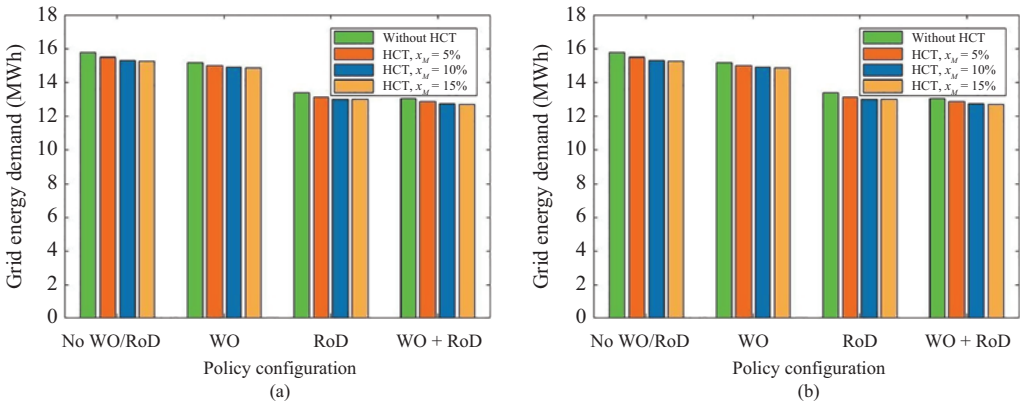


Figure 5.21 (a) Yearly energy demand from the grid and (b) yearly energy cost under different policy configurations, for different values of x_M [45]

on behalf of the mobile network (HC Trade) and/or to offer WO service. RoD may be applied as well on the MNO side. The green bars represent the case in which HCs do not agree to postpone any load to the MNO benefit. The other bars correspond to different levels of contribution by the HCs. The parameter x_M denotes the fraction of load that the HCs agree to postpone on behalf of the MNO in the case of request of decreasing the consumption from the SG.

The effect of increasing the values of x_M on the reduction of the energy bought from the grid is negligible, whereas it can give a relevant contribution to the MNO cost saving. By increasing x_M from 5% to 10%, the cost saving is almost doubled only when HC cooperation is applied alone. Higher values of this fraction, besides being unlikely, do not provide significant further benefit in terms of energy bill reduction, resulting 11%.

5.6 Conclusion

In this chapter, we focused on the introduction of RESs to power RANs. Motivated by the increase of the operational costs to power the networks, the need to reduce carbon footprint, and the interest in making power supply more independent from the power grid, the introduction of renewable-based power supply is becoming more and more attractive.

There are, however, a number of challenges and some possible solutions. MNOs must tackle various critical issues to optimize the potentiality and benefits of similar energy supply, also depending on the specific scenario where the renewable powered networks are deployed.

First, the RE system dimensioning is a very hard task, made particularly critical when the BSs should be completely independent from the electric grid. A slight

relaxation on the self-sustainability constraints combined with RoD approaches allows to remarkably reduce the size of the required PV panels, hence improving the system feasibility by reducing the occupied area and CAPEX by up to more than 40%. In the future smart city scenario, dominated by 5G technology and characterized by physical constraints typical of the urban environment, the feasibility of these RE generation systems will be further enhanced by more efficient PV modules brought into the market and by new-generation BSs featuring higher energy efficiency.

Second, network management and operation techniques can be tailored to improve the feasibility and effectiveness of these systems, and proper system operation and energy management policies are required to efficiently use the RE and guarantee the continuity of service. ML algorithms are proved to be effective in predicting mobile traffic load in real time, and this allows to dynamically adapt the radio resources to the actual traffic demand, trading off energy saving and QoS, and achieving energy consumption reductions of more than 40%.

Furthermore, the interaction with the SG provides relevant opportunities to MNOs in urban environments, since they become significant stakeholders in demand response frameworks. The use of RE properly combined with RoD techniques and WO approaches makes it possible to dynamically adapt the network consumption to the SG requests. MNOs can achieve significant cost reduction and even positive revenues by implementing proper energy management policies that do not necessarily operate by decreasing the total grid consumption but by timely reacting to the dynamic SG requests.

Finally, new smart city scenarios offer wider potentialities for achieving MNOs goals, by means of the cooperation with nonmobile users and the introduction of new business models.

While several issues and challenges need to be further investigated to make it possible to deploy on a massive scale renewable power supply for RANs, the direction is promising and will contribute to alleviate the concerns on cost, sustainability, and global deployment of the communication infrastructures.

References

[1] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2018-2023 White Paper; 2020. Available from: https://tinyurl.com/y2lwempd.

[2] Gati A, Salem FE, Galindo Serrano AM, *et al.* Key Technologies to Accelerate the ICT Green Evolution—An Operator's Point of View. arXiv preprint arXiv:190309627. 2019.

[3] Auer G, Blume O, Giannini V, *et al.* Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown Energy Aware Radio and neTwork tecHnologies (EARTH) Project; 2012 (2010).

[4] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022 White Paper; 2019. Available from: https://tinyurl.com/y2lwempd.

[5] Union IT. ITU Symposium on ICTs and Climate Change Hosted by CTIC, Quito, Ecuador, 8–10 July 2009: ITU Background Report. International Telecommunication Union Geneva; 2009.
[6] Nordman B and Lanzisera S. Electronics and network energy use: Status and prospects. In: 2011 IEEE International Conference on Consumer Electronics (ICCE). IEEE; 2011. pp. 245–246.
[7] Masoudi M, Khafagy MG, Conte A, *et al.* Green Mobile Networks for 5G and Beyond. IEEE Access. 2019;7:107270–107299.
[8] Hassan HAH, Nuaymi L, and Pelov A. Renewable energy in cellular networks: A survey. In: Online Conference on Green Communications (GreenCom), 2013 IEEE; 2013. pp. 1–7.
[9] Tala't M, Yu C, and Feng K. A survey: Green energy for cellular base station and small cell. In: 2017 2nd International Conference on the Applications of Information Technology in Developing Renewable Energy Processes Systems (IT-DREPS); 2017. pp. 1–4.
[10] Fletscher LA, Suárez LA, Grace D, *et al.* Energy-Aware Resource Management in Heterogeneous Cellular Networks With Hybrid Energy Sources. IEEE Transactions on Network and Service Management. 2019;16(1):279–293.
[11] Couto da Silva AP, Renga D, Meo M, *et al.* Small solar panels can drastically reduce the carbon footprint of radio access networks. In: 2019 31st International Teletraffic Congress (ITC 31); 2019. pp. 64–65.
[12] Pamuklu T and Ersoy C. Reducing the Total Cost of Ownership in Radio Access Networks by Using Renewable Energy Resources. Wireless Networks. 2018.
[13] Guo S, Zeng D, Gu L, *et al.* When Green Energy Meets Cloud Radio Access Network: Joint Optimization Towards Brown Energy Minimization. Mobile Networks and Applications. 2019;24(3):962–970.
[14] Wang L and Zhou S. Flexible Functional Split and Power Control for Energy Harvesting Cloud Radio Access Networks. IEEE Transactions on Wireless Communications. 2020;19(3):1535–1548.
[15] Jiang J, Gao L, Jin J, *et al.* Sustainability Analysis for Fog Nodes With Renewable Energy Supplies. IEEE Internet of Things Journal. 2019;6(4):6725–6735.
[16] Morgera AF and Lughi V. Frontiers of photovoltaic technology: A review. In: Clean Electrical Power (ICCEP), 2015 International Conference on; 2015. pp. 115–121.
[17] Fraunhofer ISE. Photovoltaics Report; 2015. https://tinyurl.com/gt7rpob.
[18] IEA-ETSAP, IRENA. Solar Photovoltaics: Technology Brief, Tech Rep. Energy Technology Systems Analysis Programme and International Renewable Energy Agency; 2013.
[19] Marsan MA, Bucalo G, Caro AD, *et al.* Towards zero grid electricity networking: Powering BSs with renewable energy sources. In: Communications Workshops (ICC), 2013 IEEE International Conference on; 2013. pp. 596–601.
[20] Renga D and Meo M. Modeling renewable energy production for base stations power supply. In: 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm); 2016. pp. 716–722.

[21] Renga D, Hassan HAH, Meo M, *et al.* Improving the interaction of a green mobile network with the smart grid. In: 2017 IEEE International Conference on Communications (ICC); 2017. pp. 1–7.

[22] Couto da Silva AP, Renga D, Meo M, *et al.* The Impact of Quantization on the Design of Solar Power Systems for Cellular Base Stations. IEEE Transactions on Green Communications and Networking. 2018;2(1):260–274.

[23] Jafari M, Platt G, Malekjamshidi Z, *et al.* Technical issues of sizing lead-acid batteries for application in residential renewable energy systems. In: Electric Power and Energy Conversion Systems (EPECS), 2015 4th International Conference on; 2015. pp. 1–6.

[24] Miozzo M, Zordan D, Dini P, *et al.* SolarStat: Modeling photovoltaic sources through stochastic Markov processes. In: 2014 IEEE International Energy Conference (ENERGYCON); 2014. pp. 688–695.

[25] Dalmasso M, Meo M, and Renga D. Radio resource management for improving energy self-sufficiency of green mobile networks. In: Performance Evaluation Review. vol. 44; 2016. pp. 82–87. DOI: 10.1145/3003977.3004001.

[26] Meo M, Zhang Y, Gerboni R, *et al.* Dimensioning the power supply of a LTE macro BS connected to a PV panel and the power grid. In: Communications (ICC), 2015 IEEE International Conference on; 2015. pp. 178–184.

[27] Chamola V and Sikdar B. Outage estimation for solar powered cellular base stations. In: Communications (ICC), 2015 IEEE International Conference on; 2015. pp. 172–177.

[28] Han F, Zhao S, Zhang L, *et al.* Survey of Strategies for Switching Off Base Stations in Heterogeneous Networks for Greener 5G Systems. IEEE Access. 2016;4:4959–4973.

[29] Ali M, Meo M, and Renga D. WiFi offloading for enhanced interaction with the Smart Grid in green mobile networks. In: 2017 IEEE International Conference on Networking, Sensing and Control (ICNSC), Calabria, Southern Italy; 2017.

[30] Chamola V and Sikdar B. Dimensioning stand-alone cellular base station using series-of-worst-months meteorological data. In: Communication Systems (ICCS), 2014 IEEE International Conference on; 2014. pp. 339–343.

[31] Ku M, Li W, Chen Y, *et al.* Advances in Energy Harvesting Communications: Past, Present, and Future Challenges. IEEE Communications Surveys Tutorials. 2016 Second quarter;18(2):1384–1412.

[32] Kimball JW, Kuhn BT, and Balog RS. A System Design Approach for Unattended Solar Energy Harvesting Supply. IEEE Transactions on Power Electronics. 2009;24(4):952–962.

[33] Gharavi H and Ghafurian R. IEEE recommended practice for sizing lead-acid batteries for stand-alone photovoltaic (PV) systems IEEE Std 1013-2007. In: Proc. IEEE. vol. 99; 2011. pp. 917–921.

[34] Han D, Li S, Peng Y, *et al.* Energy Sharing-Based Energy and User Joint Allocation Method in Heterogeneous Network. IEEE Access. 2020;8:37077–37086.

[35] Stevens JW and Corey GP. A study of lead-acid battery efficiency near top-of-charge and the impact on PV system design. In: Photovoltaic Specialists Conference, 1996. Conference Record of the Twenty Fifth IEEE; 1996. pp. 1485–1488.

[36] Al-Sheikh H and Moubayed N. Health status and diagnosis of batteries in renewable energy systems: An overview. In: Electrical and Power Engineering (EPE), 2012 International Conference and Exposition on; 2012. pp. 922–927.

[37] Hua S, Zhou Q, Kong D, *et al.* Application of Valve-Regulated Lead-Acid Batteries for Storage of Solar Electricity in Stand-Alone Photovoltaic Systems in the Northwest Areas of China. Journal of Power Sources. 2006;158(2): 1178–1185.

[38] Wu J, Zhang Y, Zukerman M, *et al.* Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey. IEEE Communications Surveys Tutorials. 2015 Second quarter;17(2):803–826.

[39] Prasanth TS, Kumar MSS, and Shankar T. A survey on techniques related to base station sleeping in Green communication and CoMP analysis. In: 2016 IEEE International Conference on Engineering and Technology (ICETECH); 2016. pp. 1059–1067.

[40] Ismail M, Zhuang W, Serpedin E, *et al.* A Survey on Green Mobile Networking: From The Perspectives of Network Operators and Mobile Users. IEEE Communications Surveys Tutorials. 2015 Third quarter;17(3):1535–1556.

[41] Hassan HAH, Pelov A, and Nuaymi L. Integrating Cellular Networks, Smart Grid, and Renewable Energy: Analysis, Architecture, and Challenges. IEEE Access. 2015;3:2755–2770.

[42] Al Haj Hassan H, Renga D, Meo M, and Nuaymi L. A Novel Energy Model for Renewable Energy-Enabled Cellular Networks Providing Ancillary Services to the Smart Grid. IEEE Transactions on Green Communications and Networking. 2019;3(2):381–396.

[43] Renga D, Al Haj Hassan H, Meo M, *et al.* Energy Management and Base Station On/Off Switching in Green Mobile Networks for Offering Ancillary Services. IEEE Transactions on Green Communications and Networking. 2018;2(3):868–880.

[44] Ali M, Meo M, and Renga D. Cost saving and ancillary service provisioning in green mobile networks. In: The Internet of Things for Smart Urban Ecosystems (IoT4SUE). Springer; 2018.

[45] Meo M, Renga D, and Sarti MT. Household users' cooperation to reduce cost in green mobile networks. In: 2018 International Telecommunications Energy Conference (INTELEC); 2018.

[46] Meo M, Zhang Y, Gerboni R, *et al.* Dimensioning the power supply of a LTE macro BS connected to a PV panel and the power grid. In: 2015 IEEE International Conference on Communications (ICC); 2015. pp. 178–184.

[47] Debaillie B, Desset C, and Louagie F. A flexible and future-proof power model for cellular base stations. In: 81st IEEE Vehicular Technology Conference. IEEE; 2015.

[48] Vallero G, Renga D, Meo M, *et al.* Greener RAN Operation Through Machine Learning. IEEE Transactions on Network and Service Management. 2019;16(3):896–908.

[49] Ruas GIS, Bragatto TAC, Lamar MV, *et al.* Electrical energy demand prediction using artificial neural networks and support vector regression. In: 2008 3rd International Symposium on Communications, Control and Signal Processing; 2008. pp. 1431–1435.

[50] Zhang D, Hu G, Lu J, *et al.* Short-term load forecasting based on GAD-BP neural network. In: 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC); 2019. pp. 1525–1529.

[51] Sun G, Cheng P, and Liu R. A hybrid short-term load prediction with input selection for the distribution system. In: 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT); 2019. pp. 826–830.

[52] Peng B, Wang C, and Tang X. Research on power system load forecasting model based on machine learning. In: 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA); 2019. pp. 477–480.

[53] Song J, Xue G, Pan X, *et al.* Hourly Heat Load Prediction Model Based on Temporal Convolutional Neural Network. IEEE Access. 2020;8:16726–16741.

[54] Liu Y, Dong J, Zuo D, *et al.* Experimental analysis and comparison of load prediction algorithms in cloud data center. In: 2019 IEEE 19th International Conference on Software Quality, Reliability and Security (QRS); 2019. pp. 197–203.

[55] Pan H, Liu J, Zhou S, *et al.* A block regression model for short-term mobile traffic forecasting. In: 2015 IEEE/CIC International Conference on Communications in China (ICCC). IEEE; 2015. pp. 1–5.

[56] Lee KY, Cha YT, and Park JH. Short-Term Load Forecasting Using an Artificial Neural Network. IEEE Transactions on Power Systems. 1992;7(1):124–132.

[57] Torres P, Marques H, Marques P, *et al.* Using deep neural networks for forecasting cell congestion on LTE networks: A simple approach. In: International Conference on Cognitive Radio Oriented Wireless Networks. Springer; 2017. pp. 276–286.

This page intentionally left blank

Chapter 6

Coverage and secrecy analysis of RF-powered Internet-of-Things

Mustafa A. Kishk[1], Mohamed A. Abd-Elmagid[2] and Harpreet S. Dhillon[2]

The Internet-of-Things (IoT) is an emerging *digital fabric* that will integrate our physical world into computer networks by connecting billions of things, such as small sensors, wearables, vehicles, and actuators, to the Internet. Owing to its massive scale, it is considered highly inefficient and even impractical to replace or recharge batteries of IoT devices, especially the ones that are deployed at hard-to-reach places, such as under the ground or in tunnels. This has naturally led to the consideration of energy harvesting to circumvent or supplement conventional power sources, such as replaceable batteries, in these devices. Due to its ubiquity and cost-efficient implementation, radio frequency (RF)-energy harvesting has quickly emerged as an appealing solution for powering IoT devices (majority of which are tiny devices, such as sensors, with very low energy requirement).

The objective of this chapter is to provide a comprehensive performance analysis of RF-powered IoT using tools from stochastic geometry. In order to capture the cyber-physical nature of IoT, our emphasis is on the metrics that jointly characterize the wireless, energy harvesting, and secrecy aspects. In the first part of this chapter, we characterize the joint probability of receiving strong enough signal and harvesting sufficient energy to operate the link. We term this the *joint coverage probability*. In this analysis, we assume that the locations of the sources of RF signals and the locations of the IoT devices are modeled using two independent Poisson point processes (PPPs). For this setup, we derive insightful mathematical expressions for key performance metrics, which collectively provide insights into the effect of the different system parameters on the overall system performance and how these parameters can be tuned to achieve the performance of a regular battery-powered system. In the second part of this chapter, we also incorporate the secrecy aspect in our analysis. In particular, we study the secrecy of RF signals when the RF-powered IoT devices are placed close to the sources of RF signals. Rigorous mathematical expressions are derived for various performance metrics, which provide several useful system design insights.

[1]CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. This work was done when M.A. Kishk was with Wireless@VT, Department of ECE, Virginia Tech.
[2]Department of ECE, Virginia Tech, Blacksburg, United States

6.1 Introduction

Owing to its ability to transform current towns and cities into smart and connected communities of tomorrow, IoT is widely regarded as one of the next frontiers in information and communications technology. The grand vision for an IoT network is to tightly integrate the cyber and physical worlds by connecting billions of things, such as small sensors, wearables, vehicles, and actuators, to the Internet. This will naturally enable numerous applications across many industrial verticals, such as home automation, intelligent transportation systems, public safety, agriculture, and medicine. A key hindrance in making this vision a reality is the energy-constrained nature of a majority of IoT devices [1]. It is well known that most of these devices will have to be battery powered, and replacing or recharging these batteries may not be a viable option for many of them. In order to overcome this issue, energy harvesting solutions have been proposed to supplement or even circumvent the use of replaceable batteries in the IoT devices. Due to its ubiquity and cost-efficient implementation, ambient RF-energy harvesting has recently been considered as an appealing solution for powering IoT devices.

Ambient RF-energy harvesting IoT devices (sometimes referred to as wirelessly powered IoT devices) can, in principle, have a *dedicated charging network*, which is deployed specifically for charging IoT devices through broadcasting RF-energy signals. For instance, one can envision a network of *power beacons* (PBs) deployed solely to charge the IoT devices. However, a far more realistic and cost-efficient alternative for the IoT devices is to simply rely on the existing communication infrastructure, such as Wi-Fi access points (APs) or cellular base stations (BSs), for both charging and communication. Given its dual purpose, this network needs to be carefully designed in order to (i) deliver the amount of RF-energy required at each IoT device and (ii) maintain reliable communication links between the APs or the BSs and the IoT devices. Since the same set of BSs (or APs) is used for both charging and communication, the amount of harvested energy and the communication signal quality (signal-to-interference-plus-noise ratio or SINR) are highly correlated. This, in turn, complicates the performance analysis of such system setup. In the first part of this chapter, we use tools from stochastic geometry to rigorously analyze the performance of this setup. In particular, we derive the joint energy and SINR coverage probability and use the derived expressions to provide several system-level insights.

In the second part of this chapter, we focus on secrecy in addition to the energy harvesting and wireless performance. Specifically, we focus on a scenario in which an RF-powered IoT network coexists with a primary network that is also the sole source of RF-energy for the IoT network. We further assume that the coexisting primary network is using the guard zone technique [2,3] to preserve the privacy of its transmitted signals. In such scenarios, RF-energy harvesting can be challenging since the IoT devices do not belong to the primary network and may hence be considered as potential eavesdroppers by the primary network. We use tools from stochastic geometry to analyze the performance of the two coexisting networks. In particular, we study the secrecy performance of the primary network and the energy harvesting performance of the IoT devices. Furthermore, we investigate how the performance of

the two networks is affected by the key system parameters, such as guard zone radius and the deployment density of the IoT devices, which allows us to obtain useful insights on the coexistence of the two networks.

6.1.1 Literature review

As already noted, this chapter considers two aspects of the performance of RF-powered IoT devices: (i) energy and SINR coverage of IoT devices and (ii) secrecy of RF signals used for energy harvesting. We use tools from stochastic geometry to study each of these aspects. The existing literature on the use of stochastic geometry for the analysis of energy harvesting wireless networks has focused mainly on the coverage aspect. On the other hand, the works that discussed the secrecy problem have focused mainly on simpler setups composed of a single point-to-point link or a fixed set of nodes, unlike the work in this chapter that considers large-scale networks. Therefore, the literature review will focus on two main research directions: (i) stochastic geometry-based analysis of energy harvesting wireless networks with emphasis on coverage analysis and (ii) secrecy analysis of RF signals used to charge RF-powered wireless networks.

Stochastic geometry has emerged as a promising mathematical tool for the system-level performance analysis of wireless networks [4–7]. Not surprisingly, there have also been a lot of recent works that use stochastic geometry to study the performance of large-scale energy harvesting wireless networks [8–12]. This literature has focused mainly on the analysis of a setup where a wireless device harvests energy and then uses it to transmit information to a receiver. The main performance metric of interest in these papers is the joint probability of energy and SINR coverage, which is the joint probability of (i) harvesting sufficient energy to be able to transmit the information signal to the receiver and (ii) successfully decoding the information signal transmitted by the energy harvesting device. These two events are independent when the locations of RF-energy sources are independent from the locations of the information receivers. For instance, in [8], authors considered a system where a sensor harvests RF-energy from the ambient signals emitted by TV and radio stations, as well as cellular networks. The sensor harvests the RF-energy and then uses it to communicate with a data sink. The authors modeled the locations of the sources of RF-energy using Ginibre α-determinantal point process (DPP). The main advantage of this point process over a PPP is its capability of capturing repulsion among the locations of TV, radio stations as well as cellular BSs. Given the technical challenges in deriving exact expressions for DPP, authors derived useful bounds for the energy coverage probability and the average harvested energy by considering the worst case scenario that focuses on the energy harvested from the nearest RF source only. The concept of having a dedicated charging network with the sole purpose of providing RF-energy was proposed in [9] for a simple point-to-point setup. In this setup, the charging network is represented by a set of PBs. This setup was then extended to a large-scale network in [10].

When the IoT device harvests RF-energy from a given network and then uses it to transmit information to the same network, the energy and SINR coverage events

become correlated. This is due to implicit correlation between the locations of successfully powered nodes (interferers during information transmission) and the locations of the RF sources. However, capturing this correlation is highly challenging and hence is typically ignored in the literature to maintain tractability [11,12]. For instance, authors in [11] considered a k-tier cellular network where the users harvest RF-energy from the signals emitted by the cellular BSs and then use this energy to communicate with the nearest BS in the uplink channel. The authors assumed the existence of a limited-capacity battery at each user and derived the energy coverage probability as well as uplink SINR coverage probability by modeling the battery level at each user as a Markov Chain followed using tools from stochastic geometry.

Another set of relevant works in the first direction considered the scenario where the energy harvesting device uses the harvested energy to activate its processing units and enable information reception [13–22]. This is particularly important in wireless devices with limited power resources, such as the IoT devices. In such devices, the energy consumption during information reception should not be neglected [23–25]. Authors in [13] considered a power-splitting architecture for the RF-energy harvesting devices. This architecture splits the received signal into two portions. One portion is used to charge the device, and the other portion is decoded as an information-carrying signal. This work considered a large-scale network of energy harvesting devices, where the distance between each device and its information transmitter was assumed to be fixed. The objective of this system is to maintain the average amount of harvested energy above a predefined threshold while maximizing the SINR coverage probability. For this setup, the SINR and the amount of harvested energy are analyzed separately. The joint analysis of the amount of harvested energy and the SINR at the energy harvesting device is much more challenging due to the high correlation between the two random variables. This correlation is induced by the fact that the same transmitters are used for information reception and energy harvesting. Hence, the amount of harvested energy and the SINR are both functions of the same point process. This problem has not received as much attention in the literature [26,27]. For instance, authors in [26] derived an upper bound on the joint energy and SINR coverage probability in order to provide tractable expressions that enable drawing system-level insights. In [16,18,20,21], the authors of this chapter proposed a simple approximation that assisted in deriving this joint probability and provided several useful system-level insights. The first part of this chapter will be based on these recent developments.

In the second part of this chapter, we use tools from stochastic geometry to study the scenario where the IoT devices harvest energy from the RF signals transmitted by a coexisting wireless network. This coexisting network is assumed to adopt a secrecy-enhancing technique to maintain the confidentiality of its transmitted messages. It is instructive to note that, until recently, this problem was only studied in the literature for the point-to-point setup or a setup with a fixed number of transmitters, RF-powered devices, and legitimate receivers [28–32]. The general theme in these works is the implicit assumption that the transmitter aims to ensure secrecy while providing RF-energy for the RF-powered device. For instance, authors in [28,29] studied a system of one transmitter–receiver pair with the coexistence of one RF-powered device. In order

to satisfy both secrecy and RF-energy delivery needs, the transmitter uses artificial noise to maintain the amount of RF-energy while reducing the probability of decoding the information signal by the RF-energy receiver. Authors in [30] proposed to use a jammer to increase the amount of RF-energy while degrading the information signal quality at the RF-powered device. Authors in [31] considered a setup of a single transmitter–receiver pair with K coexisting RF-powered devices. A more general setup with K transmitters, N receivers, and M RF-powered devices was considered in [32]. One of the few works that considered secrecy in large-scale networks with RF-powered devices using stochastic geometry is [33]. This work studied a system setup in which the legitimate transmitters are RF-powered, which is different from the setup considered in this chapter. In [17,19], the authors of this chapter have recently extended the point-to-point setup to a more general system of two coexisting large-scale networks: (i) the IoT network and (ii) the secrecy-enhancing network. These recent works will form the basis of the second part of this chapter.

6.2 RF-energy harvesting from a coexisting cellular network

6.2.1 System setup

We consider a system of RF-powered IoT devices and a charging network (cellular network) that dedicates a subset of its resources to serve the IoT devices. The locations of the IoT devices and the cellular BSs are modeled by two independent homogeneous PPPs $\Phi_E \equiv \{x_i\} \subset \mathbb{R}^2$ with density λ_E and $\Phi_R \equiv \{y_i\} \subset \mathbb{R}^2$ with density λ_R, respectively. Hence, without loss of generality, we focus our analysis on a typical IoT device located at the origin due to the stationarity of PPP. To enable simultaneous charging and communication, we consider a time-switching architecture for the IoT device. In particular, each time-slot is divided into two sub-slots: (i) charging sub-slot with duration $\tau_c T$ and (ii) downlink sub-slot with duration $\tau_d T$, where T is the total duration of each time-slot and $\tau_c + \tau_d = 1$. In the charging sub-slot, all the BSs emit RF signals with the purpose of charging the IoT devices. In the downlink sub-slot, each IoT device associates with its nearest BS and receives information-carrying signals from this BS. We assume that fading gains among all sets of BS-IoT device links are independent and exponentially distributed with mean 1. In addition, we assume that for a given link, the fading gains in the charging and downlink sub-slots are independent.

In the charging sub-slot, the amount of power received by the typical device from a BS located at $y \in \Phi_R$ is $\rho g_y \|y\|^{-\alpha}$, where ρ is the transmission power of the BS, $g_y \sim \exp(1)$ is the fading gain in the charging sub-slot, and α is the path-loss exponent. Hence, given that the duration of the charging sub-slot is $\tau_c T$ and that all the BSs are active during the charging sub-slot, the total energy harvested by the typical IoT device in the charging sub-slot is

$$E_H = \eta \tau_c T \sum_{y \in \Phi_R} \rho g_y \|y\|^{-\alpha}, \tag{6.1}$$

where η is the RF-DC conversion efficiency. We assume that any leftover energy from the previous slots is no longer available for use in the current time slot.

During the downlink sub-slot, the value of the SINR at the typical IoT device is

$$\text{SINR} = \frac{\rho h_{y_1} \|y_1\|^{-\alpha}}{\sum_{y \in \Phi_R \setminus y_1} \rho h_y \|y\|^{-\alpha} + \sigma^2}, \tag{6.2}$$

where σ^2 is the noise power, y_1 is the location of the nearest BS to the typical IoT device, and $h_y \sim \exp(1)$ is the fading gain in the downlink sub-slot.

6.2.2 Performance metrics

For the system setup described previously, our design goal is to ensure that the IoT devices are harvesting sufficient energy and the downlink SINR is above a predefined threshold. However, due to the correlation arising from relying on the same set of BSs for both charging and communication, the analysis of IoT charging cannot be separated from that of downlink communication. Hence, we focus on deriving the joint energy and SINR coverage probability, which is defined next.

Definition 6.1 (Joint coverage probability). *For a given time-slot, the IoT device needs to satisfy two conditions: (i) $E_H \geq \mathscr{E}$ and (ii)* SINR $\geq \beta$*, where $\mathscr{E}$ is the minimum threshold on E_H to ensure acceptable energy harvesting performance and β is the minimum threshold on* SINR *required for successful decoding. The probability of satisfying both conditions is defined as the* joint energy and SINR coverage probability*, which can be mathematically represented as follows:*

$$P_{\text{joint}} = \mathbb{P}(E_H \geq \mathscr{E}, \text{SINR} \geq \beta). \tag{6.3}$$

Another metric that is typically studied in the literature of RF-powered wireless networks is the energy coverage probability, which is defined as

$$P_{\text{energy}} = \mathbb{P}(E_H \geq \mathscr{E}). \tag{6.4}$$

Obviously, the energy coverage probability is a special case of the joint coverage probability, i.e., the joint coverage probability reduces to the energy coverage probability when $\beta = 0$.

Another important metric to quantify the performance of this setup is the average throughput. We assume that when the IoT device fails to harvest the minimum required amount of energy $\mathscr{E}$, it cannot communicate in the downlink sub-slot. Given that the link is only used for communication for a τ_{d} fraction of time, it is important to study how the parameters τ_{c} and τ_{d} should be selected in a way that ensures energy coverage while maximizing the downlink average throughput.

Definition 6.2 (Average throughput). *The average downlink throughput in bits/s/Hz is*

$$\begin{aligned} D_{\text{avg}} &= \tau_{\text{d}} \mathbb{E}\left[\log_2(1+\beta) \mathbb{1}(\text{SINR} \geq \beta) \mathbb{1}(E_{\text{H}} \geq \mathscr{E})\right] \\ &= \tau_{\text{d}} \log_2(1+\beta) P_{\text{joint}}, \end{aligned} \tag{6.5}$$

where $\mathbb{1}(\Xi) = 1$ if the event Ξ happens and $\mathbb{1}(\Xi) = 0$ otherwise.

6.2.3 Analysis and main results

We aim to derive the joint coverage probability described in Definition 6.1, which is the joint probability of the two events: (i) $E_H \geq \mathscr{E}$ and (ii) SINR $\geq \beta$. Recalling (6.1) and (6.2), we can observe that both E_H and SINR depend on Φ_R, which leads to high correlation between the two random variables. In particular, Φ_R models the locations of the sources of RF-energy signals in E_H as well as the locations of the interferers in SINR. Building on that observation, we can treat E_H and SINR as two independent random variables when we condition our analysis on Φ_R. Hence, we can rewrite (6.3) as follows:

$$P_{\text{joint}} = \mathbb{E}_{\Phi_R}[\mathbb{P}(E_H \geq \mathscr{E}|\Phi_R)\mathbb{P}(\text{SINR} \geq \beta|\Phi_R)]. \tag{6.6}$$

Based on the previous expression, we first need to derive each of the energy and SINR coverage probabilities conditioned on the point process Φ_R. Next, we need to take the expectation of their product over Φ_R. Before proceeding with this approach, we propose an efficient approximation that has been shown to be remarkably accurate in the literature [34–39]. This approximation is provided next:

$$\sum_{y\in\Phi_R} g_y \|y\|^{-\alpha} = g_{y_1}\|y_1\|^{-\alpha} + g_{y_2}\|y_2\|^{-\alpha} + \mathscr{M}(y_1, y_2), \tag{6.7}$$

where y_1 and y_2 are the locations of the two nearest BSs to the typical IoT device, respectively, and $\mathscr{M}(y_1, y_2) = \mathbb{E}\big[\sum_{y\in\Phi_R\setminus y_1,y_2} g_y\|y\|^{-\alpha}|y_1, y_2\big]$. In (6.7), we simply approximate the summation of received RF signal powers by the summation of RF signal powers received from the nearest two BSs and the expectation of the summation of RF signal powers received from the rest of the BSs conditioned on the location of the nearest two BSs. This is motivated by the power-law path-loss, because of which the RF signals received from the nearby BSs dominate the total received power. Using this approach, we can approximate the value of E_H in (6.1). This leads to the following approximate expression:

$$E_H = \eta\tau_c T\rho\big(g_{y_1}\|y_1\|^{-\alpha} + g_{y_2}\|y_2\|^{-\alpha} + \mathscr{M}(y_1, y_2)\big). \tag{6.8}$$

Using similar approach, we can approximate the value interference term in the dominator of the SINR expression in (6.2). This leads to the following approximate expression:

$$\text{SINR} = \frac{\rho h_{y_1}\|y_1\|^{-\alpha}}{\rho\big(h_{y_2}\|y_2\|^{-\alpha} + \mathscr{M}(y_1, y_2)\big) + \sigma^2}. \tag{6.9}$$

As will be evident shortly, this lends tractability to the analysis of the joint distribution in (6.3). In particular, this approximation reduces (6.6) to:

$$P_{\text{joint}} = \mathbb{E}_{y_1,y_2}[\mathbb{P}(E_H \geq \mathscr{E}|y_1, y_2)\mathbb{P}(\text{SINR} \geq \beta|y_1, y_2)]. \tag{6.10}$$

In the following lemma, we derive the first term inside the expectation in (6.10), namely, $\mathbb{P}(E_H \geq \mathscr{E}|y_1, y_2)$.

Lemma 6.1 (Conditional energy coverage probability). *Probability that the harvested energy during the charging sub-slot is greater than $\mathscr{E}$ conditioned on y_1, y_2 is*

$$\mathbb{P}\left(E_{\mathrm{H}} \geq \mathscr{E} \middle| y_1, y_2\right) = \frac{r_2^{\alpha} \exp\left(-r_1^{\alpha}\left[\mathscr{F}(r_2)\right]^{+}\right)}{r_2^{\alpha} - r_1^{\alpha}} - \frac{r_1^{\alpha} \exp\left(-r_2^{\alpha}\left[\mathscr{F}(r_2)\right]^{+}\right)}{r_2^{\alpha} - r_1^{\alpha}}, \quad (6.11)$$

while the unconditioned probability is

$$\begin{aligned}\mathbb{P}\left(E_{\mathrm{H}} \geq \mathscr{E}\right) &= 1 - \pi\lambda_R \mathscr{A}^2 \exp\left(-\pi\lambda_R \mathscr{A}^2\right) - \exp\left(-\pi\lambda_R \mathscr{A}^2\right) \\ &\quad + \int_{\mathscr{A}}^{\infty} \int_{0}^{r_2} \left(\frac{r_2^{\alpha} \exp\left(-r_1^{\alpha}\mathscr{F}(r_2)\right)}{r_2^{\alpha} - r_1^{\alpha}} - \frac{r_1^{\alpha} \exp\left(-r_2^{\alpha}\mathscr{F}(r_2)\right)}{r_2^{\alpha} - r_1^{\alpha}} \right) \\ &\quad \times f_{R_1,R_2}(r_1, r_2)\mathrm{d}r_1\mathrm{d}r_2, \end{aligned} \quad (6.12)$$

where $\mathscr{A} = \left(\frac{2\pi\lambda_R}{(\alpha-2)C(\tau_{\mathrm{c}})}\right)^{\frac{1}{\alpha-2}}$, $r_1 = \|y_1\|$, $r_2 = \|y_2\|$, $\mathscr{F}(r_2) = \left[C(\tau_{\mathrm{c}}) - \frac{2\pi\lambda_R}{\alpha-2} r_2^{2-\alpha}\right]$, $C(\tau_{\mathrm{c}}) = \frac{\mathscr{E}}{\tau_{\mathrm{c}} T \eta \rho}$, $[x]^{+} = \max\{0, x\}$, *and* $f_{R_1,R_2}(r_1, r_2) = (2\pi\lambda_R)^2 r_1 r_2 e^{-\lambda_R \pi {r_2}^2}$.

Proof. The value of $\mathscr{M}(y_1, y_2)$ can be derived as follows:

$$\begin{aligned}\mathscr{M}(y_1, y_2) &= \mathbb{E}\left[\sum_{y \in \Phi_R \setminus y_1, y_2} g_y \|y\|^{-\alpha} \middle| y_1, y_2\right] \overset{(a)}{=} \mathbb{E}\left[\sum_{y \in \Phi_R \setminus y_1, y_2} \|y\|^{-\alpha}\right] \\ &\overset{(b)}{=} 2\pi\lambda_R \int_{r_2}^{\infty} \frac{1}{r^{\alpha}} r \mathrm{d}r = \frac{2\pi\lambda_R}{\alpha - 2} ({r_2}^{2-\alpha}), \end{aligned} \quad (6.13)$$

where (a) follows from the assumption that all $\{g_y\}$ are independent and exponentially distributed random variables with mean one, and (b) follows from Campbell's theorem [40] with conversion from Cartesian to polar coordinates and using $r_2 = \|y_2\|$. Using the approximation introduced in (6.7), the conditional energy coverage probability can be expressed as

$$\begin{aligned}\mathbb{P}\left(E_{\mathrm{H}} \geq \mathscr{E} \middle| y_1, y_2\right) &= \mathbb{P}\left(\tau_{\mathrm{c}} T \eta \rho \left(g_{y_1} r_1^{-\alpha} + g_{y_2} r_2^{-\alpha} + \frac{2\pi\lambda_R}{\alpha - 2} r_2^{2-\alpha}\right) \geq \mathscr{E}\right) \\ &= \mathbb{P}\left(g_{y_1} r_1^{-\alpha} + g_{y_2} r_2^{-\alpha} \geq C(\tau_{\mathrm{c}}) - \frac{2\pi\lambda_R}{\alpha - 2} r_2^{2-\alpha}\right) \\ &= \mathbb{P}\left(g_{y_1} r_1^{-\alpha} + g_{y_2} r_2^{-\alpha} \geq \mathscr{F}(r_2)\right) \\ &\overset{(c)}{=} \frac{r_2^{\alpha} \exp(-r_1^{\alpha}[\mathscr{F}(r_2)]^{+}) - r_1^{\alpha} \exp(-r_2^{\alpha}[\mathscr{F}(r_2)]^{+})}{r_2^{\alpha} - r_1^{\alpha}}, \end{aligned} \quad (6.14)$$

where step (c) is due to hypo-exponential distribution of $g_{y_1} r_1^{-\alpha} + g_{y_2} r_2^{-\alpha}$ (sum of two exponential random variables with rates r_1^{α} and r_2^{α}), $C(\tau_{\mathrm{c}}) = \frac{\mathscr{E}}{\tau_{\mathrm{c}} T \eta \rho}$, and $[x]^{+} = \max\{0, x\}$. This concludes the proof of (6.11). Given that $\mathbb{P}\left(E_{\mathrm{H}} \geq \mathscr{E} \middle| y_1, y_2\right) = 1$

when $\mathscr{F}(r_2) \leq 0$, we define two sets: $\mathscr{N}_{r_2} = \{r_1 : \mathscr{F}(r_2) \leq 0,\ r_1 < r_2\}$ and $\mathscr{P}_{r_2} = \{r_1 : \mathscr{F}(r_2) \geq 0,\ r_1 < r_2\}$. We note that the set $\mathscr{N}_{r_2}$ is an empty set for $r_2 \geq \mathscr{A}$, while for $r_2 \leq \mathscr{A}$ the set reduces to $\mathscr{N}_{r_2} = \{r_1 : r_1 \leq r_2\}$. Similarly, the set $\mathscr{P}_{r_2}$ is an empty set for $r_2 \leq \mathscr{A}$, while for $r_2 \geq \mathscr{A}$ the set reduces to $\mathscr{P}_{r_2} = \{r_1 : r_1 \leq r_2\}$. Using these observations and integrating over r_1 and r_2 with $f_{R_1,R_2}(r_1, r_2) = (2\pi\lambda_R)^2 r_1 r_2 e^{-\lambda_R \pi r_2^2}$ [41], the result in (6.12) follows. □

Remark 6.1. *The energy coverage probability is significantly affected by the duration of the charging sub-slot. This intuitive insight is captured clearly in the previous theorem by $C(\tau_c)$, which is a decreasing function of τ_c. As this value decreases, the energy coverage probability in (6.12) increases.*

Now, in the next lemma, we derive the second term inside the expectation in (6.10), which is $\mathbb{P}(\text{SINR} \geq \beta | y_1, y_2)$.

Lemma 6.2 (Conditional SINR coverage probability). *Probability that the downlink SINR at the typical IoT device exceeds β, conditioned on y_1 and y_2, is*

$$\mathbb{P}\left(\text{SINR} \geq \beta \middle| y_1, y_2\right) = \exp(-\mathscr{G}(r_1, r_2)) \frac{1}{1 + \frac{\beta r_1^\alpha}{r_2^\alpha}}, \tag{6.15}$$

where $r_1 = \|y_1\|$, $r_2 = \|y_2\|$, $\mathscr{G}(r_1, r_2) = \frac{\beta\sigma^2 r_1^\alpha}{\rho} + \frac{2\pi\lambda_R \beta r_1^\alpha}{(\alpha-2) r_2^{\alpha-2}}$.

Proof. Using the definition of SINR in (6.9), we get:

$$\begin{aligned}
\mathbb{P}(\text{SINR} \geq \beta | y_1, y_2) &= \mathbb{P}\left(\frac{\rho h_{y_1} r_1^{-\alpha}}{\rho h_{y_2} r_2^{-\alpha} + \rho \mathscr{M}(y_1, y_2) + \sigma^2} \geq \beta \middle| y_1, y_2 \right) \\
&\overset{(d)}{=} \mathbb{P}\left(\frac{\rho h_{y_1} r_1^{-\alpha}}{\rho h_{y_2} r_2^{-\alpha} + \rho \frac{2\pi\lambda_R r_2^{2-\alpha}}{\alpha-2} + \sigma^2} \geq \beta \middle| y_1, y_2 \right) \\
&= \mathbb{P}\left(h_{y_1} r_1^{-\alpha} \geq \frac{\beta\sigma^2}{\rho} + \frac{2\pi\lambda_R \beta r_2^{2-\alpha}}{\alpha-2} + \beta h_{y_2} r_2^{-\alpha} \middle| y_1, y_2 \right) \\
&\overset{(e)}{=} \mathbb{E}_{h_{y_2}}\left[\exp\left(-r_1^\alpha \left(\frac{\beta\sigma^2}{\rho} + \frac{2\pi\lambda_R \beta r_2^{2-\alpha}}{\alpha-2} + \beta h_{y_2} r_2^{-\alpha} \right) \right) \right] \\
&\overset{(f)}{=} \exp(-\mathscr{G}(r_1, r_2)) \frac{1}{1 + \beta \frac{r_1^\alpha}{r_2^\alpha}},
\end{aligned} \tag{6.16}$$

where (d) follows from substituting for $\mathscr{M}(y_1, y_2)$ as derived in (6.13), and steps (e) and (f) follow from the assumption that $h_y \sim \exp(1)$, and defining $\mathscr{G}(r_1, r_2) = \frac{\beta\sigma^2 r_1^\alpha}{\rho} + \frac{2\pi\lambda_R \beta r_2^{2-\alpha} r_1^\alpha}{\alpha-2}$. □

We can now proceed to the final step of deriving the joint coverage probability described in Definition 6.1. As can be observed from (6.10), the only remaining step

is to take the expectation of the product of $\mathbb{P}(E_H \geq \mathscr{E}|y_1, y_2)$ and $\mathbb{P}(\text{SINR} \geq \beta|y_1, y_2)$ over y_1 and y_2. The final result is provided in the following theorem.

Theorem 6.1 (Joint coverage probability). *The joint coverage probability, introduced in Definition 6.1, is given by*

$$P_{\text{joint}} = \int_0^{\mathscr{A}} \int_0^{r_2} f_{R_1,R_2}(r_1, r_2) \exp(-\mathscr{G}(r_1, r_2)) \frac{1}{1 + \frac{\beta r_1^\alpha}{r_2^\alpha}} \mathrm{d}r_1 \mathrm{d}r_2 + \int_{\mathscr{A}}^{\infty} \int_0^{r_2} f_{R_1,R_2}(r_1, r_2) \exp(-\mathscr{G}(r_1, r_2)) \times \frac{r_2^\alpha \exp\left(-r_1^\alpha \mathscr{F}(r_2)\right) - r_1^\alpha \exp\left(-r_2^\alpha \mathscr{F}(r_2)\right)}{\left(r_2^\alpha - r_1^\alpha\right)\left(1 + \frac{\beta r_1^\alpha}{r_2^\alpha}\right)} \mathrm{d}r_1 \mathrm{d}r_2, \tag{6.17}$$

where $\mathscr{G}(r_1, r_2)$ is defined in Lemma 6.2, $\mathscr{A}$, and $\mathscr{F}(r_2)$ are defined in Lemma 6.1.

Proof. This result follows directly by substituting (6.11) and (6.15) in (6.10) and integrating over r_1 and r_2 using the joint distribution $f_{R_1,R_2}(r_1, r_2)$ as defined in [41, (28)]. □

Remark 6.2. *As the duration of the charging sub-slot increases, the probability of satisfying the energy coverage condition increases. Hence, for the large enough τ_c, the joint coverage probability reduces to the probability of satisfying the SINR coverage condition. This insight is captured in the previous theorem through the value of $\mathscr{A}$, which is an increasing function of τ_c. Increasing the value of $\mathscr{A}$ decreases the second term in (6.17) and increases the first term. When $\mathscr{A}$ approaches ∞, the expression in (6.17) reduces to the SINR coverage probability.*

Remark 6.3. *The value of $\mathscr{A}$ represents a threshold on r_2. In particular, for a given Φ_R, the energy coverage condition is satisfied when $r_2 \leq \mathscr{A}$. This can be observed from the integration limits in (6.17). The value of $\mathscr{A}$ can be used to optimize the deployment of RF-powered IoT devices in order to ensure high energy coverage probability. Recalling the expression of $\mathscr{A}$ provided in Lemma 6.1, we observe that increasing its value can be achieved through increasing the density of λ_R or the duration of the charging sub-slot.*

6.2.4 Numerical results and discussion

In this section, we demonstrate the accuracy of the derived expressions, verify the insights provided in the remarks, and use the numerical results to draw other important system-level insights and performance trends. The values of system parameters used in the simulation setup are $\mathscr{E} = 1$ μJ, $\rho = 1$, $\lambda_R = 10^{-4}$ m^{-2}, $\alpha = 4$, and $\eta = 0.75$.

In Figure 6.1, we plot the energy coverage probability. Clearly, the value of the energy coverage probability increases as we increase τ_c, which agrees with our comments in Remark 6.1. The perfect match between the theoretical and simulation

results demonstrates the remarkable accuracy of the approximation in (6.7). We also observe the high influence of the value of λ_R on the energy coverage probability.

In Figure 6.2, we plot the joint coverage probability, which was derived in Theorem 6.1. We observe that as τ_c increases, the joint coverage probability starts saturating to a fixed value, which coincides with the SINR coverage probability. This is due to the high energy coverage probability at higher values of τ_c, as observed from Figure 6.1, which, in turn, reduces the P_{joint} to $\mathbb{P}(\text{SINR} \geq \beta)$. We also note that reducing the value of λ_R, which reduces the energy coverage probability, also reduces the joint coverage probability.

Finally, the average throughput, described in Definition 6.2, is plotted in Figure 6.3. The results show the existence of an optimum value for τ_c that maximizes

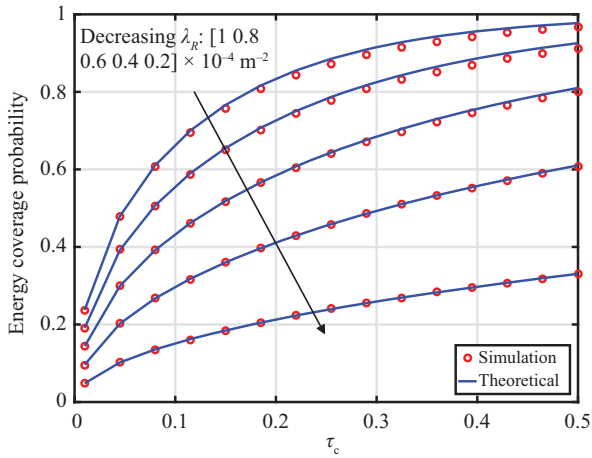


Figure 6.1 Energy coverage probability against different values of τ_c

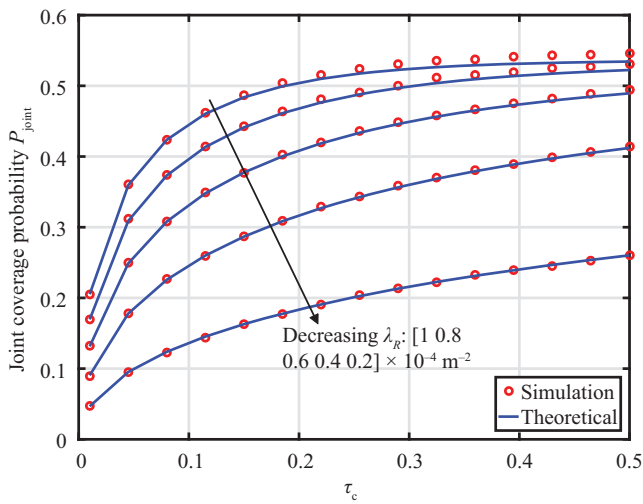


Figure 6.2 Joint coverage probability against different values of τ_c

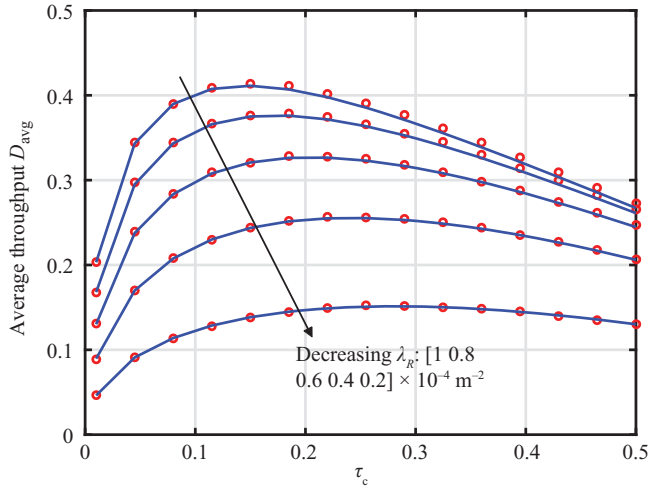


Figure 6.3 Average throughput against different values of τ_c

the average throughput. We note that this optimal value increases as we reduce λ_R, due to the decrease in the density of RF-chargers, which increases the amount of time needed for charging.

6.3 RF-energy harvesting from a coexisting, secrecy-enhancing network

In the previous section, we studied a scenario where there was a coexisting cellular network that allocates some of its resources for powering the IoT devices. On the contrary, we study a scenario where the IoT devices rely on a coexisting *primary network* for harvesting RF-energy. The primary network is composed of primary transmitters and receivers (PTs and PRs) and uses some secrecy-enhancing transmission policy to ensure a certain level of secure communication probability for the primary communication links (between PTs and PRs). As described next in detail, the IoT devices would be secondary devices for this network. Our objective in this section is to investigate the impact of the use of secrecy-enhancing technique on the energy harvesting performance of the RF-powered IoT devices.

6.3.1 System setup

Similar to the previous section, we model the locations of the IoT devices and the PTs using two independent PPPs $\Phi_E \equiv \{x_i\} \subset \mathbb{R}^2$ and $\Phi_R \equiv \{y_i\} \subset \mathbb{R}^2$, with densities λ_E and λ_R, respectively. Each PT aims to communicate with its associated PR and transmit confidential messages. In order to maintain tractability, we consider a Poisson bipolar model where the distance between each PT and its associated PR is r_1. Due to the stationarity of PPP, we focus our analysis on a typical PT–PR pair with the PR

located at the origin. Different from the system setup studied in the previous section, the charging network (the primary network) in this setup (i) does not allocate any resources for solely charging the IoT devices and (ii) is using a secrecy-enhancing technique that affects the energy harvesting performance of the IoT devices. The secrecy-enhancing technique used by the primary network is described next.

There are multiple secrecy-enhancing techniques studied in literature such as (i) beamforming [42–44], (ii) protected zones [45,46], (iii) artificial noise addition [47,48], and (iv) guard zones [2,3]. The guard zone technique was recently proven to outperform other techniques under some specific deployment scenarios [49]. Hence, we focus on this technique in this chapter. The guard zone technique is based on ensuring that the distance between the PT and its nearest illegitimate receiver (from the perspective of PT, any receiver except PRs is an illegitimate receiver) is above a given threshold. Otherwise, the PT stops its transmission (goes silent). While this discussion is not in the scope of the current chapter, the illegitimate receivers can be detected by the primary network using specialized devices such as metal detectors or leaked local oscillator power detectors [45]. To focus on the interaction between the IoT devices and the primary network, we assume that the IoT devices are the only existing illegitimate receivers in the system. The radius of the guard zone is the main design parameter in this secrecy-enhancing technique. The selection of this parameter is based on the performance metrics that the primary network aims to maximize. Before defining the performance metrics used in this chapter, we first provide some physical layer security preliminaries.

In order to ensure secrecy, the PT selects two transmission rates: (i) code word transmission rate $\mathscr{R}_c$ and (ii) confidential messages transmission rate $\mathscr{R}_m$, where $\mathscr{R}_c > \mathscr{R}_m$. The difference $\mathscr{R}_c - \mathscr{R}_m$ represents the cost paid for secure transmission. To ensure that the PR is able to successfully decode the confidential message, the mutual information between PT's channel input and the PR's channel output should be greater than $\mathscr{R}_c$. In addition, to ensure perfect secrecy, the mutual information between the PT's channel input and any illegitimate receiver's channel output should be less than $\mathscr{R}_c - \mathscr{R}_m$. This can be translated into two conditions. The first one is $\mathrm{SINR}_\mathrm{R} \geq 2^{\mathscr{R}_c} - 1$, to ensure successful connection between PT and PR, where:

$$\mathrm{SINR}_\mathrm{R} = \frac{\rho w_1 r_1^{-\alpha}}{\sum_{y_i \in \Phi_R \setminus y_1} \delta_i \rho w_i \|y_i\|^{-\alpha} + \sigma_P^2}, \tag{6.18}$$

σ_p^2 is the noise power, and $w_i \sim \exp(1)$ models the Rayleigh fading gain for the link between the PT located at y_i and the typical PR. We use the subscript $i = 1$ to refer to the typical PT–PR pair. The value of the indicator function $\delta_i = 1$ if the PT located at y_i is active, which means it does not have any illegitimate receivers in its guard zone. Otherwise, if its guard zone has at least one illegitimate receiver, we have $\delta_i = 0$. This captures the effect of the guard zone radius on the interference levels at both the legitimate and illegitimate receivers. Clearly, as the radius of the guard zone r_g increases, more PTs will go silent, which, in turn, leads to less interference. The expected value of this indicator function equals to the probability of the PT being active: $\mathbb{E}[\delta_i] = P_\mathrm{active}$. The value of P_active can be derived as follows. Denoting the

distance between the typical PT and its nearest IoT device by D_e, and recalling that the locations of the IoT devices are modeled by a PPP with density λ_E, we then have:

$$P_{\text{active}} = \mathbb{P}(D_e \geq r_g) = \exp(-\pi \lambda_E r_g^2). \tag{6.19}$$

The second condition that needs to be satisfied to ensure perfect secrecy is $\text{SINR}_{\text{E}}(x_j) \leq 2^{\mathscr{R}_c - \mathscr{R}_m} - 1$, where $\text{SINR}_{\text{E}}(x_j)$ is the value of SINR at the IoT device located at x_j. This condition should be satisfied at all IoT devices in order to ensure perfect secrecy. The SINR of the confidential signal transmitted by the typical PT measured at the IoT device located at x_j is

$$\text{SINR}_{\text{E}}(x_j) = \frac{\rho g_{1,j} \|y_1 - x_j\|^{-\alpha}}{\sum_{y_i \in \Phi_R \setminus y_1} \delta_i \rho g_{i,j} \|y_i - x_j\|^{-\alpha} + \sigma_S^2}, \tag{6.20}$$

where $g_{i,j} \sim \exp(1)$ is the fading gain for the link between the PT located at y_i and the IoT device located at x_j, and σ_S^2 is the noise power at the IoT device.

Clearly, only those PTs are active that are at a distance greater than r_g from all IoT devices. This can be formally defined as follows:

$$\bar{\Phi}_R = \left\{ y \in \Phi_R : x \notin \bigcup_{x \in \Phi_E} \mathscr{B}(x, r_g) \right\}, \tag{6.21}$$

where $\mathscr{B}(x, r_g)$ is a ball centered at x with radius r_g, and $\bar{\Phi}_R$ models the locations of the active PTs. The point process $\bar{\Phi}_R$ in (6.21) is nothing but a Poisson hole process (PHP) [50,51]. The density of the PHP $\bar{\Phi}_R$ that models the locations of active PTs is $P_{\text{active}} \lambda_R$.

Similar to the previous section, one of the main aspects of the performance of the IoT devices that we focus on is their ability to harvest sufficient amount of energy. The amount of energy harvested by the IoT device located at x_j, assuming an RF-energy harvesting time-slot of duration T, is

$$E_H = \eta T \rho \|y_{j,1} - x_j\|^{-\alpha} g_j, \tag{6.22}$$

where $g_j \sim \exp(1)$ models the Rayleigh fading gain of the link between the IoT device located at x_j and its nearest active PT. Here, we are focusing on the energy harvested from the nearest active PT, located at $y_{j,1}$. This is motivated by many recent studies that showed the dominance of the RF-energy harvested from the nearest transmitter in the total amount of harvested energy, such as [16]. This assumption lends tractability to an otherwise intractable analysis. For instance, this enables us to study the effect of some important system parameters, such as r_g and λ_E on the statistics of the harvested energy. In particular, using this approach, we can capture an interesting behavior in this system, which is how increasing the density of the IoT devices λ_E decreases the average distance between a typical IoT device and its nearest active PT, and hence, decreases the average amount of harvested energy. This results from (6.19) where we showed that the value of P_{active} is a decreasing function of λ_E and the density of active PTs is $\lambda_R P_{\text{active}}$.

6.3.2 Performance metrics

Our objective in this section is two-fold: (i) study the effect of the deployment density of the RF-powered IoT devices on the performance of the primary network and (ii) study the effect of the guard zone radius r_g on the energy harvesting performance of the IoT devices. For the primary network, we consider two performance metrics to capture: (i) the connectivity between the PT and its associated PR and (ii) the secrecy of the transmitted confidential signal through studying its SINR at the illegitimate receivers (i.e., the IoT devices). We define these two performance metrics next.

Definition 6.3 (Probability of successful connection). *In order to ensure successful connection between the typical PT and PR, two conditions need to be satisfied: (i) the typical PT is active and (ii) the SINR at the typical PR is greater than the threshold β_R. Therefore, the probability of successful connection is*

$$P_{\text{con}}(r_g, \lambda_E) = \mathbb{P}(D_e \geq r_g, \text{SINR}_{\text{R}} \geq \beta_R), \tag{6.23}$$

where $\beta_R = 2^{\mathscr{R}_c} - 1$

The second performance metric for the primary network focuses on the secrecy of the transmitted signals when the PT is active, which is provided next.

Definition 6.4 (Secure communication probability). *Given that a PT is active, the probability that its transmitted data is perfectly secure is*

$$P_{\text{sec}}(r_g, \lambda_E) = \mathbb{E}\left[\mathbb{1}\left(\bigcap_{x_j \in \Phi_E} \text{SINR}_{\text{E}}(x_j) \leq \beta_E \middle| D_e \geq r_g\right)\right], \tag{6.24}$$

where $\beta_E = 2^{\mathscr{R}_c - \mathscr{R}_m} - 1$

The goal of the primary network is to maximize the value of P_{con} while ensuring that P_{sec} is above a predefined threshold ϵ. The parameter tuned to achieve this objective is r_g. Hence, the value of r_g selected by the primary network is

$$\begin{aligned} r_g^* &= \arg \max_{r_g \in \mathscr{G}(\lambda_E)} P_{\text{con}}(r_g, \lambda_E), \\ \mathscr{G}(\lambda_E) &= \{r_g : P_{\text{sec}}(r_g, \lambda_E) \geq \varepsilon\}. \end{aligned} \tag{6.25}$$

There are many metrics that have been used in the literature to study the performance of the energy harvesting aspect, for instance, the average amount of harvested energy $\mathbb{E}[E_H]$ or the energy coverage probability $\mathbb{P}(E_H \geq \mathscr{E})$. In this section, we focus on a modified version of the latter (defined next), which enables us to better understand the relation between the two coexisting networks, which is one of the main objectives of this section.

Definition 6.5 (Average density of successfully charged devices). *The average density of devices that successfully harvest at least $\mathscr{E}$ amount of energy is*

$$\bar{\lambda}_E = \lambda_E \mathbb{P}(E_H \geq \mathscr{E}). \tag{6.26}$$

The goal of the IoT network is to optimize the value of λ_E with the objective of maximizing the previous performance metric. While this may seem counter intuitive,

one can make sense of it (based on the discussion provided earlier) by observing that increasing the value of λ_E reduces the density of active PTs that are the main sources of RF-energy for the IoT devices.

6.3.3 Analysis and main results

We start our analysis by deriving the performance metrics for the primary network. In the next theorem, we provide an expression for the successful connection probability, introduced in Definition 6.3.

Theorem 6.2 (Probability of successful connection). *For a given value of λ_E, the probability of successful connection introduced in Definition 6.3 is*

$$P_{\text{con}}(r_g, \lambda_E) = \exp\left(- \left[\lambda_E \pi r_g^2 + \beta_R \frac{\sigma_R^2}{\rho} r_1^{\alpha} + \frac{2\pi^2 \lambda_R P_{\text{active}} \beta_R^{\frac{2}{\alpha}} r_1^2}{\alpha \sin\left(\frac{2}{\alpha}\pi\right)} \right] \right). \tag{6.27}$$

Proof. Recalling the expression for SINR_R given in (6.18), specifically the indicator function δ_i that indicates which interferer is active and which is silent, we concluded that the locations of active PTs can be modeled by PHP $\bar{\Phi}_R$ in (6.21). However, before using $\bar{\Phi}_R$ in our analysis, we need to make it clear that δ_i for different $y_i \in \Phi_R$ are correlated. This implicit correlation arises from the dependence of δ_i for all i on the PPP Φ_E. However, capturing this correlation in our analysis will significantly reduce the tractability of the results. Hence, this correlation will be ignored here. The accuracy of this approximation will be verified in Section 6.3.4. Now, revisiting the expression of P_{con} in Definition 6.3, we note that the correlation between δ_1 at the typical PT and δ_i values at each of the interferers in the expression of SINR_R is the only source of correlation between the events $(D_e \geq r_g)$ and $(\text{SINR}_\text{R} \geq \beta_R)$. Hence, ignoring this correlation, for the reasons stated earlier, will lead to the following:

$$P_{\text{con}} = \mathbb{P}(D_e \geq r_g)\mathbb{P}(\text{SINR}_\text{R} \geq \beta_R). \tag{6.28}$$

The first term in the previous expression represents $P_{\text{active}} = \exp\left(-\pi \lambda_R r_g^2\right)$ (please recall (6.19) where P_{active} was derived). To derive the second term in the previous expression, characterizing the statistics of the interference from a PHP-modeled network at a randomly located reference point (the typical PR) is required. However, ignoring the correlation between $\{\delta_i\}$ is equivalent to approximating the PHP $\bar{\Phi}_R$ with a PPP Ψ of equivalent density $\tilde{\lambda}_R = \lambda_R P_{\text{active}}$. Defining $I = \sum_{y_i \in \Psi} w_i \|y_i\|^{-\alpha}$, then:

$$\begin{aligned}
\mathbb{P}(\text{SINR}_\text{R} \geq \beta_R) &= \mathbb{P}\left(\frac{w_1 r_1^{-\alpha}}{I + \frac{\sigma_R^2}{\rho}} \geq \beta_R \right) \\
&\overset{(a)}{=} \mathbb{E}_I \left[\exp\left(-\beta_R \left(I + \frac{\sigma_R^2}{\rho} \right) r_1^{\alpha} \right) \right] \\
&= \exp\left(-\beta_R \frac{\sigma_R^2}{\rho} r_1^{\alpha} \right) \mathbb{E}_I \left[\exp\left(-\beta_R I r_1^{\alpha} \right) \right] \\
&\overset{(b)}{=} \exp\left(-\beta_R \frac{\sigma_R^2}{\rho} r_1^{\alpha} \right) \mathscr{L}_I \left(\beta_R r_1^{\alpha} \right),
\end{aligned} \tag{6.29}$$

where $w_1 \sim \exp(1)$ leads to step (a), and in step (b) we use the definition of Laplace transform of I which is $\mathscr{L}_I(s) = \mathbb{E}[\exp(-sI)]$. The Laplace transform of the interference in PPP is a well-established result in the literature [4]. For completeness, its derivation is provided next.

$$\begin{aligned}
\mathscr{L}_I(s) &= \mathbb{E}_{\Psi,\{w_i\}}\left[\exp\left(-s\sum_{y_i\in\Psi} w_i\|y_i\|^{-\alpha}\right)\right] \\
&= \mathbb{E}_{\Psi,\{w_i\}}\left[\prod_{y_i\in\Psi}\exp\left(-sw_i\|y_i\|^{-\alpha}\right)\right] \\
&\overset{(c)}{=} \mathbb{E}_{\Psi}\left[\prod_{y_i\in\Psi}\frac{1}{1+s\|y_i\|^{-\alpha}}\right] \\
&\overset{(d)}{=} \exp\left(-\tilde{\lambda}_R\int_{y\in\mathbb{R}^2} 1-\frac{1}{1+s\|y\|^{-\alpha}}\mathrm{d}y\right) \\
&\overset{(e)}{=} \exp\left(-2\pi\tilde{\lambda}_R\int_0^{\infty}\frac{sr_y^{-\alpha}}{1+sr_y^{-\alpha}}r_y\mathrm{d}r_y\right) \\
&\overset{(f)}{=} \exp\left(-\frac{2\pi^2\tilde{\lambda}_R s^{\frac{2}{\alpha}}\csc\left(\frac{2\pi}{\alpha}\right)}{\alpha}\right),
\end{aligned} \tag{6.30}$$

where knowing that the set of fading gains w_i are i.i.d. with $w_i \sim \exp(1)$ leads to step (c), step (d) results from using the probability generating functional (PGFL) of PPP [40], step (e) results from converting to polar coordinates, and step (f) follows after some mathematical manipulations. Substituting (6.30) in (6.29) and then in (6.28) leads to the final result in Theorem 6.2. □

Remark 6.4. *The expression in Theorem 6.2 captures two important insights on the effect of r_g on P_{con}. Recalling the definition of P_{con}, we notice that successful connection requires (i) the typical PT being active (guard zone free of illegitimate receivers), and (ii) the SINR at the typical PR is above a predefined threshold β_R. The first condition gets harder to satisfy as the value of r_g increases, due to the difficulty of ensuring that the guard zone is free of ERs when its radius is large. This effect is captured in the first term inside the exponential in Theorem 6.2. For the second condition, namely, the SINR value, we observe that increasing the value of r_g decreases the density of active interferers, leading to higher values of* SINR_R. *Hence, increasing r_g makes it easier to satisfy the SINR condition. This is also captured in the third term inside the exponential in Theorem 6.2, implicitly in the expression of* P_{active}.

To further investigate the effect of r_g on P_{con}, we derive the value of $r_g = \hat{r}_g$ that maximizes P_{con} in the next theorem.

Theorem 6.3. *Defining* $\mathscr{A}_1 = \frac{2\pi^2\lambda_R\beta_R^{\frac{2}{\alpha}}r_1^2}{\alpha\sin(\frac{2}{\alpha}\pi)}$, *we have:*

- *If* $\mathscr{A}_1 \leq 1$, *then* P_{con} *is a decreasing function of* r_g, *and* $\hat{r}_g = 0$.
- *If* $\mathscr{A}_1 > 1$, *then* $\hat{r}_g = \sqrt{\frac{\ln(\mathscr{A}_1)}{\pi\lambda_E}}$.

Proof. From the expression of P_{con} in Theorem 6.2, we note that it can be rewritten as a function of $P_{\text{active}} = \exp(-\lambda_E\pi r_g^2)$ as follows:

$$P_{\text{con}} = \exp\left(-\beta_R\frac{\sigma_R^2}{\rho}r_1^{\alpha}\right)P_{\text{active}}\exp(-P_{\text{active}}\mathscr{A}_1). \tag{6.31}$$

To get more information about the behavior of P_{con} against P_{active}, we compute the first derivative (with respect to P_{active}). Given that P_{active} is a decreasing function of r_g (recall (6.19)), we conclude the following:

1. If $1 - \mathscr{A}_1 P_{\text{active}} \geq 0$, then P_{con} is a decreasing function of r_g.
2. If $1 - \mathscr{A}_1 P_{\text{active}} < 0$, then P_{con} is an increasing function of r_g.

Consequently, we can infer that since $0 \leq P_{\text{active}} \leq 1$, P_{con} is a decreasing function of r_g as long as $\mathscr{A}_1 \leq 1$. In the case of $\mathscr{A}_1 \geq 1$, the relation between P_{con} and r_g can be explained as follows: (i) P_{con} is an increasing function of r_g as long as $P_{\text{active}} \geq \frac{1}{\mathscr{A}_1}$ (or $r_g \leq \sqrt{\frac{\ln(\mathscr{A}_1)}{\lambda_E\pi}}$) and (ii) P_{con} is a decreasing function of r_g as long as $P_{\text{active}} \leq \frac{1}{\mathscr{A}_1}$ (or $r_g \geq \sqrt{\frac{\ln(\mathscr{A}_1)}{\lambda_E\pi}}$). This concludes the proof. □

Remark 6.5. *Consistent with the intuition, we observe from the previous theorem that when* $\mathscr{A}_1 \leq 1$, *the effect of* r_g *on the event* $D_E \geq r_g$ *dominates its effect on the density of interferers, because of which* P_{con} *is a decreasing function of* r_g. *This is because* $\mathscr{A}_1$ *is an increasing function of each of* λ_R, r_1, *and* β_R. *At lower values of* $\mathscr{A}_1$, *one or more of these parameters are small enough to ignore the effect of the interference level on* P_{con}.

Now we derive the secure communication probability, introduced in Definition 6.4, in the following theorem.

Theorem 6.4 (Secure communication probability). *For a given value of* r_g *and* λ_E, *the probability of secure communication is*

$$P_{\text{sec}}(r_g, \lambda_E) = \exp\left(-2\pi\lambda_E\int_{r_g}^{\infty}\exp\left(-\frac{\sigma_E^2\beta_E r_x^{\alpha}}{\rho}\right)\mathscr{L}_{I_2}(\beta_E r_x^{\alpha})r_x\mathrm{d}r_x\right), \tag{6.32}$$

where $\mathscr{L}_{I_2}(s) = \exp\left(-2\pi\lambda_R P_{\text{active}}\int_0^{sr_g^{-\alpha}}\frac{s^{\frac{2}{\alpha}}}{\alpha(1+z)z^{\frac{2}{\alpha}}}\mathrm{d}z\right)$.

Proof. From Definition 6.4 of P_{sec}, we observe that we need to jointly analyze the values of $\text{SINR}_{\text{E}}(x_j)$ at all the locations $x_j \in \Phi_E$. Despite the usual assumption throughout most of the stochastic geometry-based literature on secrecy analysis that these values are uncorrelated, this is actually not precise. The reason for that is the dependence of $\text{SINR}_{\text{E}}(x_j)$, by definition, on the PPP Φ_R for all $x_j \in \Phi_E$. Some recent works focused on characterizing the correlation between interference levels at different locations [52]. However, most of these works focus on characterizing the correlation between only two locations assuming the knowledge of the distance between them. Unfortunately, these results are not directly applicable to our analysis. Hence, aligning with the existing literature, we will ignore this correlation in our analysis with the knowledge that this will provide an approximation. Furthermore, the accuracy of this approximation is expected to get worse as the value of λ_E increases. This is due to the fact that the distances between ERs decrease as λ_E increases, which was shown in [52] to increase the correlation. For notational simplicity, and without any loss of generality due to the stationarity of PPP, we will assume that the typical PT is placed at the origin, i.e., $y_1 = o$, in the rest of this proof. All the analyses provided in this section are conditioned on the event $D_e \geq r_g$. Following the same approach as in the proof of Theorem 6.2 of approximating the PHP $\bar{\Phi}_R$ with a PPP Ψ, and defining $I_2(x_j) = \sum_{y_i \in \Psi \setminus y_1} g_{i,j} \|y_i - x_j\|^{-\alpha}$, P_{sec} can be derived as follows:

$$
\begin{aligned}
P_{\text{sec}} &= \mathbb{E}_{\Phi_E, I_2, \{g_{1,j}\}} \left[\mathbb{1} \left(\bigcap_{x_j \in \Phi_E} \frac{g_{1,j} \|x_j\|^{-\alpha}}{I_2(x_j) + \frac{\sigma_E^2}{\rho}} \leq \beta_E \right) \right] \\
&\overset{\text{(g)}}{=} \mathbb{E}_{\Phi_E, I_2, \{g_{1,j}\}} \left[\prod_{x_j \in \Phi_E} \mathbb{1} \left(\frac{g_{1,j} \|x_j\|^{-\alpha}}{I_2(x_j) + \frac{\sigma_E^2}{\rho}} \leq \beta_E \right) \right] \\
&\overset{\text{(h)}}{=} \mathbb{E}_{\Phi_E, I_2} \left[\prod_{x_j \in \Phi_E} \left(1 - \exp\left(- \frac{\beta_E \left(I_2(x_j) + \frac{\sigma_E^2}{\rho} \right)}{\|x_j\|^{-\alpha}} \right) \right) \right] \\
&\overset{\text{(i)}}{=} \mathbb{E}_{\Phi_E} \left[\prod_{x_j \in \Phi_E} \left(1 - \exp\left(- \frac{\beta_E \left(\frac{\sigma_E^2}{\rho} \right)}{\|x_j\|^{-\alpha}} \right) \mathbb{E}\left[\exp\left(- \frac{\beta_E I_2(y_j)}{\|x_j\|^{-\alpha}} \right) \right] \right) \right],
\end{aligned}
\tag{6.33}
$$

where step (g) (and step (i)) follow from assuming that the values of $\text{SINR}_{\text{E}}(x_j)$ (and $I_2(x_j)$) are uncorrelated. Step (h) is due to assuming the set of fading gains $\{g_{1,j}\}$ to be i.i.d. with $g_{1,j} \sim \exp(1)$. Defining the Laplace transform of $I_2(x_j)$ by $\mathscr{L}_{I_2(x_j)}(s) = \mathbb{E}[\exp(-sI_2(x_j))]$, we note that there is only one difference in the derivation of $\mathscr{L}_{I_2(x_j)}(s)$ compared to that of $\mathscr{L}_I(s)$ in the proof of Theorem 6.2. The difference is in the reference point from where we are observing the interference. In the proof of Theorem 6.2, the reference point was the PR that does not have a minimum distance from any active interfering PT. In the current derivation, the reference point is an IoT device that has a minimum distance of r_g from any active interfering PT. Hence, the derivation

of $\mathscr{L}_{I_2(x_j)}(s)$ will be exactly the same as in (6.30) until step (e), where the minimum distance effect will appear in the lower limit of the integral as follows:

$$\mathscr{L}_{I_2(x_j)}(s) = \exp\left(-2\pi\tilde{\lambda}_R \int_{r_g}^{\infty} \frac{s r_y^{-\alpha}}{1 + s r_y^{-\alpha}} r_y \mathrm{d}r_y\right). \tag{6.34}$$

Note that the previous expression is not a function of x_j, so we drop it from the notation of Laplace transform. The final expression for $\mathscr{L}_{I_2}(s)$ as provided in Theorem 6.4 follows after simple mathematical manipulations. Substituting (6.34) in (6.33), we get

$$\begin{aligned} P_{\text{sec}} &= \mathbb{E}_{\Phi_E}\left[\prod_{x_j \in \Phi_E}\left(1 - \exp\left(-\beta_E\left(\frac{\sigma_E^2}{\rho}\right)\|x_j\|^{\alpha}\right)\mathscr{L}_{I_2}\left(\beta_E\|x_j\|^{\alpha}\right)\right)\right] \\ &\overset{(k)}{=} \exp\left(-2\pi\lambda_E \int_{x \in \mathbb{R}^2 \cap \mathscr{B}(o,r_g)} \exp\left(-\beta_E\left(\frac{\sigma_E^2}{\rho}\right)\|x\|^{\alpha}\right)\mathscr{L}_{I_2}(\beta_E\|x\|^{\alpha})\,\mathrm{d}x\right), \end{aligned} \tag{6.35}$$

where step (k) results from applying PGFL of PPP, and the integration is over $y \in \mathbb{R}^2 \cap \mathscr{B}(o, r_g)$ because the analysis in this section is conditioned on the event $D_e \geq r_g$, which means that the typical PT is active. Since we assumed that the typical PT is placed at the origin in this derivation, the ball $\mathscr{B}(o, r_g)$ is clear of ERs. Converting from Cartesian to polar coordinates leads to the final result in Theorem 6.4. □

Remark 6.6. *As stated in the earlier remarks, the value of r_g has a significant effect on the density of active interferers. Hence, the total interference at any receiver, legitimate or not, decreases as we increase the value of r_g. This is captured in the Laplace transform term of the expression derived in Theorem 6.4. In addition, since transmitting signals are already conditioned on the guard zone being free of the illegitimate receivers, there is a minimum distance r_g between the typical PT and the nearest illegitimate receiver. Hence, increasing the value of r_g reduces the quality of the confidential signal transmitted by the PT at the illegitimate receivers. This is captured in Theorem 6.4 in the integration interval, which decreases as we increase r_g. This trade-off in the effect of r_g on the secure communication probability will be further investigated and visualized with the aid of numerical results in the next subsection.*

In the following theorem, we provide the main performance metric for the IoT devices, which is the density of successfully charged devices.

Theorem 6.5 (Density of successfully charged devices). *For a given value of r_g and λ_E, the density of successfully charged IoT devices is*

$$\bar{\lambda}_E = \lambda_E \int_{r_g}^{\infty} 2\pi\lambda_R P_{\text{active}} r_p \exp\left(-\pi\lambda_R P_{\text{active}}\left(r_p^2 - r_g^2\right) - \frac{\mathscr{E} r_p^{\alpha}}{\rho\eta}\right)\mathrm{d}r_p. \tag{6.36}$$

Proof. The density of successfully charged IoT devices can be derived as follows:

$$P_{\text{energy}} = \lambda_E \mathbb{P}\left(\eta\rho R_p^{-\alpha} g \geq \mathscr{E}\right) \overset{\text{(l)}}{=} \mathbb{E}_{R_p}\left[\exp\left(-\frac{\mathscr{E} R_p^{\alpha}}{\eta\rho}\right)\right], \tag{6.37}$$

where R_p is the distance between the ER and its nearest active PT, and step (l) is due to $g \sim \exp(1)$. The distance R_p represents the contact distance of a PHP observed from a hole center. Unfortunately, the exact distribution of this distance is unknown. However, the approach of approximating the PHP $\bar{\Phi}_R$ with a PPP Ψ is known to provide fairly tight approximation of the contact distance distribution of PHP [53]. Given that the nearest active PT to the ER is at a distance of at least r_g, the distribution of R_p is

$$f_{R_p}(r_p) = 2\pi\tilde{\lambda}_R \exp\left(-\pi\tilde{\lambda}_R (r_p - r_g)^2\right), \quad r_p \geq r_g. \tag{6.38}$$

Using this distribution to compute the expectation in (6.37) leads to the final result in Theorem 6.5. □

6.3.4 *Numerical results and discussion*

In this section, unless otherwise specified, we use the following values for the simulation parameters: $\eta = 0.75$, $\varepsilon = 0.9$, $\lambda_R = 10^{-1}$ m^{-2}, $\rho = 1$, $T = 1$, $\alpha = 4$, $\beta_R = 3$ dB, $\beta_E = 0$ dB, and $\mathscr{E} = 1$ μJ. We also refer to the SNR value at the IoT devices as $\gamma = \frac{\rho}{\sigma_E^2}$.

In Figure 6.4, we plot the successful connection probability for different values of r_g. Given that the simulation setup captures a scenario with $\mathscr{A}_1 < 1$ (introduced

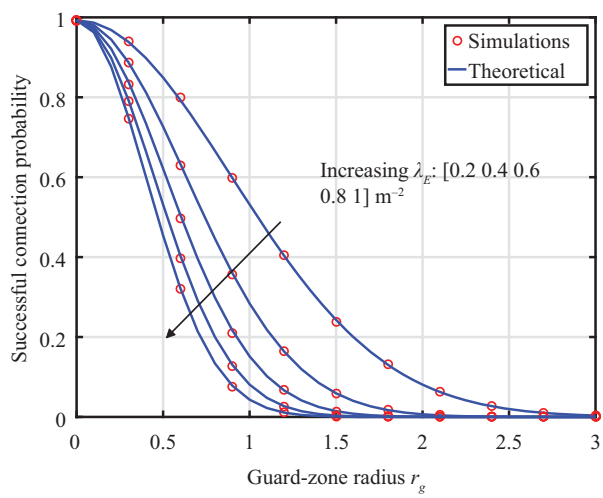


Figure 6.4 Successful connection probability against different values of r_g

in Theorem 6.3), P_{con} is a decreasing function of r_g. This is consistent with our observations in Theorem 6.3 and the following remarks. Hence, recalling (6.25), we conclude that for the considered simulation setup, r_g^* is the minimum value of r_g that ensures $P_{\text{sec}} \geq \varepsilon$.

In Figure 6.5, we plot the secure communication probability for different values of r_g. The effect of r_g on the secure communication probability, discussed in Theorem 6.4 and the following remarks, can be observed in Figure 6.5. At low values of r_g, the performance is dominated by the high interference at the illegitimate receivers, due to the high density of the active PTs. This reflects the stronger effect of λ_S on P_{sec} at lower values of r_g. At high values of r_g, despite the low density of active PTs, the performance is dominated by the large distance between the PT and its nearest illegitimate receiver. We observe the existence of a minimum value for P_{sec}, at which the value of r_g is not small enough to result in high level of interference at the illegitimate receivers, nor is it high enough to result in large distance between the active PT and its nearest illegitimate receiver. We also observe that increasing the value of λ_E increases the value of r_g^* that ensures $P_{\text{sec}} = \varepsilon$, where $\varepsilon = 0.9$.

In Figure 6.6, we study the effect of $\gamma = \frac{\rho}{\sigma_E^2}$ on the value of P_{sec}. Unlike λ_E, the effect of γ on P_{sec} is more prominent at higher values of r_g. As stated earlier, this is due to the dominance of the interference on the performance at lower values of r_g, leading to negligible effect of γ in this regime.

In Figure 6.7, we plot the density of successfully charged devices for different values of λ_E. As reported earlier, there exists an optimal value of λ_E that maximizes this density. Furthermore, we observe that increasing the value of r_g has a clear negative effect on the performance of the IoT devices. This can be noticed from the decrease in the value of the maximum achievable density of successfully charged devices as we increase the value of r_g.

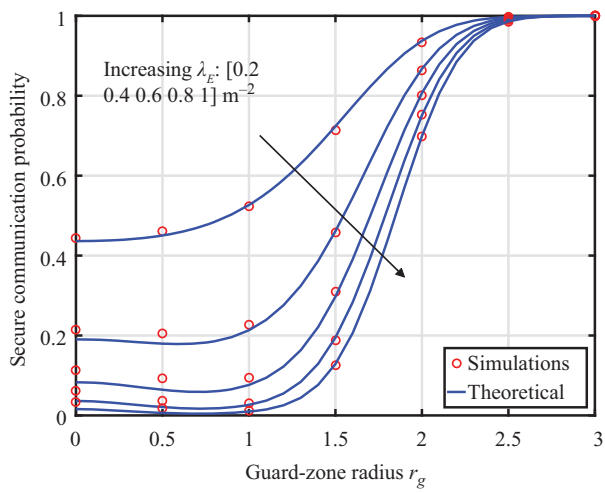


Figure 6.5 Secure communication probability against different values of r_g and λ_E

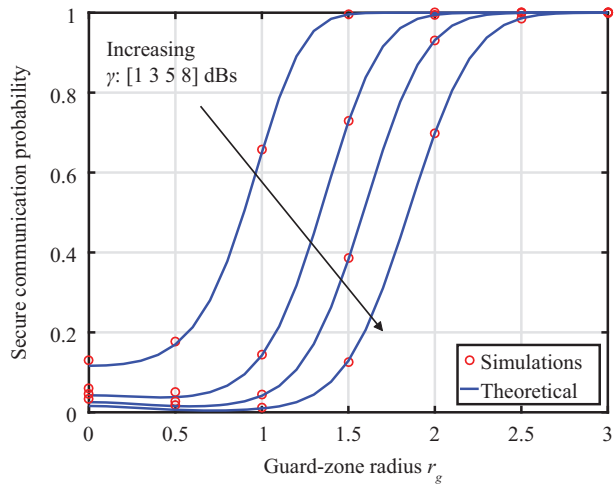


Figure 6.6 Secure communication probability against different values of r_g and γ

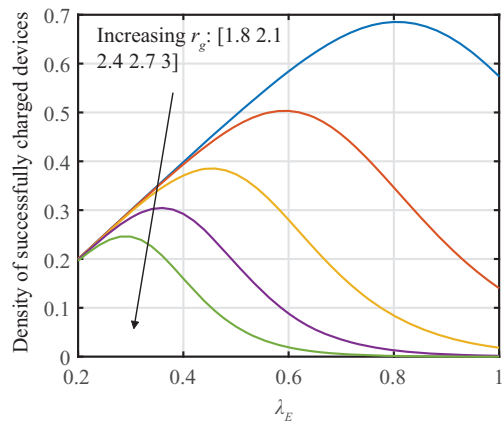


Figure 6.7 Density of successfully charged devices against different values of λ_E and r_g

We can observe from Figures 6.5 and 6.7 that the parameter selection of the two networks (r_g selection by the primary network and λ_E selection by the IoT devices) is intertwined. This interaction can actually be modeled as a two-player noncooperative game. The Nash equilibrium of such game would capture a state where the primary network has no incentive to change the value of r_g given the value of λ_E, and the IoT devices have no incentive to change the value of λ_E given the value of r_g. However, the main complexity of analyzing such a game arises from the relatively complicated expressions of P_{sec} and $\bar{\lambda}_E$, because of which it is challenging to prove the existence

of a Nash equilibrium. This actually is one of the main challenges that arise whenever the objective is to optimize an expression derived using stochastic geometry tools. Interested readers are advised to refer to [19] for more details.

6.4 Summary

This chapter focused on the application of stochastic geometry to the performance analysis of RF-powered IoT networks. In particular, we focused on two scenarios of general interest: (i) the same wireless network provides connectivity and RF charging to the IoT devices and (ii) IoT devices rely on a coexisting, secrecy-enhancing network for harvesting RF-energy. In the first part of this chapter, we considered an IoT network that relies on the cellular infrastructure for communication and RF charging. For this setup, we studied the joint probability of harvesting sufficient energy and maintaining sufficiently high downlink SINR. We proposed a dominant-interferer approximation that enabled the derivation of the joint probability and resulted in several system-level insights. One of the main insights obtained from this analysis is the existence of an optimal charging slot duration that maximizes the downlink average throughput. We also derived a *tuning parameter* that captures the effect of the system parameters on the system performance, such as the density of the BSs.

In the second part of this chapter, we considered an RF-powered IoT network coexisting with a secrecy-enhanced primary network. The IoT network relies on the RF transmissions of the primary network for RF charging. The primary network is assumed to use the guard zone technique. This technique maintains a minimum distance between any active PT and its nearest IoT device. For that setup, we derived the secrecy performance metrics of the primary network and the energy harvesting performance metrics of the IoT network. We showed that the performance of both networks is correlated, because both secrecy and energy harvesting performance metrics depend on the deployment density of the IoT devices and the guard-zone radius of the primary network. A useful insight obtained from this study is the existence of an optimal deployment density for the IoT network that maximizes the density of successfully charged devices.

Acknowledgment

The support of the US NSF (Grants CPS-1739642 and CNS-1814477) is gratefully acknowledged.

References

[1] Dhillon HS, Huang H, and Viswanathan H. Wide-Area Wireless Communication Challenges for the Internet of Things. IEEE Communications Magazine. 2017;55(2):168–174.

[2] Zhou X, Ganti RK, Andrews JG, *et al.* On the Throughput Cost of Physical Layer Security in Decentralized Wireless Networks. IEEE Transactions on Wireless Communications. 2011;10(8):2764–2775.

[3] Pinto PC, Barros J, and Win MZ. Secure Communication in Stochastic Wireless Networks—Part I: Connectivity. IEEE Transactions on Information Forensics and Security. 2012;7(1):125–138.

[4] Andrews JG, Gupta AK, and Dhillon HS. A Primer on Cellular Network Analysis Using Stochastic Geometry; 2016, Available online: arxiv.org/abs/1604.03183.

[5] ElSawy H, Hossain E, and Haenggi M. Stochastic Geometry for Modeling, Analysis, and Design of Multi-tier and Cognitive Cellular Wireless Networks: A Survey. IEEE Communications Surveys and Tutorials. 2013;15(3):996–1019.

[6] Mukherjee S. Analytical Modeling of Heterogeneous Cellular Networks: Geometry, Coverage, and Capacity. New York: Cambridge University Press; 2014.

[7] ElSawy H, Sultan-Salem A, Alouini MS, *et al.* Modeling and Analysis of Cellular Networks Using Stochastic Geometry: A Tutorial. IEEE Communications Surveys Tutorials. 2017 First quarter;19(1):167–203.

[8] Flint I, Lu X, Privault N, *et al.* Performance Analysis of Ambient RF Energy Harvesting With Repulsive Point Process Modeling. IEEE Transactions on Wireless Communications. 2015;14(10):5402–5416.

[9] Zhong C, Chen X, Zhang Z, *et al.* Wireless-Powered Communications: Performance Analysis and Optimization. IEEE Transactions on Communications. 2015;63(12):5178–5190.

[10] Huang K and Lau VK. Enabling Wireless Power Transfer in Cellular Networks: Architecture, Modeling and Deployment. IEEE Transactions on Wireless Communications. 2014;13(2):902–912.

[11] Sakr AH and Hossain E. Analysis of K-Tier Uplink Cellular Networks With Ambient RF Energy Harvesting. IEEE Journal on Selected Areas in Communications. 2015;33(10):2226–2238.

[12] Che YL, Duan L, and Zhang R. Spatial Throughput Maximization of Wireless Powered Communication Networks. IEEE Journal on Selected Areas in Communications. 2015;33(8):1534–1548.

[13] Krikidis I. Simultaneous Information and Energy Transfer in Large-Scale Networks With/Without Relaying. IEEE Transactions on Communications. 2014;62(3):900–912.

[14] Zhou X, Zhang R, and Ho CK. Wireless Information and Power Transfer: Architecture Design and Rate-Energy Tradeoff. IEEE Transactions on Communications. 2013;61(11):4754–4767.

[15] Zhou S, Chen T, Chen W, *et al.* Outage Minimization for a Fading Wireless Link with Energy Harvesting Transmitter and Receiver. IEEE Journal on Selected Areas in Communications. 2015;33(3):496–511.

[16] Kishk MA and Dhillon HS. Downlink Performance Analysis of Cellular-Based IoT Network With Energy Harvesting Receivers. In: IEEE GLOBECOM; 2016.

[17] Kishk MA and Dhillon HS. Modeling and Analysis of Ambient RF Energy Harvesting in Networks With Secrecy Guard Zones. In: 2017 IEEE Wireless Communications and Networking Conference (WCNC); 2017. pp. 1–6.
[18] Kishk MA and Dhillon HS. Joint Uplink and Downlink Coverage Analysis of Cellular-Based RF-Powered IoT Network. IEEE Transactions on Green Communications and Networking. 2018;2(2):446–459.
[19] Kishk MA and Dhillon HS. Coexistence of RF-Powered IoT and a Primary Wireless Network With Secrecy Guard Zones. IEEE Transactions on Wireless Communications. 2018;17(3):1460–1473.
[20] Abd-Elmagid MA, Kishk MA, and Dhillon HS. Joint Energy and SINR Coverage in Spatially Clustered RF-Powered IoT Network. IEEE Transactions on Green Communications and Networking. 2019;3(1):132–146.
[21] Abd-Elmagid MA, Kishk MA, and Dhillon HS. Coverage Analysis of Spatially Clustered RF-Powered IoT Network. In: Proc., IEEE ICC; 2018. pp. 1–7.
[22] Kishk MA and Dhillon HS. A Comparison of Secrecy Enhancing Techniques for Secure SWIPT System. In: Proc., IEEE ICC workshops; 2018. pp. 1–6.
[23] Ganesan K, Grover P, and Rabaey J. The Power Cost of Over-designing Codes. In: Proc., IEEE Workshop on Signal Processing Systems (SiPS); Oct 2011. pp. 128–133.
[24] Blake CG and Kschischang FR. Energy Consumption of VLSI Decoders. IEEE Transactions on Information Theory. 2015;61(6):3185–3198.
[25] Grover P. Bounds on the Tradeoff Between Decoding Complexity and Rate for Sparse-Graph Codes. In: Proc., IEEE Information Theory Workshop (ITW); 2007. pp. 196–201.
[26] Di Renzo M and Lu W. System-Level Analysis and Optimization of Cellular Networks With Simultaneous Wireless Information and Power Transfer: Stochastic Geometry Modeling. IEEE Transactions on Vehicular Technology. 2017;66(3):2251–2275.
[27] Khan TA, Alkhateeb A, and Heath RW. Millimeter Wave Energy Harvesting. IEEE Transactions on Wireless Communications. 2016;15(9):6048–6062.
[28] Li Q, Ma WK, and So AMC. Robust Artificial Noise-Aided Transmit Optimization for Achieving Secrecy and Energy Harvesting. In: Proc., IEEE Intl. Conf. on Acoustics, Speech, and Sig. Proc. (ICASSP); 2014. pp. 1596–1600.
[29] Xing H, Liu L, and Zhang R. Secrecy Wireless Information and Power Transfer in Fading Wiretap Channel. IEEE Transactions on Vehicular Technology. 2016;65(1):180–190.
[30] Liu M and Liu Y. Power Allocation for Secure SWIPT Systems With Wireless-Powered Cooperative Jamming. IEEE Communications Letters. 2017;21(6):1353–1356.
[31] Liu L, Zhang R, and Chua KC. Secrecy Wireless Information and Power Transfer With MISO Beamforming. IEEE Transactions on Signal Processing. 2014;62(7):1850–1863.

[32] Ren Y, Lv T, Gao H, *et al.* Secure Wireless Information and Power Transfer in Heterogeneous Networks. IEEE Access. 2017;5:4967–4979..

[33] Liu Y, Wang L, Zaidi SAR, *et al.* Secure D2D Communication in Large-Scale Cognitive Cellular Networks: A Wireless Power Transfer Model. IEEE Transactions on Communications. 2016;64(1):329–342.

[34] Weber S, Andrews JG, and Jindal N. The Effect of Fading, Channel Inversion, and Threshold Scheduling on Ad Hoc Networks. IEEE Transactions on Information Theory. 2007;53(11):4127–4149.

[35] Madhusudhanan P, Restrepo JG, Liu Y, *et al.* Downlink Performance Analysis for a Generalized Shotgun Cellular System. IEEE Transactions on Wireless Communications. 2014;13(12):6684–6696.

[36] Schloemann J, Dhillon HS, and Buehrer RM. Toward a Tractable Analysis of Localization Fundamentals in Cellular Networks. IEEE Transactions on Wireless Communications. 2016;15(3):1768–1782.

[37] Bhandari T, Dhillon HS, and Buehrer RM. The Impact of Proximate Base Station Measurements on Localizability in Cellular Systems. In: Proc., IEEE SPAWC; 2016.

[38] Chetlur VV and Dhillon HS. Downlink Coverage Probability in a Finite Network of Unmanned Aerial Vehicle (UAV) Base Stations. In: Proc., IEEE SPAWC; 2016.

[39] Chetlur VV and Dhillon HS. Downlink Coverage Analysis for a Finite 3-D Wireless Network of Unmanned Aerial Vehicles. IEEE Transactions on Communications. 2017;65(10):4543–4558.

[40] Haenggi M. Stochastic Geometry for Wireless Networks. Cambridge University Press; 2012.

[41] Moltchanov D. Distance Distributions in Random Networks. Ad Hoc Networks. 2012;10(6):1146–1166.

[42] Khandaker MR and Wong KK. Robust Secrecy Beamforming With Energy-Harvesting Eavesdroppers. IEEE Wireless Communications Letters. 2015;4(1):10–13.

[43] Ng DWK, Lo ES, and Schober R. Robust Beamforming for Secure Communication in Systems With Wireless Information and Power Transfer. IEEE Transactions on Wireless Communications. 2014;13(8):4599–4615.

[44] Xu J, Liu L, and Zhang R. Multiuser MISO Beamforming for Simultaneous Wireless Information and Power Transfer. IEEE Transactions on Signal Processing. 2014;62(18):4798–4810.

[45] Chae SH, Choi W, Lee JH, *et al.* Enhanced Secrecy in Stochastic Wireless Networks: Artificial Noise With Secrecy Protected Zone. IEEE Transactions on Information Forensics and Security. 2014;9(10):1617–1628.

[46] Liu W, Ding Z, Ratnarajah T, *et al.* On Ergodic Secrecy Capacity of Random Wireless Networks With Protected Zones. IEEE Trans. on Vehicular Technology. 2016;65(8):6146–6158.

[47] Zhou X and McKay MR. Secure Transmission With Artificial Noise Over Fading Channels: Achievable Rate and Optimal Power Allocation. IEEE Transactions on Vehicular Technology. 2010;59(8):3831–3842.

[48] Zhang M, Liu Y, and Zhang R. Artificial Noise Aided Secrecy Information and Power Transfer in OFDMA Systems. IEEE Transactions on Wireless Communications. 2016;15(4):3085–3096.

[49] Kishk MA and Dhillon HS. Stochastic Geometry-Based Comparison of Secrecy Enhancement Techniques in D2D Networks. IEEE Wireless Communications Letters. 2017;6(3):394–397.

[50] Yazdanshenasan Z, Dhillon HS, Afshang M, *et al.* Poisson Hole Process: Theory and Applications to Wireless Networks. IEEE Transactions on Wireless Communications. 2016;15(11):7531–7546.

[51] Lee Ch and Haenggi M. Interference and Outage in Poisson Cognitive Networks. IEEE Transactions on Wireless Communications. 2012;11(4):1392–1401.

[52] Krishnan S and Dhillon HS. Spatio-Temporal Interference Correlation and Joint Coverage in Cellular Networks. IEEE Transactions on Wireless Communications. 2017;16(9):5659–5672.

[53] Kishk MA and Dhillon HS. Tight Lower Bounds on the Contact Distance Distribution in Poisson Hole Process. IEEE Wireless Communications Letters. 2017;6(4):454–457.

Chapter 7

Backscatter communications for ultra-low-power IoT: from theory to applications

Seung-Woo Ko[1], Kaifeng Han[2], Bruno Clerckx[3] and Kaibin Huang[4]

Internet-of-Things (IoT) is expected to connect tens of billions of devices anytime and anywhere and enable a wide range of services such as smart city, connected vehicles, and health care [1]. Recent advancements have driven the rapid growth of IoT in 5G technologies along with cloud- and edge-computing-enabled big-data analytics. However, one typical drawback of the existing IoT solution is the limited lifetime due to the massive number of IoT devices being powered by batteries with finite capacities. Therefore, keeping a large number of energy-constrained IoT devices alive poses a key design challenge. To this end, Backscatter Communication (BackCom) has emerged as a promising technology, allowing IoT devices to transmit data with low-power consumption. Moreover, its low-complexity design and small form factor make BackCom more attractive by realizing cost-effective IoT deployment. We organize the remainder of this chapter as follows. In Section 7.1, we provide fundamental knowledge for BackCom, including the basic principles, key design parameters, and standardization. Then, we summarize several BackCom networks in Section 7.2 and introduce several advanced emerging communication technologies redesigned for BackCom in Section 7.3. In Section 7.4, we explain several performance improvement methods of BackCom. Next, we focus on the applications empowered by BackCom in Section 7.5. Last, we discuss the open issues and future directions of BackCom in Section 7.6.

7.1 BackCom basic principle

This subchapter aims at introducing critical principles of BackCom, which are required to understand more advanced designs explained in the sequel. We first

[1]Division of Electronics and Electrical Information Engineering, Korea Maritime and Ocean University, Busan, South Korea
[2]Institute of Policy and Economical Research, China Academy of Information and Communications Technology, Beijing, China
[3]Electrical and Electronics Engineering Department, Imperial College London, London, UK
[4]Department of Electrical and Electronics Engineering, The University of Hong Kong, Hong Kong, Hong Kong, China

illustrate the architecture and basic operations of BackCom, including modes and modulation. Next, we deal with the issues of determining several parameters for performance enhancement. Last, we review the current standardization of BackCom.

7.1.1 Architecture

Although there are various types of BackCom, all of them are based on a fundamental architecture consisting of two entities, a tag and a reader [2]. The tag, which is a passive device without any active component, comprises an energy harvester, a battery, an information decoder, and a modulation block, as shown in Figure 7.1. On the other hand, the reader can transmit and receive a signal with its power supply and has active Radio Frequency (RF) components to generate a single-tone sinusoidal Continuous Wave (CW). The tag reflects a portion of the incident sinusoidal CW radiated from the reader, while it harvests the remaining CW for powering the on-tag Integrated Circuit (IC). The IC modulates data onto the reflected CW by controlling impedance matching between the antenna and the load in the tag. Specifically, a reflection coefficient Γ is defined as the ratio of the incoming and reflected signal. It is given as $\Gamma = \frac{Z_L - Z_A}{Z_L + Z_A}$, where Z_L and Z_A represent the impedances of the load and the antenna, respectively. The tag has multiple load impedances each of which is all different. Each load impedance follows a different reflection coefficient, and it thus corresponds to a different symbol of the backscatter transmission. We call such a procedure a backscatter modulation.

7.1.2 Modes and modulation

Given the tag–reader pair, there are two kinds of modes: a forward information transmission from the reader to the tag and backward information transmission from the tag to the reader [3].

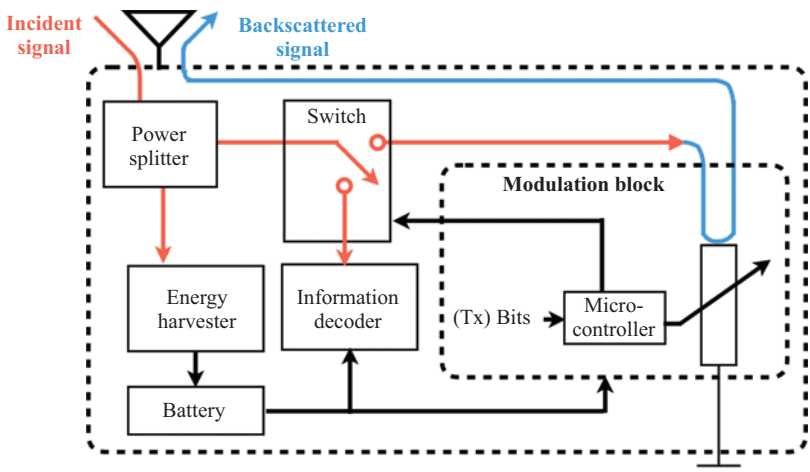


Figure 7.1 The architecture of a backscatter tag

For forward information transmission, the reader transmits a modulated signal to the tag. This signal is delivered to the tag's information decoder, and it is translated into a binary sequence based on On–Off Keying (OOK) demodulation. Specifically, the decoded bit is 1 or 0 when the received energy is high or low, respectively. It is worth recalling that the reason why BackCom uses this primitive OOK is due to the hardware limitation without any power-hungry component, i.e., RF chains and oscillators.

For the backward information transmission, the CW radiated from the reader is delivered into the tag's modulation block. The tag's data stream is embedded onto the CW by switching over tag's impedances according to the stream, corresponding to forming a constellation. For example, in the case of two impedances, Binary Phase-Shift Keying (BPSK) modulation is available by choosing either one of the two depending on the bit the tag wants to send. A higher modulation order is possible if there are more tag's impedances. In this chapter, we focus on a tag with two loads unless specified, and the tag can embed binary data onto its reflected signal. One important design criterion in the mode of backward information transmission is the energy-rate trade-off. Specifically, a backscatter tag is typically designed to switch between inactive and active states. In the inactive state, the tag harvests a large portion of energy from the incident CW by matching its load impedance to that of the antenna, and its circuit remains inactive for energy conservation.* In the active state, on the other hand, the circuit becomes activated, and the tag's data is embedded into the backscattered signal via impedance mismatching. The fraction of harvested energy thus decreases. Define a duty cycle as the ratio of the duration for the active modes, which is a crucial design parameter to optimize the energy-rate trade-off [4].

The mode of backward information transmission is dominant in most common Radio Frequency IDentification (RFID) applications due to the asymmetricity of data traffics. For example, low-rate signaling data is delivered via the forward transmission mode, whereas high-rate information-bearing information is delivered via the backward information transmission mode. However, both of the modes are important in future IoT applications due to the need for more complex signaling and the increase of direct transmissions between the massive number of devices.

7.1.3 Design parameters

The performance of BackCom can be affected by many parameters. We use the maximum available distance R as an exemplified performance metric to explain such effects. According to Friis equation, the distance R can be expressed as

$$R = \frac{\lambda}{4\pi}\sqrt{\frac{P_t G_t(\theta,\varphi) G_r(\theta,\varphi)\rho\tau}{P_{\mathrm{TH}}}}. \tag{7.1}$$

*Note that the inactive mode does not mean that a reader is in silence, since it receives a certain amount of energy in the inactive mode due to a structure mode scattering depending on the antenna's geometry and material [2].

where λ is the wavelength, P_t is the reader's transmit power, $G_t(\theta, \varphi)$ and $G_r(\theta, \varphi)$ are gains of the transmit and receive antennas on the angle of (θ, φ), respectively. Moreover, P_{TH} is the minimum threshold power required to activate the tag's circuit, and ρ is the polarization efficiency. The parameter $\tau = |\Gamma|^2$ is the power transmission efficiency depending on the impedance matching between antenna and load impedances. We explain the effect of these parameters on the distance R as follows.

7.1.3.1 Operating frequency

An operating frequency, which is inversely proportional to wavelength λ, is an essential factor to determine the size of antennas. In the case of half-dipole antennas, for example, its size can be calculated as $\frac{\lambda}{2}$, i.e., 16 cm at 915 MHz and 2.5 cm at 5.79 GHz. The most common operating frequency band of RFID systems is Ultrahigh Frequency (UHF) ranging from 860 to 960 MHz. Recently, there exist some recent works in the literature suggesting backscatter systems operating in Superhigh Frequency, i.e., 2.4–2.5 GHz and 5.725–5.875 GHz for achieving higher data rate [5] as well as reducing its form factor. On the other hand, it leads to reducing the distance R according to (7.1).

7.1.3.2 Impedance matching

The impedance matching between the antenna and load can control the states of a backscatter tag between active and inactive ones. The complex load and antenna impedances are given as $Z_L = R_L + jX_L$ and $Z_A = R_A + jX_A$. Here, R_L and R_A are the load and antenna's resistance, respectively, and X_L and X_A are the load and antenna's reactance, respectively. In general, Z_L is not easy to switch due to its dependency on the operating frequency and the received power. Consequently, the change in antenna impedance Z_A is more suitable for controlling impedance matching. These two impedances are said to be perfectly matched when the transmission efficiency τ specified in (7.1) becomes one. The efficiency τ is given as

$$\tau = \frac{4R_L R_A}{|Z_L + Z_A|^2}. \tag{7.2}$$

As a result, the antenna impedance for perfect impedance matching can be easily calculated as $Z_L = Z_A^*$.

7.1.3.3 Antenna gain

Antenna gain is defined as the ratio of the received power in the specific direction to an isotropic source. In a typical communication system, an antenna with a higher gain is used to extend the transmission range, but it is an expensive solution. In the case of BackCom, whose target distance is small and the manufacturing cost is limited, a low-gain antenna is preferable. Another vital factor to affect the antenna gain is the on-board gain penalty, defined as the loss of antenna gain due to the material attachment [4]. It depends on multifold factors, including material properties, object

geometry, frequency, and antenna type, making it challenging to calculate tractably. The most common effective method to determine the on-board gain penalty is based on simulation and measurements [6].

7.1.3.4 Polarization

Polarization is the trajectory of an end point of the vector to represent the instantaneous electric field, describing the change of the field vector's direction and magnitude. The received power is maximized when the polarization of the incident wave is matched to that of the antenna. To this end, the antennas of the reader and tag should be located in parallel. If the reader's antenna is rotated by $\frac{\pi}{2}$, on the other hand, the received power becomes zero. The problem of the polarization is critical since the orientation is usually random. Several antenna designs have been suggested to cope with the polarization issue. For example, one can use a circular array in both of the reader and the tag [7] and two linearly polarized antennas with $\frac{\pi}{4}$ orientation to avoid the complete polarization mismatch [8].

7.1.4 Standardization

A standard, which is defined as the set of rules, conditions, or requirements that the components of a system must follow to operate effectively, is an essential issue in the area of BackCom. There exist two significant initiatives regarding BackCom standardization groups, International Standard Organization (ISO) [9], and EPCglobal [10], which have different approaches introduced next. ISO aims at creating technology-oriented standards for general use. It closely works with International Electrotechnical Commission [11] responsible for the standards of all electrical, electronic, and related technologies covering various issues, e.g., air interface (e.g., ISO 18000 series), data content (e.g., ISO 15418), and conformance and performance (e.g., ISO 18046, ISO 18047). Moreover, there exist a few separate standards developed for different applications such as tracking animals (e.g., ISO 11784, ISO 11785, and ISO 14223) and contactless cards (e.g., ISO 14443 and ISO 15693). Different ISO standards are summarized in Table 7.1.

EPCglobal tends to make application-specific standards such as Electronic Product Code (EPC), a unique code used for numbering items and identifying objects. EPCglobal also categorizes standards into different classes, which help vendors to manufacture tags with different capabilities and prices. A higher class tag can provide more functionalities than a lower class one. Specifically, there are six classes such as Class 0 and 1 for all passive tags, Class 2 for higher functionality tags enabling read/write, Class 3 for semi-passive tags with an additional power sources, Class 4 for active tags with the ability of communications between the tags in the same class, Class 5 for tags having capability of powering Class 1 and 2 tags and communicate with Class 3 tags. The classification of EPCglobal is summarized in Table 7.2.

It is worth noting that the two main initiatives cooperate to help the broader adoption of BackCom. For example, EPCglobal Class 1 Gen 2 [12], which is designed to work in UHF, is developed based on the compliance with UHF air interface protocol

Table 7.1 ISO RFID standards

Standard	Details
ISO 10536	Identification cards—contactless integrated circuit(s) cards
ISO 11784	RFID of animals—code structure
ISO 11785	RFID of animals—technical concept
ISO 14443	Cards and security devices for personal identification—contactless cards proximity
ISO 15418	Information technology—automatic identification and data capture techniques
ISO 15459	Information technology—unique identification
ISO 15693	Identification cards—contactless integrated circuit cards—vicinity cards
ISO 15961	RFID for item management—data protocol: application interface
ISO 15962	RFID for item management—data encoding rules and logical memory functions
ISO 15963	RFID for item management—unique identification of RF tag
ISO 18000	RFID for the air interface for a unique frequency range: Part 1: Reference architecture, Part 2: 135 kHz, Part 3: 13.56 MHz, Part 4: 2.45 GHz, Part 5: 5.8 GHz, Part 6: 860–960 MHz, Part 7: 433.92 MHz
ISO 18046	RFID performance test methods
ISO 18047	RFID conformance test methods
ISO 24710	RFID for item management—elementary tag licence plate functionality for ISO 18000 air interface definitions
ISO 24729	RFID for item management—implementation guidelines
ISO 24730	Information technology—Real-Time Locating System (RTLS)
ISO 24752	Information technology—user interfaces—universal remote console
ISO 24753	RFID for item management—encoding and processing rules for sensors and batteries
ISO 24769	RTLS device conformance test methods
ISO 24770	RTLS device performance test methods

Table 7.2 Classification used by EPCglobal for its tags

Class	Tag classification	Feature	Programming
0	"Read only" Passive		Programmed by the manufacturer
1	"Write once-read many" Passive		Programmed by the customer; cannot be reprogrammed
2	Rewritable Passive	A passive tag with up to 65 kB of read–write memory	Programmed by the customer; cannot be reprogrammed
3	Semi-passive	Similar to a Class 2 tag but with a built-in battery to support increased read range	Reprogrammable
4	Active	An active tag that transmits and runs its chip's circuitry with the use of a built-in battery	Reprogrammable
5	Active	An active tag that can communicate with other Class 5 tags and/or other devices or tags	Reprogrammable

in ISO 18000-6 [13]. Consequently, the standard has been widely used in various applications, including inventory tracking and pallet tracking.

7.2 BackCom networks

In this section, we introduce three most popular and representative types of BackCom networks based on their different network architectures, including the point-to-point BackCom network, multi-access BackCom network, and interference BackCom network. Different types of BackCom network architectures could be adopted in and enable various applications in IoT scenarios.

7.2.1 BackCom networks

The simplest and basic setup for BackCom is the point-to-point network that can be further classified into two major types: the monostatic and bistatic BackCom networks.

7.2.1.1 Monostatic BackCom networks

In monostatic BackCom network, the CW emitter and backscatter receiver are integrated (colocated) in one device. We called it a reader in the previous section. One typical example of the monostatic BackCom network is the RFID system that consists of one backscatter reader and one backscatter tag, as shown in Figure 7.2(a).

The main advantages of monostatic BackCom network include simple architecture, low power consumption, and low implementation cost. To be specific, the

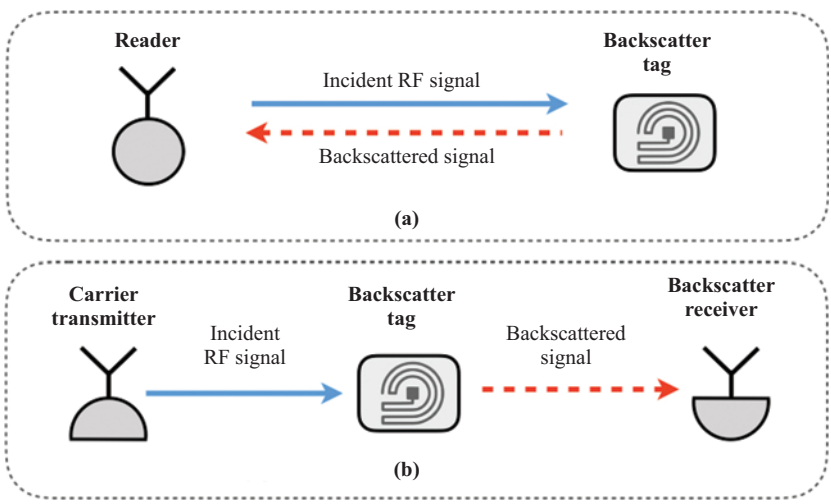


Figure 7.2 Point-to-point BackCom networks (a) monostatic and (b) bistatic

backscatter tag can communicate by leveraging the incident signals. It thus needs not the oscillator to generate active RF signals, leading to the low power consumption. For example, the backscatter tag used in [14] only consumes 10.6 W for operating its circuit. Moreover, as the backscatter tag can harvest energy from the incident signals for its operations, it does not require any battery. As a result, it can be designed in a small form size with low complexity. Therefore, the implementation cost of the monostatic BackCom network can be dramatically reduced. For example, a passive backscatter tag only costs 7–15 cents USD [15] and implementing a large-scale BackCom system consisting of 100 tags only costs 10 USD [14].

However, the main drawback of monostatic BackCom network is the short transmission distance between the reader and tag due to the round-trip pass-loss effect [16] and limited emitted power of the reader [17]. Therefore, the monostatic BackCom system is mainly used for short-range RFID applications. On the other hand, the principle of BackCom can be applied to more complex networks and integrated into different wireless communication technologies introduced in the sequel.

7.2.1.2 Bistatic BackCom networks

In the bistatic BackCom network, the CW transmitter and backscatter receiver are separated and deployed in different devices, as shown in Figure 7.2(b). In general, the carrier transmitter is the dedicated RF source such as Power Beacon (PB) [18].

Using the bistatic BackCom configuration can significantly increase the communication range between the backscatter tag and the backscatter receiver. The reason is that the path-loss of the forward link (from carrier transmitter to tag) decreases if the carrier transmitter is deployed close to the tag. Also, multiple carrier transmitters can be jointly used to power the tag to increase the power level of the backscattered signal and hence enlarge its transmission distance [19]. For example, in [16], if carrier transmitter emits power with 13 dBm and its distance to backscatter tag is 2–4 m, the BackCom range from tag to backscatter receiver could reach to 130 m. It can cover a wide range of Device-to-Device (D2D) communication area. Furthermore, the bistatic BackCom network is cost-effective since manufacturing the CW transmitter as well as the backscatter receiver is cheaper than those in the monostatic system [20].

7.2.2 Multi-access BackCom network

In practical IoT scenarios, a single reader may serve multiple backscatter tags for delivering multiple D2D BackCom links. It is called a multi-access BackCom network shown in Figure 7.3. Many IoT applications can be modeled as multi-access BackCom networks. In a smart home, for example, the central data processor could support a large number of IoT sensors and collect and aggregate the sensing data backscattered from IoT sensors simultaneously [21].

Multiple tag's concurrent transmissions can be collided, resulting in the transmission failure. For the collision avoidance, multi-access BackCom network should adopt suitable Multiple-Access Channel (MAC) schemes, including Time-/Frequency-/Code-/Space-Division Multiple Access (TDMA/FDMA/CDMA/SDMA) [3].

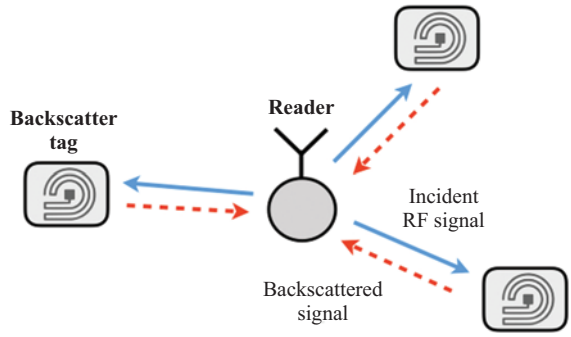


Figure 7.3 Multi-access BackCom networks

Specifically, TDMA is the most practical and straightforward MAC scheme for multi-access BackCom network that the different tags use different preassigned well-separated time slots for transmission [16]. The synchronization requirement between reader and tags could be easily satisfied due to the benefit of inherent closed-loop signaling of BackCom system. Then, in FDMA, the tag can change the frequency of the backscattered signals via RF switch [22], and thus, the reader could differentiate different signals in the frequency domain. However, applying FDMA increases the complexity significantly, power consumption, and cost of signal processing. Next, in CDMA, each tag uses a unique orthogonal or near-orthogonal code for modulating the backscattered signals for data separation at reader side. For example, in [23], the time-hopping spreading spectrum technique is used to generate the orthogonal codes for different tags to enable multi-access. It is worth mentioning that enabling power control at the tag side is vital in the CDMA system to avoid the near–far problem. Last, in SDMA, the reader is equipped with antenna arrays to form beams for scanning the around space. So, the tags within the scanning space of the reader can be distinguished via angular information [24]. The main drawback of SDMA is the high cost and high complexity due to the directional beamforming antennas.

7.2.3 Interference BackCom network

In the case with multiple tags and multiple readers, the tags reflect all incident signals, including useful RF signals as well as unwanted interference signals transmitted by other tags. The resultant interference effect could be more severe in this BackCom network and results in interference regeneration [23] as shown in Figure 7.4 (the case with two readers and two tags). Therefore, the total number of interference links received at the reader side is the square of the number of coexisting BackCom links, which is larger than that in the conventional sensor network. So effectively suppressing the interference is the one design challenge in distributed D2D or ad hoc BackCom network. In [25], a novel solution is proposed that treats the useful BackCom signals as sparse codes. It can be successfully decoded by using compressive sensing approach at the reader side.

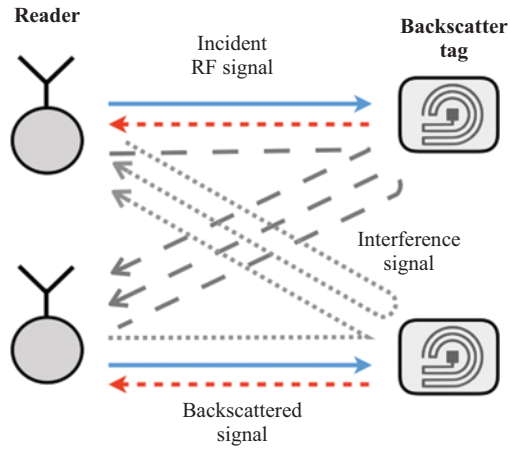


Figure 7.4 Interference BackCom networks

7.3 Emerging backscatter communication technologies

Despite the advantages mentioned earlier, the conventional BackCom system has several drawbacks. First, the transmit power of a reader is limited due to several reasons (i.e., regulation and hardware constraints). It is thus challenging to extend the coverage area of BackCom. Second, the receive antenna of a reader is typically collocated with its energy source, bringing about the self-interference between them. Third, due to its passive operation, it is challenging to apply advanced communication techniques such as multiple-access transmission and interference cancellation. These drawbacks make it challenging for BackCom to be adopted in many applications. It calls for developing ways to overcome the limitations. To this end, new types of BackCom system have been emerging for future IoT systems introduced as follows.

7.3.1 Ambient BackCom

There exist many kinds of RF sources in our environments, e.g., TV tower, FM towers, cellular base stations, and Wi-Fi Access Points (APs). We can utilize them for powering backscatter devices without the dedicated energy sources (i.e., reader). We call it an *ambient BackCom* [4]. Figure 7.5 illustrates BackCom, comprising the three main entities: an RF source, a backscatter transmitter, and a backscatter receiver. The backscatter transmitter can send information to the receiver by harvesting and reflecting the ambient signals broadcast from the RF source. By separating the RF source and the backscatter receiver, ambient BackCom provides several benefits. First, its energy consumption can be significantly reduced. For example, for the first design of ambient BackCom proposed in [26], its transmitter and receiver consume

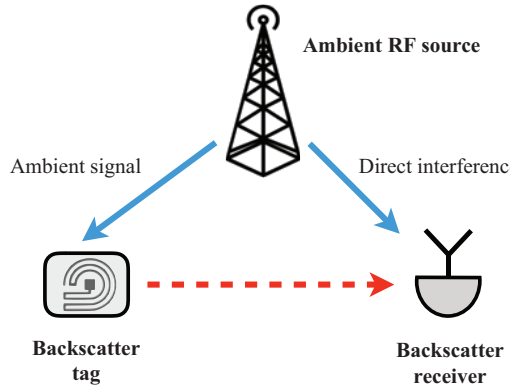


Figure 7.5 Ambient BackCom networks

0.25 and 0.54 μW, respectively. On the other hand, for Wireless Identification and Sensing Platform [27], which is one popular programmable BackCom system, the tag and reader consume 2.32 and 18 μW, respectively. Second, the manufacturing cost can also decrease since the RF source is an expensive component in conventional BackCom. Last, the ambient BackCom system does not actively transmit signals in the licensed spectrum, and the resultant interference to the legacy receiver is negligible. In other words, the ambient BackCom can operate as underlay cognitive radio without the dedicated bandwidth, helping the reduction of its operating cost.

For the ambient BackCom be practical, it is vital to develop a way to extract the desired information from the ambient backscatter signal. Contrary to the conventional BackCom using an unmodulated CW, the ambient RF source is a modulated signal. Consequently, the resultant backscatter signal is the mixture of the ambient and backscatter modulations, making the demodulation of the ambient BackCom more challenging. One simple scheme is to cancel the effect of the ambient modulation by averaging adjacent samples [28]. The backscatter signal's modulation rate is much slower than ambient ones. Therefore, the correlation between the adjacent samples of the desired backscatter modulation is almost one if it is within one symbol duration. On the other hand, the samples of the ambient signal are independent, and it is possible to cancel out the ambient signals by the averaging algorithm. However, this scheme is only valid when the concerned data rate is low. Besides, it suffers from direct interference from RF sources, bringing about the significant degradation of the received signal quality. Recently, several works have been proposed to improve the data rate by utilizing other RF sources such as FM broadcasting [28], Wi-Fi AP [29], and Orthogonal Frequency Division Multiplexing (OFDM) [30]. Especially in [28], the authors developed an algorithm to cancel out the direct-link interference based on the knowledge of the OFDM signal structure. It leads to increasing the data rate as well as improving the received signal quality.

7.3.2 Wirelessly powered BackCom

In the future massive IoT, the readers are energy limited and thus may not be used to power other tags for BackCom over sufficiently long ranges. Consequently, the conventional monostatic BackCom architecture is not suitable. Although the ambient BackCom system introduced in Section 7.3.1 could leverage the rich ambient RF signals to empower the BackCom, it has the following two major limitations [19]. First, it has a much lower data rate than that of ambient signal links. Second, it does not have the scalability since it depends on other networks as energy sources. The limitations mentioned earlier motivate the design of using the dedicated energy source (e.g., PB) to wirelessly power and enable the BackCom links between tags and receivers. It is called a wirelessly powered BackCom network. Figure 7.6 shows a typical example of a wirelessly powered BackCom network. Specifically, the low-complexity PB is deployed for wirelessly powering its nearby backscatter tags via either beamforming or isotropic transmission. The tags will first harvest energy from the incident signal and then modulate and backscatter the signal to their paired receivers. By using this network design, the data rate and communication distance are further increased.

Authors in [19] made the first attempt to model and optimize a large-scale wirelessly powered BackCom network by jointly using stochastic geometry and convex optimization. To be specific, the network topology is modeled as a Poisson cluster process where PBs are the cluster centers, and the tags are distributed around PBs as the cluster members. This modeling is motivated by the fact that only the tags that are close enough to the PBs could harvest sufficient energy for BackCom. Then, based on the proposed model, the network performance, including network coverage probability and capacity, is optimized in terms of BackCom parameters such as duty cycle and reflection coefficient. In [31], a hybrid wirelessly powered BackCom network is proposed to increase the transmission distance of BackCom further. The tags could be powered by either ambient signals (e.g., TV tower) or dedicated RF signals (e.g., carrier transmitter). In [32], the sum-throughput of a wirelessly powered BackCom

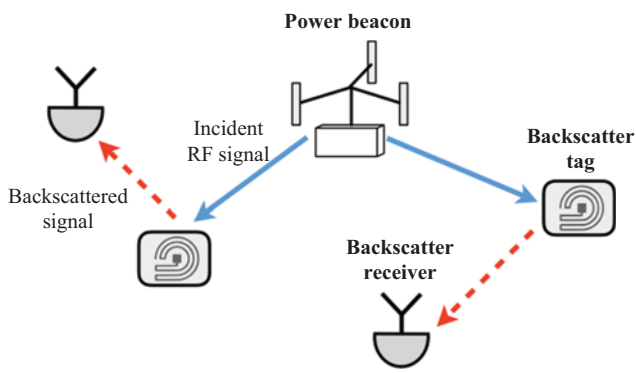


Figure 7.6 Wirelessly powered BackCom networks

network is maximized by optimizing the transmission policies such as time allocation and working mode permutation of BackCom users.

The advantages of wirelessly powered BackCom networks can provide flexibility to apply BackCom to a wide range of IoT scenarios. For example, authors in [33] proposed a novel backscatter sensing framework that uses BackCom for efficient sensing data transmission by sensors and applies statistical learning for accurate detection and inference at the receiver.

7.3.3 Full-duplex BackCom

Recall that the basic BackCom is designed for a unidirectional data transmission from a tag to its paired reader. On the other hand, it is expected in the future IoT services that bidirectional communications are essential to support frequent and low-latency data exchange between IoT devices. It is thus natural to consider Full-Duplex (FD) BackCom by allowing the IoT devices to speak and listen simultaneously. It is worth noting that in conventional communication systems, an essential technique for enabling FD is self-interference cancellation based on a knowledge of the signal the device is transmitting [34]. However, this approach is unsuitable for low-cost and low-complexity IoT devices, since sophisticated analog and digital signal processing units are required. On the other hand, BackCom's simultaneous information and energy transmissions make it much easier to implement FD BackCom without concerning the self-interference cancellation.

There exist some recent works designing FD BackCom. In [35], a simple FD BackCom is first proposed, where high-rate and low-rate transmissions in the opposite direction are superimposed. However, this method is based on the rate asymmetricity, and its usage is limited if both forward and backward links require high data rates. In [23], Time-Hopping Spread-Spectrum (THSS) is adopted in FD BackCom, where a THSS bit sequence is transmitted from a reader to a tag, which plays a role to suppress interference. Besides, it enables the reader to harvest more energy, which facilitates noncoherent energy detections.

7.3.4 Visible-light-BackCom

To deliver the reliable backscatter data links in RF-limited scenarios (e.g., hospitals or tunnels), Visible-Light-BackCom (VL-BackCom) system has been proposed as a complementary solution. It exploits the benefits of visible light such as its sufficient spectrum and good directionality [36]. In general, the VL-BackCom has a similar key principle to the conventional RF-enabled BackCom. The main difference is that the VL-BackCom leverages the visible light signals, instead of RF signals, for data transmission.

The primary setting of VL-BackCom system includes the VL-BackCom transmitter, VL-BackCom tag, and VL-BackCom receiver, shown in Figure 7.7. Specifically, the VL-BackCom tag first uses its solar panel to harvest energy from the visible light emitted by VL-BackCom transmitter and then modulates and reflects the visible light signal to the dedicated VL-BackCom receiver for demodulation. The modulation

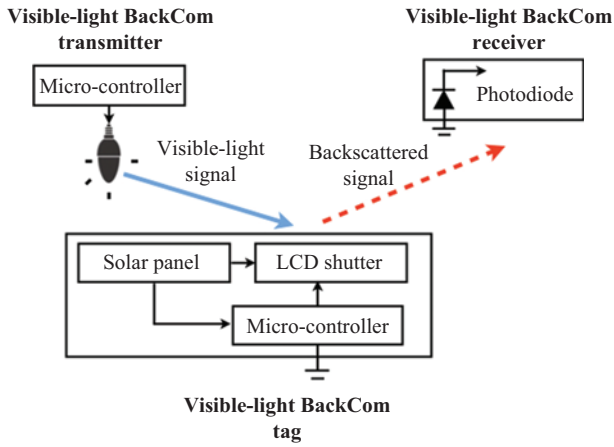


Figure 7.7 Basic setting of VL-BackCom system

procedure at tag is done by switching (i.e., pass or block) the Liquid Crystal Display shutter. The demodulation at receiver is achieved by using photodiode.

VL-BackCom has been a hot topic in green IoT and further enriches the functionality of BackCom. For example, in [37], an ambient VL-BackCom system is designed and implemented. It is shown in the experimental results that the system could deliver 0.5 and 10 kbps data rates for uplink and downlink, respectively, over a communication distance of 2.4 m. To further increase the data rate, several novel modulation schemes, named 8 Pulse Amplitude Modulation and trend-based modulation, are proposed in [36,38], respectively. Specifically, the data rate of uplink could be improved up to 1 kbps, which is four times higher than that in [37]. In addition, a large-scale VL-BackCom network that consists of multiple D2D VL-BackCom links is modeled and analyzed in [39] by using stochastic geometry. It is shown that the network performance could be optimized in terms of BackCom parameters such as duty cycle.

7.3.5 BackCom system with technology conversion

To enable the BackCom between various types of commercialized devices (e.g., Bluetooth and Wi-Fi devices), it is necessary to study and set up the BackCom system with different technology conversions.

Figure 7.8 shows one technology conversion that using the Bluetooth signal transmits data from a tag to a Wi-Fi device [40], following a bistatic BackCom architecture. Specifically, a smart watch, acting as the carrier transmitter, will emit the Bluetooth signal via Gaussian Frequency-Shift Keying (FSK) to the backscatter tag (e.g., a sensor). Then, the tag could shift the Bluetooth signal (carrier frequency is 2,426 MHz) to the Wi-Fi channel (carrier frequency is 2,462 MHz) by performing an FSK modulation via turning the mismatch level between antenna's impedances, and backscatter to the Wi-Fi device (e.g., a smartphone) for decoding. This technology conversion has been verified and implemented in [40].

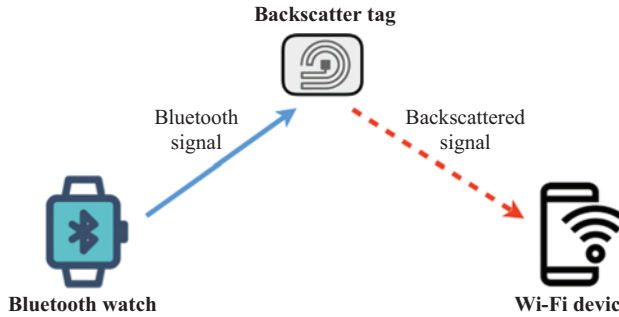


Figure 7.8 BackCom system with technology conversion

Technology conversion can also be implemented by using a multi-access BackCom architecture that includes a Wi-Fi AP and multiple backscatter tags [41]. Specifically, after receiving the Wi-Fi signals, tags could modulate their information onto the Wi-Fi signals and reflect them back to AP for decoding. Therefore, the tags may access the Internet, even in the presence of access by legal Wi-Fi clients such as laptops and smartphones. In summary, BackCom system with technologies conversion is envisioned to interconnect everything for smart IoT.

7.4 Performance enhancements of backscatter communication

The simple and low-cost designs of BackCom can help large-scale deployments for future IoT system, but they limit its performance such as communication range and error probability. Besides, the existing advanced signal processing techniques adopted in other communication systems may be unsuitable for BackCom due to its primitive architecture. As a result, it is essential to develop new techniques for BackCom introduced as follows.

7.4.1 Waveform design

A waveform refers to a specific shape of a signal expressed in terms of voltage. Recall that BackCom commonly considers a sinusoidal CW as its waveform. Such a CW is simple to generate, but more advanced waveforms can be used to boost the performance of the system.

To this end, we can leverage the significant progress made on the design of efficient communications and signal strategies for Wireless Power Transfer (WPT) [42]. In particular, multi-sine waveforms adaptive to the Channel State Information (CSI) have been shown particularly powerful in exploiting the rectifier nonlinearity and the frequency selectivity of the channel to maximize the amount of harvested DC power and extend the range of WPT [43,44]. Such a design can be leveraged to design efficient BackCom waveform. Interestingly, BackCom waveform design is slightly different from WPT waveform design, since the waveform design influences

not only the amount of energy delivered to the tag (as in WPT) but also the quality of the communication at the reader, measured e.g., in terms of Signal-to-Noise Ratio (SNR). This leads to a trade-off between the SNR at the reader and the amount of energy harvested at the tag, and the waveform design aims at identifying this trade-off and maximizing the amount of energy delivered at the tag subject to a minimum SNR requirement at the reader, or inversely [45].

7.4.1.1 Single-tag case

To study such trade-off more concretely, consider a multi-sine waveform (with N sine waves) transmitted at time t over a single antenna

$$x(t) = Re\left[\sum_{n=0}^{N-1} w_n e^{j2\pi f_n t}\right], \tag{7.3}$$

with $w_n = s_n e^{j\varphi_n}$ where s_n and φ_n refer to the amplitude and the phase of sine wave n at frequency f_n, respectively. The transmit waveform propagates through a multipath channel and is received at the tag as

$$y(t) = Re\left[\sum_{n=0}^{N-1} h_n w_n e^{j2\pi f_n t}\right], \tag{7.4}$$

where h_n is the forward channel frequency response at frequency f_n. This signal is absorbed by the tag (and not reflected to the reader) whenever the reflection coefficient is zero. Then it is conveyed to a rectifier that converts the incoming RF signal into DC. The amount of DC power harvested at the output of the rectifier P_{DC} is a nonlinear function $f_{NL}(y)$ of the input signal $y(t)$ [43]. On the other hand, when the reflection coefficient is 1, the signal is reflected to the reader (and not absorbed by the tag) such that the received signal at the reader can be written as

$$z(t) = m \cdot Re\left[\sum_{n=0}^{N-1} h_{r,n} h_n w_n e^{j2\pi f_n t}\right] + n(t), \tag{7.5}$$

where m equals 0 or 1 when the reflection coefficient at the tag is 0 and 1, respectively. $n(t)$ is an additive white Gaussian noise at the reader, and $h_{r,n}$ is the frequency response of the channel between the tag and the reader ($h_{r,n}$ could be different from h_n if the transmitter and reader are not colocated). After applying a product detector to each frequency and assuming an ideal low-pass filtering, the baseband signal on each frequency n is obtained and the N observations are combined using Maximum Ratio Combining. The SNR at the reader can then be computed.

Assuming that all channel coefficients have been estimated in advance, the waveform can then be optimized and the SNR-energy trade-off (so-called SNR-energy region) be characterized by formulating an optimization problem that aims at finding the set of complex coefficients w_n (hence, magnitude and phase) that maximizes the harvested energy $P_{DC} = f_{NL}(y)$ subject to an average transmit power constraint, and the SNR at the tag being larger than a minimum threshold [45]. Such optimization problem can be solved, and results highlight that as the SNR threshold is low, the transmit power is allocated over multiple frequencies (as a consequence of the

nonlinearity of the rectifier), while for large SNR threshold, the transmit power is dominantly allocated to a single frequency [45].

7.4.1.2 Multi-tag case

In a more general setting, it is likely that a transmitter has to serve multiple tags. In that case, all tags will compete to have access to the resources. This leads to a multiuser (or multi-tag) BackCom problem design. In [46], the problem of waveform design was studied for a multiuser BackCom. In contrast to the single-tag setup, in the presence of multiple tags, energy needs to be delivered to each of them and the waveform design for one given tag may not be suitable for another tag. Hence, there is trade-off between the amounts of energy delivered at the different tags. One valid metric is to consider a weighed sum of harvested DC power, where the higher the weight of one tag, the higher it is prioritized to receive energy. Moreover, since all the tags reflect data simultaneously, multiuser interference is created at the reader, and the communication quality can be measured in terms of Signal-to-Interference-plus-Noise Ratio (SINR) (instead of SNR as in single tag). The problem is therefore to identify the best transmit waveform that leads to the best trade-off between the weighted, harvested DC power and the SINRs (the so-called SINR-energy region). To that end, a multi-sine waveform at the transmitter is optimized dynamically as a function of the wireless channels so as to maximize a weighted sum of the harvested DC power, subject to SINR constraints at the tags [46]. The numerical results demonstrate the benefits of accounting for the harvester nonlinearity, multiuser diversity, frequency diversity, and multi-sine waveform adaptive to the channel state to enlarge the SINR-energy region. Results also highlight that such a simultaneous multiuser transmission using optimized waveform outperforms a TDMA approach.

7.4.2 Multi-antenna transmissions

Due to the two-way propagation, the BackCom requires a different channel model from other wireless systems. To be specific, it follows a cascaded channel of the forward and backward links, which can be modeled as the product of the two wireless channels. Consequently, the BackCom suffers from a double-propagation loss. According to [5], the link budget of BackCom is proportional to R^{-4}, where R represents the tag–reader distance. Noting that the link budget of other wireless systems with a one-way propagation is proportional to R^{-2}, its maximum distance is shorter than the conventional systems. One solution to extend the range is to deploy more antennas at both a reader and a tag, making it possible to use space-time coding for reliable data transmission. Besides, it helps one to increase the efficiency of the energy harvesting via energy beamforming and increasing receive antenna apertures [47].

7.4.2.1 Space-time coding

Let us consider a backscatter tag with L antennas and reader with M transmit and N receive antennas. Compared with a standard M by N Rayleigh Multiple-Input–Multiple-Output (MIMO) channel, the distinguishable feature of the backscatter channel is that all multipath signals from the reader are combined at the tag's antenna, called a pinhole effect. Then, the combined signal is returned back to the reader's

antenna as if it is generated from a single source. We define it as a Dyadic Backscatter Channel (DBC) and represent it as (M, L, N) DBC. In [48], the Bit Error Rate of uncoded BPSK over the DBC channel is analyzed, providing three interesting observations. First, SNR gain can be achieved when the tag's number of antennas L increases, even though the reader's number of receive antennas N is 1. Second, diversity gain can be achieved when both L and the reader's receive antennas N increase. In other words, increasing the number N cannot guarantee the reliability of the backward data transmission, which is different from the conventional MIMO channels. Third, increasing the number of the reader's transmit antennas M leads to coding gain. In [49], the pairwise error probability of BackCom is analyzed when Orthogonal Space-Time Block Codes [50] is used. It is shown that the diversity order is L if N is less than L. In [2], the Diversity-Multiplexing Trade-off (DMT) of BackCom is derived such that when the optimal DMT is given as $d^*(r) = L - r$ when $N \geq L$, and $d^*(r) = L(1 - r/N)$ when $N < L$, which is different from that of the conventional MIMO channel in [51].

7.4.3 Energy beamforming

Energy beamforming refers to one multi-antenna technique making the electromagnetic energy into a narrow beam for efficient energy transfer. Specifically, the reader transmits a narrow energy beam to the tag's direction (forward channel). It enables the tag to harvest more energy that are used to activate the tag's circuit more frequently and transmit data to the reader (backward channel) more reliably. For the energy beamforming to be more efficient, one prerequisite is to obtain CSI. Contrary to conventional wireless communications who use pilot symbol transmissions from the transmitter to the receiver to perform CSI estimation, it is difficult to estimate the forward and backward channels individually due to its cascade channel between the two. When the pilot symbol is transmitted from the reader, it only enables to obtain the product of the forward and backward CSIs, called a Backscatter-Channel CSI (BS-CSI). In [52], energy beamforming is optimized based on BS-CSI for a single-tag case. In [47], energy beamforming for a multi-tag case is considered where a new energy beamforming technique is developed by only using BS-CSI, and the optimal resource allocation scheme is derived for the proportional-fair-energy maximization.

7.5 Applications empowered by backscatter communications

Backscatter tags connected to the IoT based on BackCom will find a wide range of applications ranging from autonomous driving to logistics management. Several applications are introduced as follows based on the survey articles in [3,53].

7.5.1 BackCom-assisted positioning

Autonomous driving is envisioned as a disruptive technology for next-generation smart transportation. The technology is being developed at leading companies such as Google and Tesla and was recently demonstrated in real-life applications. Autonomous

vehicles are expected to be widely deployed in the near future, giving rise to an emerging market of tens of billions of dollars. An essential operation in autonomous driving technology is positioning, namely, recognizing the car's absolute and relative positions concerning other objects such as buildings, pedestrians, and other vehicles. The state of the art of vehicular positioning relies on Global Positioning System (GPS) satellites for positioning so as to follow planned routes. This requirement of line-of-sight to satellites limits the positioning to be in environments with few blockages such as rural areas or highways. Unfortunately, autonomous driving services are mostly needed in urban areas where GPS signals are frequently blocked by high-rise buildings. To overcome the limitation, Nvidia has developed a solution for Unmanned Aerial Vehicle (UAV) navigation without GPS by using visual recognition and deep learning. The technology is prone to accidents caused by visual errors under hostile weathers (i.e., fog, snow, and heavy rain) or in a poorly lighted environment. Furthermore, a required powerful computer for deep learning adds to the UAV weight and power consumption, which shortens the delivery ranges. An alternative technology, backscatter-tag-assisted autonomous driving (see Figure 7.9), is a promising solution that relies on the infrastructure's support to avoid the need of GPS satellites for positioning.

BackCom-based positioning has been extensively studied in the area of indoor positioning (e.g., [54–57]), while there exist a few recent works considering BackCom positioning in a vehicular environment. In [58,59], for example, tag-assisted vehicular positioning systems are proposed, where a large number of tags are deployed on the road surfaces embedding the corresponding location-related information. An on-board reader installed at bottom of a vehicle attempts to read tags that it passes by to know its location. Unfortunately, this approach may be infeasible in practice due to the following reasons. First, due to the limited coverage of BackCom, it requires high tag density and thereby high costs to achieve high positioning accuracy. Second, the tags on the road surfaces are unlikely to be durable because heavy vehicles frequently press down the tags. Besides, it is not easy to replace broken tags

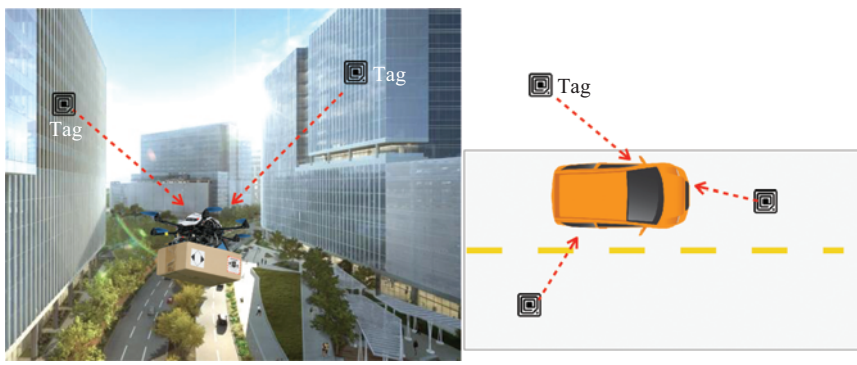


Figure 7.9 Backscatter-tag-assisted positioning for autonomous driving

because of safety issues on highway. Third, a vehicle with high velocity frequently fails to read the information of tags due to its short contact duration.

Tag-assisted vehicular positioning involves the embedding of a large number of low-cost backscatter tags into the ambient environment and a vehicle positioning itself by communicating with the tags. The principle of tag-assisted positioning is to (1) detect the relative location of a vehicle with respect to a tag; (2) read the absolute position stored in the tag; and (3) combine the relative and tag's absolute positions to give the absolute position of the vehicle. Developing this technology faces numerous challenges. For example, to reduce the required latency, positioning and BackCom should be integrated and calls for the development of new algorithms. As another example, Doppler shift at high mobility compromises positioning accuracy and needs to be coped with in the design.

7.5.2 Smart home and cities

Low-power or passive BackCom devices with energy-harvesting capabilities can be densely deployed to provide pervasive and uninterrupted sensing and computing services that provide a platform for implementing applications for smart homes/cities. In a smart home, a large number of passive BackCom sensors can be placed at flexible locations (e.g., embedded in walls, ceilings, and furniture). They are freed from the constraints due to recharging or battery replacements as one or multiple in-house PBs can be deployed to simultaneously power all the sensors or otherwise they can operate on ambient energy harvesting. The tasks performed by the sensors have a wide range such as detection of gas leak, smoke and Carbon Dioxide (CO_2), monitoring movements, indoor positioning, and surveillance (see Figure 7.10). As an example, BackCom-based smart dustbins are able to monitor their trash levels

Figure 7.10 BackCom-powered smart home

and communicate the information with passing-by garbage trucks by backscattering, streamlining the trash-collection process. Another example is that household robots are able to use the backscattered signals from the tags located on doors and furniture for indoor navigation. In a smart city, ubiquitous BackCom sensor nodes can be placed in every city corner such as buildings, bridges, trees, streetlamps, and parking areas. They can streamline the city operations and improve our life quality via, e.g., monitoring of air/noise pollution and traffic and parking-availability indicating. The efficient sensing data fusion and wireless power for BackCom sensors can be realized by the deployment of integrated PBs and APs at fixed locations or mounted on autonomous ground vehicles or UAVs, providing full-city coverage without costly backhaul networks.

7.5.3 Logistics

BackCom for logistics is a very attractive proposition due to the ultra-low manufacturing cost of simple and passive BackCom tags, as illustrated in Figure 7.11. For example, as early as 2007, the biggest 100 suppliers of the global renowned chain commercial group Wal-Mart have used the BackCom technology for logistics tracking. The technology has been helping the companies to substantially reduce operational cost, guarantee product quality, and accelerate the processing speed. In the past decade, the popularization and the application of BackCom have brought revolutionary changes to the logistics industry, due to its advantages compared with the conventional bar code technology such as reduced manual control, long service lives, long reading distances, and encryptable and rewritable data.

Figure 7.11 Backscatter-tag-assisted logistics management

Looking into the future, apart from the existing BackCom techniques for logistics tracking and management, BackCom-based three-dimensional orientation tracking is an emerging technique. By attaching an array of low-cost passive BackCom tags as orientation sensors on the surface of the target objects, three-dimensional orientation information is available at the reader by analyzing the relative phase offset between different tags. In this way, human workers can be warned when the angle of a cargo is larger than a threshold.

7.5.4 Biomedical applications

IoT biomedical applications such as plant/animal monitoring, wearable, and implantable human health monitoring require tiny and low heat-radiation communication devices. BackCom devices, which do not rely on any active RF component, can meet such requirements and thereby avoid causing any significant effect on the plants, animals, tissues, or organs being monitored. These advantages make BackCom a promising solution for IoT biomedical applications. One example is the BackCom-based smart Google Contact Lens, as illustrated in Figure 7.12. The lens was invented in Google in 2014 for the purpose of assisting people with diabetes by constantly measuring the glucose levels in their tears (once per second). The device consists of a miniaturized glucose sensor and a tiny BackCom tag. The tag is able to provide energy to the sensor by RF energy harvesting from a wireless controller and also backscatter the measured blood sugar level to the wireless controller for diagnosing purpose. Looking into the future, we envisage that BackCom will find a wide range of biomedical applications. In particular, implantable tiny BackCom neural devices with ultra-low power consumption and heat radiation may be placed on the surface of the patient's brain to help the study, diagnosis, and treatment of diseases such as epilepsy and Parkinson's disease, where the BackCom implants act as the brain–computer interface.

Figure 7.12 Backscatter-tag-enabled smart contact lens

7.6 Open issues and future directions

We discuss a few open research problems that have not fully studied in the literature.

7.6.1 From wireless information and power transmission to BackCom

BackCom is closely related to the general research area of Wireless Information and Power Transmission (WIPT), and Wirelessly Powered BackCom is actually one of the possible types of WIPT [60]. This implies that much work and ideas from the area of WPT and WIPT can be leveraged to analyze and enhance the performance of BackCom. Some examples were already discussed in Section 7.4. Progress not only in WPT regarding energy harvester modeling, energy beamforming for WPT, channel acquisition, power region characterization in multiuser WPT, waveform design with linear and nonlinear energy receiver model, safety and health issues of WPT, massive MIMO and millimeter wave (mmWave)-enabled WPT, wireless charging control, and wireless power and communication system codesign, as discussed in [42], but also in the RF design and the interplay between RF and signal design for WPT [61] can certainly play a role to enhance the WPT performance of Wirelessly Powered BackCom. Similarly, progress in the WIPT area, including energy harvester and receiver models and their impact on signal and system designs, rate-energy region characterization, transmitter and receiver architecture, waveform, modulation, beamforming and input distribution optimizations, resource allocation, and RF spectrum use, as discussed in [60], can also play a role to enhance the communication performance along with the energy delivery/harvesting performance of a BackCom system.

7.6.2 Security and jamming issues

The lightweight protocol and low-energy transmission of BackCom provide many advantages for future IoT devices explained so far. On the other hand, its simple architecture makes it impractical to apply advanced security and anti-jamming protocols, which results in exposing many malicious attacks. For example, when a jamming device intentionally sends RF signals during the BackCom procedure, the received signal quality can be significantly degraded since the BackCom signal's strength is too weak to suppress the jamming signal. Besides, an eavesdropper overhears the BackCom transmission without difficulty due to the lack of conventional security solutions such as encryption and digital signature. To overcome the limitations, there exist some works adopting physical-layer security approaches, which exploit wireless channel characteristics such as fading and noises, traditionally considered as obstacles, for defending BackCom against malicious attacks [62]. The authors in [63] proposed an artificial noise that is injected with the aid of the reader to protect the BackCom from unauthorized eavesdroppers. Specifically, a reader generates a random noise signal and adds it to the CW signal, which obscures the eavesdroppers to decode the BackCom information signal. The authors of [64] extend this approach into multi-antenna BackCom where the energy supply power and the precoding matrix of the artificial noise are jointly optimized. Using the additional degree-of-freedom of

multi-antenna channel, it helps the BackCom from jamming devices as well as eavesdroppers. The authors of [65] study the physical-layer security of multi-tag BackCom system under the consideration of channel correlation between forward and backward links. However, this artificial noise injection approach is only viable in the traditional BackCom where the reader plays a role of CW emitter and BackCom receiver, since the assumption of the noise injection knowledge is a prerequisite. In cases of other BackCom systems such as ambient BackCom, this assumption is invalid, making it challenging to apply the artificial noise injection approach directly. It opens a new research direction to develop a new physical-layer security approach for different types of BackCom systems.

7.6.3 mmWave-based BackCom

To enrich the limited spectrum resource under sub-6 GHz, mmWave communications explore a huge amount of available bandwidth, ranging from 24 to 100 GHz, to provide high data rate and high network capacity for 5G wireless networks [66]. Although the mmWave signal is previously assumed unsuitable for wireless transmission due to its high propagation loss, now it can be leveraged to offer high-speed-low-latency services based on the advanced signal processing technologies [67] (i.e., beamforming) and high gain antenna arrays [68].

Consider the massive IoT network of 5G. mmWave could also further enhance the performance of BackCom, including increasing the energy-efficiency, peak data rate, and enriching its functionality. Specifically, in [69], an mmWave-enabled monostatic BackCom system is designed and implemented to deliver 4 Gigabit backscatter transmission rates but just consume 0.15 pJ/bit energy. In [70], an indoor tag-assisted localization technique is proposed by transmitting mmWave-backscattered signals. The reader is equipped with high-directional beam-steering antennas to achieve reliable positioning performance. In addition, a novel backscatter-tag-assisted vehicular positioning approach is proposed in [71] by transmitting the mmWave signal to enable both BackCom and ranging purposes simultaneously.

Acknowledgment

This chapter is written on the basis of Prof. Kaibin Huang's work [3] published in EURASIP Journal on Wireless Communications and Networking 2019(69). He retains copyright, licensing the article under a Creative Commons license.

References

[1] Gubbim J, Buyya R, Marusic S, *et al.* Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer System. 2013;29(7):1645–1660.

[2] Boyer C and Roy S. Backscatter communication and RFID: Coding, energy, and MIMO analysis. IEEE Transactions on Communications. 2014;62(3): 770–785.
[3] Liu W, Huang K, Zhou X, *et al.* Next generation backscatter communications: systems, techniques, and applications. EURASIP Journal on Wireless Communications and Networking. 2019;2019(69):1–11.
[4] Huynh NV, Hoang DT, Lu X, *et al.* Ambient backscatter communications: A contemporary survey. IEEE Communications Surveys & Tutorials. 2018 Fourth quarter;20:2889–2922.
[5] Dardari D, D'Errico R, Robin C, *et al.* Ultrawide bandwidth RFID: The next generation? IEEE Proceeding. 2010;98(9):1570–1582.
[6] Griffin JD. High-frequency modulated-backscatter communications using multiple antennas. School Elect. Comput, Eng., Georgia Inst. Techn; 2009.
[7] Nikitin PV and Ra KVS. Antennas and propagation in UHF RFID systems. In: Proc. IEEE Int. Conf. RFID; 2008.
[8] Griffin JD and Durgin GD. Complete link budgets for backscatter radio and RFID systems. IEEE Antennas and Propagation Magazine. 2009;51(2): 11–25.
[9] International Organization for Standardization. Available from: https://www.iso.org.
[10] EPCglobal Standards. Available from: https://www.gs1.org/epcglobal.
[11] International Electrotechnical Commission. Available from: https://www.iec.ch.
[12] EPCglobal. EPC™ Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz – 960 MHz, Oct. 2008. Available from: https://www.gs1.org/sites/default/files/docs/epc/uhfc1g2_1_2_0-standard-20080511.pdf.
[13] ISO/IEC. Information Technology – Radio Frequency Identification for Item Management – Part 6: Parameters for Air Interface Communications at 860 MHz to 960 MHz. General, Jan. 2013, Available from https://www.iso.org/standard/59644.html.
[14] Konstantopoulos C, Koutroulis E, Mitianoudis N, *et al.* Converting a plant to a battery and wireless sensor with scatter radio and ultra-low cost. IEEE Transactions on Instrumentation and Measurement. 2016;65(2):388–398.
[15] How Much Does an RFID Tag Cost Today? RFID Journal. Available from: https://www.rfidjournal.com/faq/how-much-does-an-rfid-tag-cost-today.
[16] Kimionis J, Bletsas A, and Sahalos JN. Increased range bistatic scatter radio. IEEE Transactions on Communications. 2014;62(3):1091–1104.
[17] Choi SH and Kim DI. Backscatter radio communication for wireless powered communication networks. In: Proc. 21st Asia–Pac. Conf. Commun. (APCC); 2015.
[18] Huang K and Zhou X. Cutting the last wires for mobile communications by microwave power transfer. IEEE Communications Magazine. 2015;53(6): 86–93.

[19] Han K and Huang K. Wirelessly powered backscatter communication network: Modeling, coverage and capacity. IEEE Transactions on Wireless Communications. 2017;16(4):2548–2561.

[20] Lu X, Niyato D, Jiang H, *et al.* Ambient backscatter assisted wireless powered communications. IEEE Wireless Communications Journal. 2018;25(2): 170–177.

[21] Liu W, Liang YC, Li Y, *et al.* Backscatter multiplicative multiple-access systems: Fundamental limits and practical design. IEEE Transactions on Wireless Communications. 2018;17(9):2548–2561.

[22] Xu C, Yang L, and Zhang P. Practical backscatter communication systems for battery-free Internet of Things: A tutorial and survey of recent research. IEEE Signal Processing Magazine. 2018;35(5):16–27.

[23] Liu W, Huang K, Zhou X, *et al.* Full-duplex backscatter interference networks based on time-hopping spread spectrum. IEEE Transactions on Wireless Communications. 2017;16(7):4361–4377.

[24] Bletsas A, Siachalou S, and Sahalos J. Anti-collision backscatter sensor networks. IEEE Transactions on Wireless Communications. 2009;8(10):5018–5029.

[25] Wang J, Hassanieh H, Katabi D, *et al.* Efficient and reliable low-power backscatter networks. In: Proc. ACM SIGCOMM; 2012. pp. 61–72. Available from: http://doi.acm.org/10.1145/2342356.2342364.

[26] Liu V, Parks A, Talla V, *et al.* Ambient backscatter: Wireless communication out of thin air. In: Proc. of the ACM SIGCOMM; 2013. pp. 39–50. Available from: http://doi.acm.org/10.1145/2486001.2486015.

[27] Sample A, Yeager D, Powledge P, *et al.* Design of an RFID-based battery-free programmable sensing platform. IEEE Transactions on Instrumentation and Measurement. 2008;57(11):2608–2615.

[28] Daskakakis SN, Kimionis J, Collado A, *et al.* Ambient backscatterers using FM broadcasting for low cost and low power wireless applications. IEEE Transactions on Microwave Theory and Techniques. 2017;65(12):5251–5252.

[29] Bharadia D, Joshi KR, Kotaru M, and Katti S. BackFi: High throughput WiFi backscatter. In: Proc. of the 2015 ACM SIGCOMM. SIGCOMM'15. New York, NY, USA: ACM; 2015. pp. 283–296. Available from: http://doi.acm.org/10.1145/2785956.2787490.

[30] Yang G, Liang YC, Zhang R, *et al.* Modulation in the air: Backscatter communication over ambient OFDM carrier. IEEE Transactions on Communications. 2018;66(3):1219–1233.

[31] Kim S and Kim D. Hybrid backscatter communication for wireless-powered heterogeneous networks. IEEE Transactions on Wireless Communications. 2017;16(10):6557–6570.

[32] Lyu B, Yang Z, Gui G, *et al.* Wireless powered communication networks assisted by backscatter communication. IEEE Access. 2017;5:7254–7262.

[33] Zhu G, Ko SW, and Huang K. Inference from randomized transmissions by many backscatter. IEEE Transactions on Wireless Communications. 2018;17(5):3111–3127.

[34] Sabharwal A, Schniter P, Guo D, *et al.* In-band full-duplex wireless: Challenges and opportunities. IEEE Journal on Selected Areas in Communications. 2014;32(9):1637–1652.
[35] Liu V, Talla V, and Gollakota S. Enabling instantaneous feedback with full-duplex backscatter. In: Proc. of the ACM MobiCom. ACM; 2014. pp. 67–78.
[36] Xu X, Shen Y, Yang J, *et al.* PassiveVLC: Enabling practical visible light backscatter communication for battery-free IoT applications. In: Proc. 23rd Annu. Int. Conf. Mobile Computing and Networking; 2017.
[37] Li J, Liu A, Shen G, *et al.* Retro-VLC: Enabling battery-free duplex visible light communication for mobile and IoT applications. In: Proc. Int. Workshop Mobile Comput. Syst. Appl.; 2015.
[38] Shao S, Khreishah A, and Elgala H. Pixelated VLC-backscattering for self-charging indoor IoT devices. IEEE Photonics Technology Letters. 2017;29(2):177–180.
[39] Wang X, Han K, and Zhang M. Modeling the large-scale visible light backscatter communication network. In: Proc. of 23rd Asia-Pacific Conference on Communications; 2017.
[40] Iyer V, Talla V, Kellogg B, *et al.* Inter-technology backscatter: Towards internet connectivity for implanted devices. In: Proc. ACM SIGCOMM; 2016.
[41] Zhang P, Bharadia D, Joshi K, *et al.* Hitchhike: Practical backscatter using commodity WiFi. In: Proc. of ACM Conf. Embedded Netw. Sensor Sys.; 2016.
[42] Zeng Y, Clerckx B, and Zhang R. Communications and signals design for wireless power transmission. IEEE Transactions on Communications. 2017;65(5):2264–2290.
[43] Clerckx B and Bayguzina E. Waveform design for wireless power transfer. IEEE Transactions on Signal Processing. 2016;64(23):6313–6328.
[44] Huang Y and Clerckx B. Large-scale multi-antenna multi-sine wireless power transfer. IEEE Transactions on Signal Processing. 2017;65(21):5812–5827.
[45] Clerckx B, Zawawi ZB, and Huang K. Wirelessly powered backscatter communications: Waveform design and SNR-energy tradeoff. IEEE Communications Letters. 2017;21(10):2234–2237.
[46] Zawawi ZB, Huang Y, and Clerckx B. Multiuser wirelessly powered backscatter communications: Nonlinearity, waveform design and SINR-energy tradeoff. IEEE Transactions on Wireless Communications. 2019;18(1):241–253.
[47] Yang G, Ho CK, and Guan YL. Multi-antenna wireless energy transfer for backscatter communication systems. IEEE Journal on Selected Areas in Communications. 2015;33(2):2974–2987.
[48] Griffin J and Durgin G. Gains for RF tags using multiple antennas. IEEE Transactions on Antennas and Propagation. 2008;56(2):563–570.
[49] Boyer C and Roy S. Space time coding for backscatter RFID. IEEE Transactions on Wireless Communications. 2013;12(5):2272–2280.
[50] Tarokh V, Jafarkhani H, and Calderbank A. Space-time block codes from orthogonal design. IEEE Transactions on Information Theory. 1999;45(5): 1456–1467.

[51] Zheng L and Tse DNC. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. IEEE Transactions on Information Theory. 2003;49(5):1073–1096.
[52] Arnitz D and Reynolds M. Multi transmitter wireless power transfer optimization for backscatter RFID transponders. IEEE Antennas and Wireless Propagation Letters. 2013;12:849–852.
[53] Ko SW, Chae H, Han K, *et al.* V2X-based vehicular positioning: opportunities, challenges, and future direction. Arxiv preprint Arxiv:190804606. Available from: http://arxiv.org/abs/1908.04606.
[54] Ni LM, Liu Y, Lau YC, *et al.* LANDMARC: Indoor location sensing using active RFID. In: Proc. of IEEE PerCom; 2003. pp. 407–415.
[55] Saab SS and Nakad ZS. A standalone RFID indoor positioning system using passive tags. IEEE Transactions on Industrial Electronics. 2011;58(5): 1961–1970.
[56] Kotaru M, Zhang P, and Katti S. Localizing low-power backscatter tags using commodity Wifi. In: Proc. ACM CoNEXT; 2017. pp. 251–262.
[57] Liu T, Liu Y, Yang L, *et al.* BackPos: High accuracy backscatter positioning system. IEEE Transactions on Mobile Computing. 2016;15(3):586–598.
[58] Jing T, Wei X, Cheng W, *et al.* An efficient scheme for tag information update in RFID systems on roads. IEEE Transactions on Vehicular Technology. 2016;65(4):2435–2444.
[59] Qin H, Peng Y, and Zhang W. Vehicles on RFID: Error-cognitive vehicle localization in GPS-less environments. IEEE Transactions on Vehicular Technology. 2017;66(11):9943–9957.
[60] Clerckx B, Zhang R, Schober R, *et al.* Fundamentals of wireless information and power transfer: From RF energy harvester models to signal and system designs. IEEE Journal on Selected Areas in Communications. 2019;37(1): 4–33.
[61] Clerckx B, Costanzo A, Georgiadis A, *et al.* Toward 1G mobile power networks: RF, signal, and system designs to make smart objects autonomous. IEEE Microwave Magazine. 2018;19(6):69–92.
[62] Mukherjee A, Fakoorian SAA, Huang J, *et al.* Principles of physical layer security in multiuser wireless networks: A survey. IEEE Communications Surveys and Tutorials. 2014 Third quarter;16(3).
[63] Saad W, Zhou X, Han Z, *et al.* On the physical layer security of backscatter wireless system. IEEE Transactions on Wireless Communications. 2016;15(11):7547–7560.
[64] Yang Q, Wang HM, Zhang Y, *et al.* Physical layer security in MIMO backscatter wireless systems. IEEE Transactions on Wireless Communications. 2016;15(11):7547–7560.
[65] Zhang Y, Gao F, Fan L, *et al.* Secure communications for multi-tag backscatter system. IEEE Wireless Communications Letters. 2019;8(4):1146–1149.
[66] Rappaport T, Heath R, Daniels JR, *et al.* Millimeter wave wireless communications. Prentice Hall Press; 2014.

[67] Heath R, Gonzalez-Prelcic N, Rangan S, *et al.* An overview of signal processing techniques for millimeter wave MIMO systems. IEEE Journal of Selected Topics in Signal Processing. 2016;10(3):436–453.

[68] Rangan S, Rappaport T, and Erkip E. Millimeter-wave cellular wireless networks: Potentials and challenges. IEEE Proceeding. 2014;102(3):365–385.

[69] Kimionis J, Georgiadis A, and Tentzeris MM. Millimeter-wave backscatter: A quantum leap for gigabit communication, RF sensing, and wearables. In: Proc. IEEE MTT-S Int. Microw. Symp. vol. 812–815; 2017.

[70] Guidi F, Decarli N, Dardari D, *et al.* Millimeter-wave beamsteering for passive RFID tag localization. IEEE Journal of Radio Frequency Identification. 2018;2(1):9–14.

[71] Han K, Ko SW, Lee S, *et al.* Joint frequency-and-phase modulation for multi-antenna backscatter vehicular positioning. In: Proc. IEEE Int. Workshop Signal Process; 2019.

This page intentionally left blank

Chapter 8

Age minimization in energy harvesting communications

Ahmed Arafa[1], Songtao Feng[2], Jing Yang[2], Sennur Ulukus[3] and H. Vincent Poor[4]

Latency assessment in communication systems is commonly approached through measuring throughput (the amount of data that could be transmitted in a certain amount of time), or transmission delay (the amount of time it takes to transmit a certain amount of data). In this chapter, we introduce an alternative perspective on assessing latency in energy harvesting communication systems, namely, through the notion of the age-of-information (AoI) metric. Different from throughput and transmission delay, AoI measures the amount of time elapsed, since the latest amount of data has reached its destination. Therefore, it provides a mathematical measure of data freshness and timeliness at the destinations, and hence, is very suitable to assess latency for applications in which a fresh stream of data is continuously required over a period of time, such as in surveillance videos, remote sensing systems, and vehicular networks.

Minimizing AoI, however, leads to relatively different characteristics for optimal policies when compared to those maximizing throughput or minimizing transmission delay. This chapter discusses and characterizes AoI-optimal policies in the context of energy harvesting communications, in which transmitters do not have enough energy to transmit data all the time and maintain its freshness at the receivers.

The notion of AoI is introduced first, along with some related works. Then, the focus shifts to single transmitter–receiver pair systems. For these, the effects of having different battery sizes on the optimal policies are shown for perfect (zero-error) channels and erasure channels. This chapter is concluded by some takeaways and future directions for this active line of research.

[1]Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, United States
[2]School of Electrical Engineering and Computer Science, The Pennsylvania State University, State College, United States
[3]Department of Electrical and Computer Engineering, University of Maryland, College Park, United States
[4]Electrical Engineering Department, Princeton University, Princeton, United States

8.1 Introduction: the age-of-information (AoI)

Real-time sensing applications in which time-sensitive measurements of some physical phenomenon (status updates) are sent to a destination (receiver) call for careful transmission scheduling policies under proper metrics that assess the updates' timeliness and freshness. The AoI metric has recently attracted attention as a suitable candidate for such a purpose. AoI is defined as the time spent, since the latest measurement update has reached the destination, and hence it basically captures delay from the destination's perspective. Mathematically, at time t, the AoI $a(t)$, or merely age, is given by

$$a(t) = t - u(t), \tag{8.1}$$

where $u(t)$ denotes the time stamp of the last measurement received at the destination. *To keep the data fresh at the destination, one needs to minimize the AoI.*

The AoI metric has been studied in the literature under various settings: mainly through modeling the update system as a queuing system and analyzing the long-term average AoI, and through using optimization tools to characterize optimal status updating policies, see, e.g., [1–25].

8.1.1 Status updating under energy harvesting constraints

When sensors (transmitters) rely on energy harvested from nature to transmit their status updates, they cannot transmit continuously, so that they do not run out of energy and risk having overly stale status updates at the destination. Therefore, the fundamental question of how to optimally manage the harvested energy to send timely status updates needs to be addressed.

8.1.1.1 Summary of related works

There have been a number of works studying this setting under various assumptions [26–55]. With the exception of [31], an underlying assumption in these works is that energy expenditure is normalized, i.e., sending one status update consumes one energy unit. References [26,27] consider a sensor with infinite battery, with [26] focusing on online policies under stochastic service times, and [27] focusing on both off-line and online policies with zero service times, i.e., with updates reaching the destination instantly.* Reference [28] studies the effect of sensing costs on AoI with an infinite battery sensor transmitting through a noisy channel. Using a harvest-then-use protocol, [28] presents a steady-state analysis of AoI under both deterministic and stochastic energy arrivals. The off-line policy in [27] is extended to nonzero, but fixed, service times in [29] for both single and multi-hop settings, and in [31] for energy-controlled variable service times.

The online policy in [27] is analyzed through a dynamic programming approach in a discretized time setting and is shown to have a threshold structure, i.e., an update is sent only if the age grows above a certain threshold and energy is available for transmission. Motivated by such results for the infinite battery case, [32] then studies

*In fact, most works focus on the zero service time case (cf. Table 8.1).

the performance of online threshold policies for the finite battery case under zero service times. Reference [33] proves the optimality of online threshold policies under zero service times for the special case of a unit-sized battery, via tools from renewal theory. It also shows the optimality of best effort online policies, where updates are sent over uniformly spaced time intervals if energy is available, for the infinite battery case. Reference [30] shows that such best effort policy is optimal in the online case of multi-hop networks, thereby extending the off-line work in [29]. Best effort is also shown to be optimal, for the infinite battery case, when updates are subject to erasures, with and without erasure feedback, in [34–36].

Under the same system model of [34], study of [37] analyzes the performances of the best effort and the save-and-transmit online policies, where the sensor saves some energy in its battery before attempting transmission, for the purpose of coding to combat channel erasures. A slightly different system model is considered in [38], in which status updates' arrival times are exogenous, i.e., their measurement times are not controlled by the sensor. With a finite battery, and stochastic service times, [38] employs tools from stochastic hybrid systems to analyze the long-term average AoI. The work in [39] considers a similar queuing framework as in [38] and studies the value of preemption in service on AoI. Reference [40] also considers a similar approach as in [38,39] under general energy and data buffer sizes. An interesting approach is followed in [41] where the idea of sending extra information, on top of the measurement status updates, is introduced and analyzed for unit batteries and zero service times.

Optimality of threshold policies for finite batteries with online energy arrivals has been shown in [42–44] using tools from renewal theory and a Lagrangian framework, which provides closed-form solutions of the optimal thresholds. This has also been shown independently and concurrently in [45] using tools from the optimal stopping theory. Reference [46] shows the optimality of threshold policies under general age-penalty functionals. Online policies for unit batteries with update erasures have been shown to also have a threshold structure in [47,48].

Other frameworks where AoI with energy harvesting has been studied include works that focus on wireless power transfer [49], multiple access channels [50], cognitive radio systems [51], monitoring with priority [52], operational and sensing costs [53,54], and trade-offs between AoI and distortion [55].

8.1.1.2 Categorization

We categorize the previous related works according to three main aspects:

- The first is the battery size B, which can either be *finite* or *infinitely* large.
- The second is the energy arrival process knowledge, which can either be *offline*, i.e., predictable before the energy arrives (is harvested), or *online*, i.e., can only be known causally after the energy arrives.
- The third is the service time d, denoting the time for an update to traverse through the communication channel and reach the destination; this can take multiple forms but is mainly categorized into *deterministic* (zero or nonzero) services times and *stochastic* service times.

Using the previous three categories (B, offline/online, d), we summarize the related works [26–55] in Table 8.1.

8.1.2 Chapter outline and focus

Over the next two sections, we will discuss two main categories of status updating under energy harvesting constraints to further illustrate how AoI affects the optimal transmission policies. The first is related to updating over perfect (zero-error) communication channels [33,42–44] in Section 8.2, and the second is related to updating over erasure channels [34–36,47,48] in Section 8.3. We will be discussing works with various battery sizes ($B = \infty$, $B = 1$ and $B < \infty$), with online knowledge of the energy arrivals, along with zero service times ($d = 0$). One main takeaway from the findings of these works is emphasized next.

Greedy is not always optimal

To minimize the long-term average AoI, depending on the energy arrival profile, it is not always optimal to send a new status update *whenever energy is available.* Rather, it is optimal to evenly spread out the status updates over time, up to the extent allowed by the energy availability.

Table 8.1 Summary of works on AoI with energy harvesting

Battery size B	**Energy arrival knowledge**	**Service time d**	**Specifics/keywords**	**Ref.**
∞	Online	Stochastic	Lazy scheduling	[26]
∞	Offline and online	0	Equispaced updating	[27]
∞	Offline and online	Fixed	Sensing costs	[28]
∞	Offline and online	Fixed	Single-hop and multi-hop	[29,30]
∞	Offline	Variable	Energy controls service time	[31]
$<\infty$	Online	0	Update erasures	[32]
$\infty, <\infty, 1$	Online	0	Best effort and threshold policies	[33]
∞	Online	0	Update erasures	[34–36]
∞	Online	0	Coding for update erasures	[37]
$<\infty$	Online	Stochastic	Queuing framework	[38–40]
1	Online	0	Simult. updates and information	[41]
$<\infty$	Online	0	Threshold policies	[42–45]
$<\infty$	Online	0	General AoI functionals	[46]
1	Online	0	Update erasures	[47,48]
$<\infty$	Online	1	Wireless power transfer	[49]
∞	Online	1	Multiple access	[50]
$<\infty$	Online	1	Cognitive radio setting	[51]
$<\infty$	Online	1	Priority monitoring	[52]
$\infty, 1$	Online	0	Operational costs	[53]
1	Online	Stochastic	Sensing costs	[54]
∞	Offline and online	1	Distortion effects	[55]

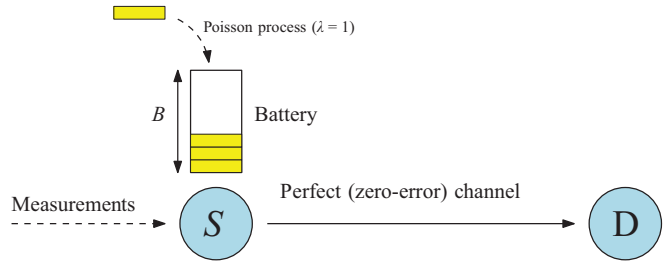


Figure 8.1 System model for status updating over perfect (zero-error) channels

The abovementioned takeaway is one fundamental observation that makes AoI minimization different from (conventional) transmission delay minimization or throughput maximization approaches.

8.2 Status updating over perfect channels

In this section, we discuss the results reported in [33,44] in greater detail. In these works, energy arrives in units according to a Poisson process with normalized rate $\lambda = 1$ arrival per unit time (see Figure 8.1). In addition, energy expenditure is also normalized in the sense that one update transmission consumed one energy unit.[†] Let s_i denote the time at which the sensor acquires (and transmits) the ith measurement update, and let $\mathscr{E}(t)$ denote the amount of energy in the battery at time t. We then have the following energy causality constraint:

$$\mathscr{E}\left(s_i^-\right) \geq 1, \quad \forall i. \tag{8.2}$$

We assume that we begin with an empty battery at time 0. The battery evolves as follows over time:

$$\mathscr{E}\left(s_i^-\right) = \min\left\{\mathscr{E}\left(s_{i-1}^-\right) - 1 + \mathscr{A}\left(x_i\right), B\right\}, \tag{8.3}$$

where $x_i \triangleq s_i - s_{i-1}$, and $\mathscr{A}(x_i)$ denotes the number of energy arrivals in $[s_{i-1}, s_i)$. Note that $\mathscr{A}(x_i)$ is a Poisson random variable with parameter x_i. We denote by $\mathscr{F}$, the set of feasible transmission times $\{s_i\}$ described by (8.2) and (8.3) in addition to an empty battery at time 0, i.e., $\mathscr{E}(0) = 0$.

Let $n(t)$ denote the total number of updates sent by time t. We are interested in minimizing the average AoI represented by the area under the age evolution curve versus time, see Figure 8.2 for a possible sample path with $n(t) = 3$. At time t, this area is given by

$$r(t) \triangleq \frac{1}{2}\sum_{i=1}^{n(t)} x_i^2 + \frac{1}{2}\left(t - s_{n(t)}\right)^2. \tag{8.4}$$

[†]Normalized arrival rates and transmission times are an assumption that will carry along the whole chapter. Extensions to non-normalized arrival rates and transmission times can be directly derived, at the expense of increased AoI as the arrival rate decreases or the transmission time increases.

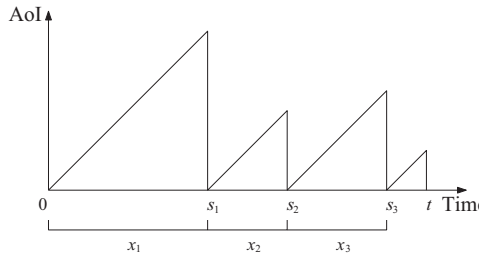


Figure 8.2 Example of the age evolution versus time with $n(t) = 3$

The goal is to choose a set of feasible transmission times $\{s_1, s_2, s_3, \dots\} \in \mathscr{F}$ such that the long-term average AoI is minimized. Equivalently, one can optimize the inter-update times $\{x_1, x_2, x_3, \dots\}$ to do so. Therefore, the goal is to characterize the optimal long-term average AoI, as a function of the battery size, $\rho(B)$ by solving:

$$\rho(B) \triangleq \min_{\{x_i\} \in \mathscr{F}} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[r(T)\right]. \tag{8.5}$$

We will first discuss the solution for $B = \infty$, followed by the special case of $B = 1$ energy unit, and then conclude the section by discussing the general case of $B < \infty$.

8.2.1 The case $B = \infty$

When the battery size is infinite, no energy overflow will happen. If we then replace the energy causality constraint in (8.2) by its long-term average (which is in this case equal to 1), one can show that [33]

$$\rho(\infty) \geq \frac{1}{2}. \tag{8.6}$$

The intuition is that one would send a new update per unit time without violating the constraints *on average*. In reality, however, sending one update per unit time is not always feasible, since (8.2) represents an instantaneous constraint. Let us now define the following policy.

Definition 8.1 (Best effort uniform updating [33]). *The sensor is scheduled to send a new update at $s_n = n$, $n = 1, 2, 3, \dots$. The sensor performs the task as scheduled if $\mathscr{E}(s_n^-) \geq 1$. Otherwise, it stays silent until the next scheduled sampling time.*

Clearly, the best effort uniform (BU) updating policy is always feasible. One of the main results of [33] is showing that it is also optimal for $B = \infty$. That is, proving the following theorem:

Theorem 8.1 ([33]). *The BU updating policy is optimal for $B = \infty$.*

Proof sketch. The main idea is to show that under the BU updating policy, $\rho(\infty) \leq \frac{1}{2}$, and hence it must be optimal since $\frac{1}{2}$ is already a lower bound. Toward that end, let us

define $K(T)$ as the number of *missed* transmission opportunities under the policy up to time T, i.e., the number of times in which the sensor could not transmit a new update as scheduled due to the lack of energy. Leveraging tools from laws of large numbers, one can show that $\lim_{T\to\infty} K(T)/T = 0$ almost surely (details in [33]). Thereby, the policy will *eventually* be transmitting a new update at the highest possible value of 1 update per unit time and achieving the lower bound. □

Through simulation results, it has also been demonstrated in [33] that the BU updating policy works quite well for other energy arrival models different from Poisson, such as Markovian energy sources. This shows that the policy can achieve desirable performances under general arrival models as well.

8.2.2 The case $B = 1$

This subsection and the next deal with finite battery cases. In order to minimize the long-term average AoI when B is finite, the status update policy should try to prevent battery overflows, since wasted energy leads to performance degradation. On the other hand, owing to the nature of AoI, one should also try to send updates as uniformly as possible (as seen in the $B = \infty$ case). As we will see, the optimal policy would then strike a balance in between these two premises.

One technical condition is needed for the treatment of this finite battery case, namely, that *optimal status update policies have inter-update times of bounded second moment.* Policies that abide by this technical condition are denoted as *uniformly bounded policies* in [33]. For such policies, we have the following structural result.

Theorem 8.2 ([33]). *The optimal status updating policy has a renewal structure, in which the update times $\{s_i\}$ constitute a renewal process.*

The proof of the theorem mainly relies on the memoryless property of the exponential distribution governing energy inter-arrival times. The reader is referred to [33] for more details. Focusing on renewal policies, one can vastly reduce the complexity of the problem. It basically states that history does not really matter in between updates.

Further, by [33, Theorem 4], it is shown that the optimal renewal policy admits a *threshold* structure, in which a new update is only transmitted if the AoI grows above a certain threshold. The following is a proof of such statement, which is also stated in [44, Section III].

Let τ_i denote the time until the next energy arrival since the $(i-1)$th update time, s_{i-1}. Since the arrival process is Poisson with rate 1, τ_is are independent and identically distributed (i.i.d.) exponential random variables with parameter 1. Under renewal policies, the ith inter-update time x_i should not depend on the events before s_{i-1}; it can only be a function of τ_i. Moreover, under any feasible policy, $x_i(\tau_i)$ cannot be smaller than τ_i, since the battery is empty at s_{i-1}. Next, note that whenever an update occurs, both the battery and the AoI drop to 0, and hence the system resets. This constitutes a renewal event, and therefore using the strong law of large numbers

of renewal processes (renewal reward theorem) [56, Theorem 3.6.1], $\rho(1)$ becomes equal to

$$\rho(1) = \min_{x(\tau)\geq\tau} \quad \frac{\mathbb{E}\left[x(\tau)^2\right]}{2\mathbb{E}[x(\tau)]}, \tag{8.7}$$

where expectation is over the exponential random variable τ.

In order to make problem (8.7) more tractable to solve, we introduce the following parameterized problem:

$$p_1(\lambda) \triangleq \min_{x(\tau)\geq\tau} \quad \frac{1}{2}\mathbb{E}\left[x(\tau)^2\right] - \lambda\mathbb{E}[x(\tau)]. \tag{8.8}$$

One can show that $p_1(\lambda)$ is decreasing in λ, and the optimal solution of problem (8.7) is given by λ^* that solves $p_1(\lambda^*) = 0$ [57], which can be found by, e.g., a bisection search. Therefore, $\lambda^* = \rho(1)$. Focusing on problem (8.8), we introduce the following Lagrangian [58]:

$$\mathscr{L} = \frac{1}{2}\int_0^\infty x^2(\tau)e^{-\tau}d\tau - \lambda\int_0^\infty x(\tau)e^{-\tau}d\tau - \int_0^\infty \mu(\tau)(x(\tau)-\tau)d\tau, \tag{8.9}$$

where $\mu(\tau)$ is a nonnegative Lagrange multiplier. Taking (the functional) derivative with respect to $x(t)$ and equating to 0, we get:

$$x(t) = \lambda + \frac{\mu(t)}{e^{-t}}. \tag{8.10}$$

Now if $t < \lambda$, $x(t)$ has to be larger than t, for if it were equal, the right-hand side of the above equation would be larger than the left-hand side. By complementary slackness [58], we conclude that in this case $\mu(t) = 0$, and hence $x(t) = \lambda$. On the other hand, if $t \geq \lambda$, $x(t)$ has to be equal to t, for if it were larger, by complementary slackness, $\mu(t) = 0$ and the right-hand side of the previous equation would be smaller than the left-hand side. In conclusion, we have:

$$x(t) = \begin{cases} \lambda, & t \leq \lambda \\ t, & t > \lambda \end{cases}. \tag{8.11}$$

This means that the optimal inter-update time is threshold-based; if an energy arrival occurs before λ amount of time since the last update time, i.e., if $\tau < \lambda$, the sensor should not use this energy amount right away to send an update. Instead, it should wait for $\lambda - \tau$ extra amount of time before updating. Else, if an energy arrival occurs after λ amount of time since the last update time, i.e., if $\tau \geq \lambda$, the sensor should use that amount of energy to send an update right away. We denote this kind of policy as λ-*threshold policy*. Substituting this $x(t)$ into (8.8), we get:

$$p_1(\lambda) = e^{-\lambda} - \frac{1}{2}\lambda^2, \tag{8.12}$$

which admits a unique solution of $\lambda^* \approx 0.9012$ when equated to 0.

We can now see that the balance in updating is attained. Basically, a threshold policy *tries* to evenly spread out updates over time by delaying sending new updates if an energy arrives relatively early, i.e., before exactly 0.9012 for $B = 1$. In the next

subsection, we extend this policy to work for general $1 < B < \infty$ and show that it has a multi threshold structure, one for each possible energy level in the battery.

8.2.3 The case $B < \infty$

In this section, we focus on problem (8.5) for the general case of $1 < B < \infty$, whose solution is reported in [44].[‡] Similar to the $B = 1$ case, [44] first shows that the optimal update policy that solves problem (8.5) has a renewal structure, namely, that it is optimal to transmit updates in such a way that the inter-update delays are independent over time, and that the time durations in between the two consecutive events of transmitting an update and having $k \leq B - 1$ units of energy left in the battery are i.i.d., i.e., these events occur at times that constitute a renewal process. We first introduce some notation.

Let the pair $(\mathscr{E}(t), a(t))$ represent the state of the system at time t. Fix $k \in \{0, 1, \ldots, B - 1\}$, and consider the state $(k, 0)$, which means that the sensor has just submitted an update and has k units of energy remaining in its battery. Let l_i denote the time at which the system visits $(k, 0)$ for the ith time. We use the term *epoch* to denote the time in between two consecutive visits to $(k, 0)$. Observe that there can possibly be an infinite number of updates occurring in an epoch, depending on the energy arrival pattern and the update time decisions. For instance, in the ith epoch, which starts at l_{i-1}, one energy unit may arrive at some time $l_{i-1} + \tau_{1,i}$, at which the system goes to state $(k + 1, \tau_{1,i})$, and then the sensor updates afterward to get the system state back to $(k, 0)$ again. Another possibility (if $k \geq 1$) is that the sensor first updates at some time $l_{i-1} + x_{k,i}$, at which the system goes to state $(k - 1, 0)$, and then two consecutive energy units arrive at times $l_{i-1} + \tau_{1,i}$ and $l_{i-1} + \tau_{1,i} + \tau_{2,i}$, respectively, at which the system goes to state $(k + 1, \tau_{1,i} + \tau_{2,i})$, and then the sensor updates afterward to get the system state back to $(k, 0)$ again. Depending on how many energy arrivals occur in the ith epoch, how far apart from each other they are, and the status update times, one can determine the length of the ith epoch and how many updates it has. The next theorem provides a structural result of the optimal update policy, extending Theorem 8.2 to work for $1 < B < \infty$.

Theorem 8.3 ([44]). *The optimal status update policy for problem (8.5) in the case $1 < B < \infty$ is a renewal policy, i.e., the sequence $\{l_i\}$ denoting the times at which the system visits state $(k, 0)$, for some fixed $0 \leq k \leq B - 1$, forms a renewal process.*

The proof follows the memoryless property of the exponential distribution governing energy inter-arrival times. The reader is referred to [44] for more details. Based on Theorem 8.3, the next corollary now follows.

Corollary 8.1 ([44]). *In the optimal solution of problem (8.5), the inter-update times are independent.*

[‡]We note that [44] also investigates another energy arrival model in which one energy arrival completely fills up the battery. We do not discuss such model here for coherence of treatment and refer the reader to [44, Section IV] (and originally [42]) for more details.

Proof. Observe that whenever an update occurs, the system enters state $(j, 0)$ for some $j \leq B - 1$. The system then starts a new epoch *with respect to state* $(j, 0)$. Since the choice of k energy units in Theorem 8.3 is arbitrary, the results of the theorem now tell us that the update policy in that epoch, and therefore its length, is independent of the past history, in particular the past inter-update lengths. □

Based on Corollary 8.1, we have the following observation. Let us assume that the optimal policy is such that the state at time t is (j, τ). This means that the previous status update occurred at time $t - \tau$. By Corollary 8.1, the policy at time t is independent of the events before time $t - \tau$. However, it may depend on the events occurring in $[t - \tau, t)$. For instance, for $j \geq 1$, it may be the case that at time $(t - \tau)^+$ the sensor had $j - 1$ energy units in its battery and then received another energy unit at some time in $[t - \tau, t)$; or it may have already started with j energy units at time $(t - \tau)^+$ and received no extra energy units in $[t - \tau, t)$. The question now is whether the optimal policy at time t is the same in either of the two scenarios. The following result concludes that it is indeed the same. The proof also depends on the memoryless property of exponentials, whose details are in [44].

Lemma 8.1 ([44]). *The optimal status update policy of problem (8.5) is such that at time t the next scheduled update time is only a function of the system state* $(\mathscr{E}(t), a(t))$.

By Theorem 8.3, focusing on state $(k, 0)$ for some $k \leq B - 1$ and defining the epochs with respect to this state, problem (8.5) reduces to an optimization over a single epoch. Based on Corollary 8.1 (and Lemma 8.1), we introduce the following notation.

Once the system goes into state $(k, 0)$, for $1 \leq k \leq B - 1$, at some time l, the sensor schedules its next update after x_k time. Since x_k does not depend on the history before time l and cannot depend on the future energy arrivals by the energy causality constraint, we conclude that it is a constant. Now if the first energy arrival in that epoch occurs at time $l + \tau_1$ with $\tau_1 > x_k$, the sensor transmits the update at $l + x_k$, whence the state becomes $(k - 1, 0)$, and if $k \geq 2$ the sensor schedules its next update after x_{k-1} time, i.e., at $l + x_k + x_{k-1}$. On the other hand, if the first energy unit arrives relatively early, i.e., $\tau_1 \leq x_k$, the state becomes $(k + 1, \tau_1)$ at $l + \tau_1$, and the sensor *reschedules* the update to be at $l + y_{k+1}(\tau_1)$ instead of $l + x_k$. Note that y_{k+1} only depends on τ_1, since it does not depend on the history before time l. If the second energy arrival in that epoch occurs at time $l + \tau_1 + \tau_2$ with $\tau_2 > y_{k+1}(\tau_1)$, the sensor transmits the update at $l + y_{k+1}(\tau_1)$, whence the state returns to $(k, 0)$. On the other hand, if the second energy arrival occurs relatively early as well, i.e., $\tau_2 \leq y_{k+1}(\tau_1)$, and if $k \leq B - 2$, the state becomes $(k + 2, \tau_1 + \tau_2)$ at $l + \tau_1 + \tau_2$, and the sensor reschedules the update at $l + y_{k+2}(\tau_1 + \tau_2)$ instead of $l + y_{k+1}(\tau_1)$.

In summary, the optimal update policy is completely characterized by $B - 1$ constants: $\{x_1, x_2, \ldots, x_{B-1}\}$, and B functions: $\{y_1(\cdot), y_2(\cdot), \ldots, y_B(\cdot)\}$, where x_k represents the scheduled update time after entering state $(k, 0)$, and $y_k(t)$ represents the scheduled update time after entering state (k, t) at some time t. We emphasize the fact that by Corollary 8.1, the constants $\{x_k\}$ neither depend on each other, nor on the functions $\{y_k(\cdot)\}$.

8.2.3.1 Renewal state analysis

To analyze the optimal solution of our problem, in view of Theorem 8.3, we now need to choose some renewal state $(k,0)$, $k \leq B-1$, and define the epoch with respect to that state. We note that for all choices of $k \leq B-1$, there can possibly be an infinite number of updates in a single epoch. In the sequel, we continue our analysis with state $(0,0)$ as the renewal state and define the epochs with respect to it, i.e., an epoch from now onward denotes the time between two consecutive visits to state $(0,0)$. We note, however, that any other renewal state choice yields the same results with equivalent complexity. We use the notation $R(\boldsymbol{x},\boldsymbol{y})$ and $L(\boldsymbol{x},\boldsymbol{y})$ to denote the area under the age curve in a given epoch and its length, respectively, as a function of the constants $\boldsymbol{x} \triangleq [x_1, x_2, \ldots, x_{B-1}]$ and the functions $\boldsymbol{y} \triangleq [y_1, y_2, \ldots, y_B]$. Using the strong law of large numbers of renewal processes [56], problem (8.5) now reduces to:

$$\begin{aligned} \rho(B) = \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \frac{\mathbb{E}\left[R(\boldsymbol{x},\boldsymbol{y})\right]}{\mathbb{E}\left[L(\boldsymbol{x},\boldsymbol{y})\right]} \\ \text{s.t.} \quad & x_k \geq 0, \quad 1 \leq k \leq B-1 \\ & y_k(\tau) \geq \tau, \quad 1 \leq k \leq B. \end{aligned} \tag{8.13}$$

We now introduce the auxiliary parameterized problem:

$$\begin{aligned} p_B(\lambda) \triangleq \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \mathbb{E}\left[R(\boldsymbol{x},\boldsymbol{y})\right] - \lambda \mathbb{E}\left[L(\boldsymbol{x},\boldsymbol{y})\right] \\ \text{s.t.} \quad & \text{constraints of (8.13)} \end{aligned} \tag{8.14}$$

and solve for the unique λ^* such that $p_B(\lambda^*) = 0$. Observe that $\lambda^* = \rho(B)$. One main goal now is to express $\mathbb{E}\left[R(\boldsymbol{x},\boldsymbol{y})\right]$ and $\mathbb{E}\left[L(\boldsymbol{x},\boldsymbol{y})\right]$ explicitly in terms of $\boldsymbol{x}$ and $\boldsymbol{y}$ in order to proceed with the optimization.

For convenience, we remove the dependency on $\{\boldsymbol{x},\boldsymbol{y}\}$ in the sequel. Observe that starting from state $(0,0)$, the system can go to any other state $(j,0)$, $0 \leq j \leq B-1$, by the next status update, i.e., after only *one* update, each with some probability. Then, from state $(j,0)$, $1 \leq j \leq B-1$, the system can only go to one of the following states by the next update: $\{(j-1,0),\ (j,0), \ldots, (B-1,0)\}$, each with some probability. We denote by $p_{i,j}$ the probability of going from state $(i,0)$ to state $(j,0)$ after one update. Clearly, $p_{i,j} = 0$ for $j \leq i-2$. We also denote by $r_{i,j}$ and $\ell_{i,j}$ the area under the age curve and the time taken when the system goes from state $(i,0)$ to state $(j,0)$ in one update, respectively. Finally, since the goal is to compute the area under the age curve in an epoch together with the epoch length, we define R_j and L_j as the area under the age curve and the time taken to go from state $(j,0)$ back to $(0,0)$ again (in, however, many number of updates). See Figure 8.3 where we depict the relationships between the previous variables/notation in the form of a tree graph. The graph basically represents the transitions between different system states (nodes on the graph) after only *one* update, which occur with probabilities indicated on the arrows in the graph that connects the nodes. We emphasize that, for instance, state $(0,0)$ in the first column of the graph is no different than state $(0,0)$ in the second column, and that the arrow connecting them merely represents a loop connecting a

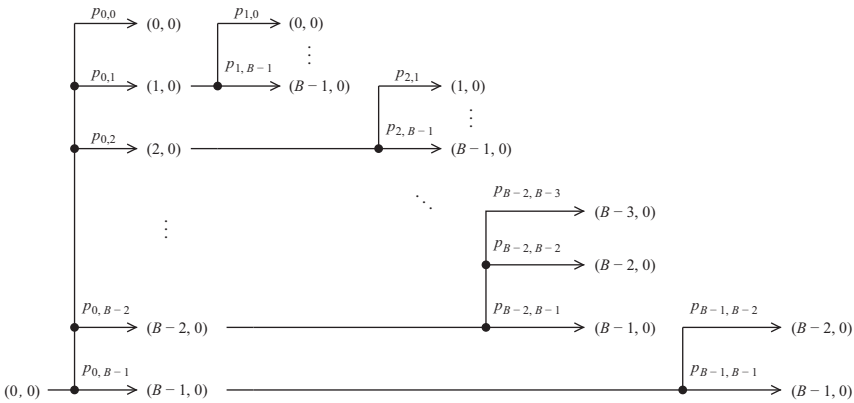


Figure 8.3 Transitions among system states after only one update. Each transition from state (i, 0) to (j, 0) occurs with probability $p_{i,j}$ as indicated on the tree branches

state to itself; we chose to expand such loop horizontally for clarity of presentation. From the graph, one can write the following equations:

$$\mathbb{E}[R] = p_{0,0}\mathbb{E}\left[r_{0,0}\right] + \sum_{j=1}^{B-1} p_{0,j}\left(\mathbb{E}\left[r_{0,j}\right] + \mathbb{E}\left[R_j\right]\right), \tag{8.15}$$

$$\mathbb{E}[L] = p_{0,0}\mathbb{E}\left[\ell_{0,0}\right] + \sum_{j=1}^{B-1} p_{0,j}\left(\mathbb{E}\left[\ell_{0,j}\right] + \mathbb{E}\left[L_j\right]\right). \tag{8.16}$$

Through meticulously involved analysis, the previous variables can be expressed explicitly in terms of $\boldsymbol{x}$ and $\boldsymbol{y}$. We refer the reader to [44] for the fine details. We are now ready to find the optimal $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$.

8.2.3.2 Multi threshold policy

One of the main findings of [44] is showing that the optimal status update policy has a multi threshold structure, in the sense that y_j is an x_j-threshold policy, $1 \leq j \leq B-1$, and that y_B is a λ-threshold policy. Moreover, the optimal thresholds x_js are all found, in closed-form, in terms of λ. For instance, for $B = 4$, this allows for expressing (see the derivation details in [44]):

$$\begin{aligned} p_4(\lambda) = {} & e^{-\lambda}\left(\frac{1}{6}\lambda^3 + \frac{3}{2}\lambda^2 + 6\lambda + 10\right) - \frac{1}{2}\lambda^2 - (x_1 - \lambda) - (x_2 - \lambda) - (x_3 - \lambda) \\ & - (x_1 + 2)e^{-x_1} - \left(\frac{1}{2}x_2^2 + 2x_2 + 3\right)e^{-x_2} - \left(\frac{1}{6}x_3^3 + x_3^2 + 3x_3 + 4\right)e^{-x_3}, \end{aligned} \tag{8.17}$$

where

$$x_3 = \log\left(\frac{1}{e^{-\lambda} - \frac{1}{2}\lambda^2}\right), \tag{8.18}$$

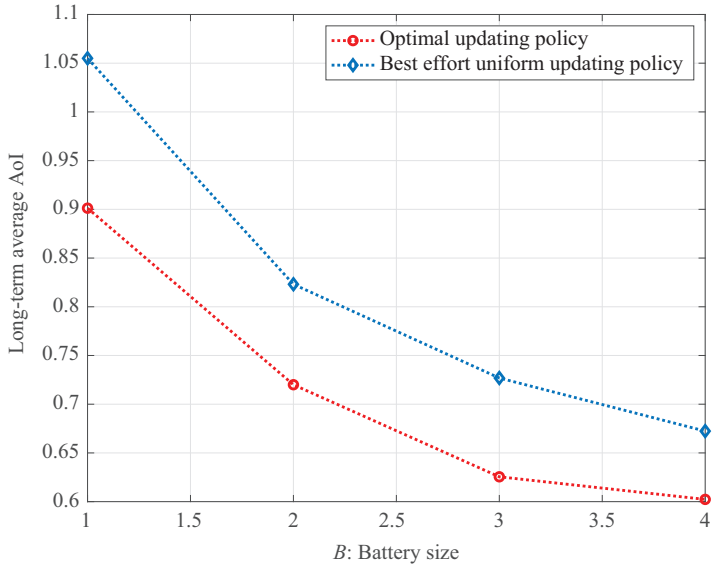


Figure 8.4 *Comparison of long-term average age versus battery size under optimal and best effort update policies*

$$x_2 = -\log\left((\lambda+1)e^{-\lambda} + \frac{1}{2}\lambda^2 + \lambda + \left(e^{-\lambda} - \frac{1}{2}\lambda^2 + 1\right)\log\left(e^{-\lambda} - \frac{1}{2}\lambda^2\right)\right), \tag{8.19}$$

$$x_1 = -\log\left(\left(\frac{1}{2}\lambda^2 + \lambda + 1\right)e^{-\lambda} - x_2\left(e^{-x_2} + 1\right) - x_3\left(\frac{1}{2}x_3 e^{-x_3} - 1\right)\right). \tag{8.20}$$

Solving the previous for $p_4(\lambda^*) = 0$, we get that $\lambda^* \approx 0.6023$, with the remaining thresholds $x_3^* \approx 1.005$, $x_2^* \approx 1.243$, and $x_1^* \approx 1.636$.

We note that the thresholds are monotonically decreasing: $x_1^* > x_2^* > x_3^* > \lambda^*$, which has an intuitive explanation; it basically says that the sensor is less eager to send an update when it has relatively lower energy available in its battery than it is when it has relatively higher energy available.

To show the gain achieved with respect to another baseline, Figure 8.4 shows the gap between the optimal multi threshold policy and the BU updating policy (which is optimal for $B = \infty$).

8.3 Status updating over erasure channels

In this section, we describe the results reported in [36,47,48] in more detail. The system model is very similar to that described in Section 8.2, yet with the main difference in that *status updates are subject to erasures.* Specifically, the communication channel between the sensor and the destination is modeled as a time-invariant noisy

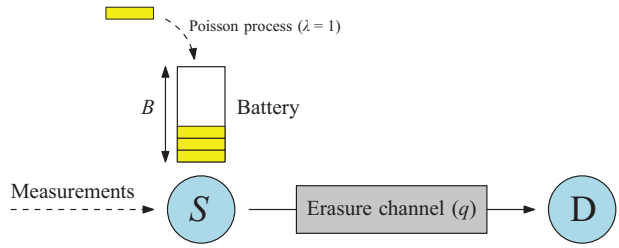


Figure 8.5 System model for status updating over erasure channels

channel, in which each update transmission gets erased with probability $q \in (0, 1)$, independently from other transmissions, see Figure 8.5. We differentiate between two main cases in our treatment:

1. *No updating feedback.* In this case, the sensor has no knowledge of whether an update is successful. It can only use the up-to-date energy arrival profile and status updating decisions as well as the statistical information, such as the energy arrival rate and the erasure probability of the channel, to decide the upcoming updating time points.
2. *Perfect updating feedback.* In this case, the sensor receives an instantaneous, error-free feedback when an update is transmitted. Therefore, it can decide when to update next based on the feedback information, along with the information it uses for the no feedback case.

Since each update transmission is not necessarily successful, we denote by $\{l_i\}$ the set of update transmission times and by $\{s_i\}$ the times of the *successful* ones of which. Therefore, in general, $\{s_i\} \subseteq \{l_i\}$. The energy causality constraint in (8.2) now becomes:

$$\mathscr{E}\left(l_i^-\right) \geq 1, \quad \forall i, \tag{8.21}$$

and the battery evolution in (8.3) becomes:

$$\mathscr{E}\left(l_i^-\right) = \min\{\mathscr{E}\left(l_{i-1}^-\right) - 1 + \mathscr{A}(x_i), B\}, \quad \forall i, \tag{8.22}$$

where $x_i \triangleq l_i - l_{i-1}$ now denotes the inter-update attempt delay. We assume $s_0 = l_0 = 0$ without loss of generality, i.e., the system starts with fresh information at time 0. We denote by $\mathscr{F}_q$, the set of feasible transmission times $\{l_i\}$ described by (8.2) and (8.3) in addition to an empty battery at time 0, i.e., $\mathscr{E}(0) = 0$.[§]

Let us denote by $y_i \triangleq s_i - s_{i-1}$ the *successful* inter-update delay and by $n(t)$ denote the number of updates that are *successfully* received by time t. We are interested in the average AoI given by the area under the age evolution curve (see Figure 8.6), which is given by

$$r(t) = \frac{1}{2}\sum_{i=1}^{n(t)} y_i^2 + \frac{1}{2}\left(t - s_{n(t)}\right)^2. \tag{8.23}$$

[§]In [36], it is assumed that $\mathscr{E}(0) = 1$ to simplify the analysis. For $\mathscr{E}(0) = 0$, the same results would follow after slightly modifying the proofs. We set $\mathscr{E}(0) = 0$ for consistency of the chapter's framework.

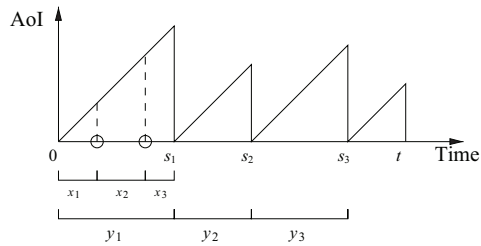


Figure 8.6 Age evolution versus time with $n(t) = 3$ successful updates. Circles denote failed attempts. In this example, the first update is successfully received after three update attempts

The goal is to choose a set of feasible transmission times $\{l_1, l_2, l_3, \dots\} \in \mathscr{F}_q$, or equivalently $\{x_1, x_2, x_3, \dots\}$, such that the long-term average AoI is minimized. Therefore, the goal is to characterize the optimal long-term average AoI, as a function of the battery size, $\rho_q^{\omega}(B)$ by solving:

$$\rho_q^{\omega}(B) \triangleq \min_{\{x_i\} \in \mathscr{F}_q} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[r(T)\right], \tag{8.24}$$

where the superscript $\omega \equiv$ noFB in the case without updating feedback, and $\omega \equiv$ wFB in the case with perfect feedback.

In the following subsections, we present the solution of (8.24) for $B = \infty$ followed by the special case of $B = 1$, in view of the previous two feedback availability cases.

8.3.1 The case $B = \infty$

For the case $B = \infty$, without updating feedback, [36] shows that the BU updating policy, as per Definition 8.1, is optimal, while for the scenario with perfect feedback, [36] proposes a *BU with retransmission (BUR) updating policy* and shows its optimality. Optimality in both the cases is shown by first evaluating a lower bound on the long-term average AoI and then showing that it can be achieved by BU updating.

Although the proposed policies are quite intuitive, their optimality is quite challenging to establish, compared with the scenario of Section 8.2.1. This is because both battery outages and updating erasures will affect the AoI under the proposed policies. While the impact of either of those two events can be analyzed relatively easily when isolated, it becomes extremely challenging when both of them are involved. Moreover, when there exists perfect updating feedback to the sensor, updating erasures under the BUR will lead to subsequent retransmissions and energy consumption, thus affecting the battery outage probability in the future. Such complicated interplay between those two events makes the problem even more complicated.

In order to overcome such difficulties, [36] proposes a novel *virtual-policy*-based approach. Specifically, for both BU and BUR updating policies, a sequence of virtual policies is constructed, which are strictly suboptimal to their original counterparts, and eventually converge to them. Leveraging the virtual policies, the effects of battery

outages and updating errors could be decoupled in the performance analysis. Finally, the long-term average AoI under the virtual policies is shown to converge to the corresponding lower bound, which implies the optimality of the original policy.

We elaborate on the earlier over the next two subsections.

8.3.1.1 Updating without feedback

A technical condition is introduced to facilitate the following analysis, which basically states that the focus is on *optimal status update policies that have inter-update attempt times that have bounded first moment.* Policies abiding by this condition are termed *bounded updating policies* in [36]. Focus on such policies arises from the fact that we are interested in policies that achieve finite long-term average AoI. Now the following lower bound holds [36]:

$$\rho_q^{\text{noFB}}(\infty) \geq \frac{1+q}{2(1-q)}. \tag{8.25}$$

The previous can be shown by replacing the energy causality constraint by its long-term average, as in the approach used to show the lower bound in (8.6).

To achieve the abovementioned bound, let us introduce the next virtual policy.

Definition 8.2 (BU-ER$_{\mathbf{T_0}}$ [36]). *The sensor performs BU updating until the battery level after sending an update becomes zero for the first time or until time T_0^+, in which case the sensor depletes its battery. After that, when the battery level becomes higher than or equal to 1 after a successful update for the first time, the sensor reduces the battery level to 1 and then repeats the process.*

Observe that as $T_0 \to \infty$, the BU-ER$_{T_0}$ updating policy becomes a BU updating policy. The next result now holds the following.

Lemma 8.2 ([36]). *For any $T_0 > 0$, BU-ER$_{T_0}$ updating policy is suboptimal to the BU updating policy.*

Proof. Note that BU-ER$_{T_0}$ updating is identical to BU updating except the energy removal at time T_0 and when $\mathscr{E}(s_n^+)$ becomes higher than 1. Given the same energy harvesting sample path, the battery level under BU is always higher than that under BU-ER$_{T_0}$. Thus, BU-ER$_{T_0}$ incurs more infeasible status updates. With the same *update erasure pattern*, the instantaneous AoI under BU-ER$_{T_0}$ updating is always greater than or equal to that under BU updating sample path wisely. Thus, the expected time-average AoI under BU-ER$_{T_0}$ is greater than or equal to that under BU, which proves the lemma. □

One can now show the following.

Theorem 8.4 ([36]). *As $T_0 \to \infty$, the long-term average AoI under the BU-ER$_{T_0}$ updating policy is upper bounded by $\frac{1+q}{2(1-q)}$.*

Proof sketch. Observe that the BU-ER$_{T_0}$ updating policy is a *renewal*-type policy, i.e., the states of the system evolve according to a renewal process. To see this, we note that the updating process under BU-ER$_{T_0}$ works in cycles, where each cycle begins

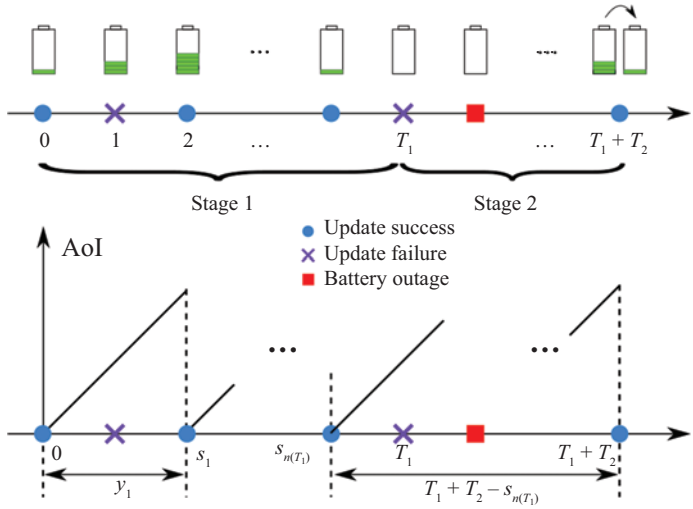


Figure 8.7 An illustration of the BU-ER$_{T_0}$ updating policy and the battery level right after each updating epoch. AoI will be reset to zero at the successful updating points

with the initial battery level to be 1 and the AoI to be 0, followed by i.i.d. battery and AoI evolution processes. Therefore, to analyze the expected long-term average AoI, it suffices to analyze the expected average AoI over one renewal interval. In the following, we will focus on the first renewal interval and show that the corresponding expected average AoI converges to the lower bound in (8.25) as T_0 increases. As illustrated in Figure 8.7, the renewal interval consists of two stages. The first stage starts at time zero and ends until the battery becomes empty for the first time or until time T_0^+. We denote T_1 as the end of the first stage. We note that all scheduled status updating points over $(0, T_1]$ are feasible. The second stage starts at T_1 and ends when the battery level becomes higher than or equal to 1 after a successful update for the first time after T_1. We denote the duration of the second stage as T_2. The second stage thus ends at $T_1 + T_2$.

By renewal theory properties of BU-ER$_{T_0}$ policy, the result can be shown. We refer the reader to [36] for more details. □

By the previous theorem, it is readily shown that the BU updating policy achieves the lower bound in (8.25) and is therefore optimal, since the BU-ER$_{T_0}$ updating policy becomes a BU updating policy as $T_0 \to \infty$.

In Figure 8.8, some numerical evaluations of the virtual policies BU-ER$_{T_0}$ for different value of T_0 are presented for $q = 0.4$. For comparison, a greedy updating policy is also evaluated, in which the sensor updates instantly when one unit of energy arrives. From the figure, greedy updating has the highest AoI and never approaches the lower bound. On the other hand, the time average AoI the BU-ER$_{T_0}$ updating policy is monotonically decreasing as T_0 increases and gradually approaches that under the BU updating policy, which is, in turn, approaching the lower bound.

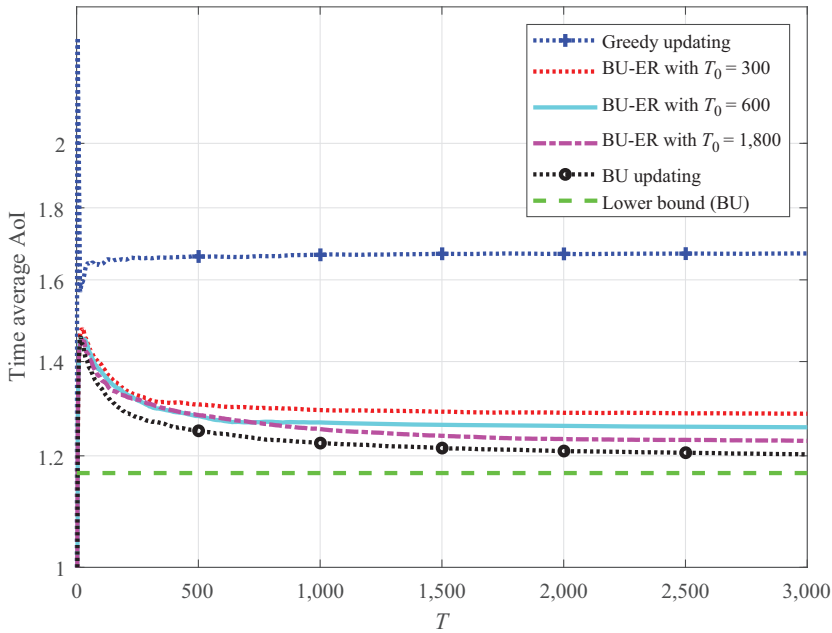


Figure 8.8 AoI under variant BU-ER updating policies, BU updating policy, and greedy updating policy versus time T. The lower bound is also plotted for comparison

8.3.1.2 Updating with perfect feedback

With perfect updating feedback, the sensor has the choice to retransmit the update immediately or wait and update later, thus leading to optimal solutions different from the no feedback case. A technical condition that will facilitate the analysis is that *optimal status update policies have inter-update attempt times of bounded second moment and such that the number of update attempts in a given time is bounded above by some constant multiplied by the length of the time frame, on average.* Such policies are denoted as uniformly bounded policies in [36].

Define K_i as the number of *attempted* updates (including the last successful one) between two successful updates at times s_{i-1} and s_i. Thus, $K_i \geq 1$. To find a lower bound, as usual, we replace the energy causality constraint by its long-term average rate. Let us denote such policies by *average energy constraint policies*. The following result now holds:

Theorem 8.5 ([36]). *The optimal average energy constraint policy is of a renewal structure, in which $\{s_i\}$ forms a renewal process. In addition, y_i only depends on K_i.*

Proof sketch. For any given average energy constraint policy, we first average y_i over sample paths with the same K_i, so that all factors other than K_i that may affect y_i are averaged out. Then, we form a linear combination of y_i and use it as the inter-update delay under the new policy. Such a policy is a renewal policy, and each renewal

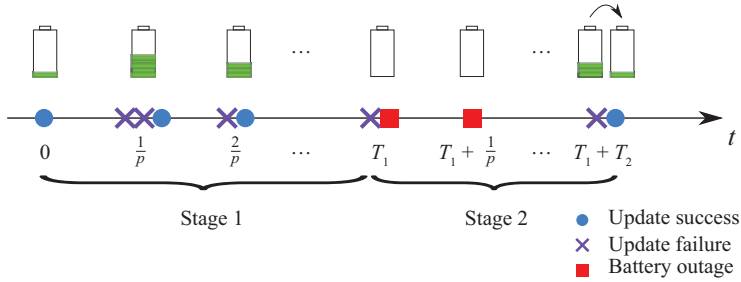


Figure 8.9 An illustration of the BUR-ER$_{T_0}$ updating policy and the battery level right after each updating epoch. AoI will be reset to zero at the successful updating points

interval only depends on K_i. Through rigorous stochastic analysis, we prove that the constructed renewal policy always outperforms the original policy. □

Focusing on renewal policies, it is then shown in [36] that the optimal $y_i = \frac{1}{1-q}$, regardless of the value of K_i. Based on that the following lower bound is established:

$$\rho_q^{wFB}(\infty) \geq \frac{1}{2(1-q)}. \tag{8.26}$$

Motivated by the structure of the constant y_i in the previous optimal renewal-type average energy constraint policy, let us define the following *BUR updating policy.*

Definition 8.3 (BUR updating [36]). *The sensor is scheduled to send a new update at $s_n = \frac{n}{1-q}$, $n = 1, 2, 3, \ldots$. The sensor keeps sending updates at s_n until an update is successful or until it runs out of battery. Otherwise, the sensor keeps silent until the next scheduled status update time.*

In order to prove that the BUR updating policy is optimal, we will first construct a sequence of policies that are suboptimal to the BUR updating policy and show that the limit of those suboptimal policies achieves the lower bound in (8.26).

Definition 8.4 (BUR with energy removal (BUR-ER$_{T_0}$) [36]). *The sensor performs BUR updating policy until the battery level after sending an update becomes zero for the first time, or until time T_0^+, in which case the sensor depletes its battery after a successful update at T_0. After that, when the battery level becomes higher than or equal to 1 after a successful update for the first time, the sensor reduces the battery level to 1 and then repeats the process.*

As in Lemma 8.2, one can show that the BUR-ER$_{T_0}$ updating policy is suboptimal to the BUR updating policy. One can now show the following.

Theorem 8.6 ([36]). *As $T_0 \to \infty$, the expected long-term average AoI under BUR-ER$_{T_0}$ updating is upper bounded by $\frac{1}{2(1-q)}$.*

Proof sketch. Note that BUR-ER$_{T_0}$ updating is a renewal policy, and Figure 8.9 is an illustration of one renewal interval. In order to analyze the expected long-term average AoI, it suffices to analyze the expected average AoI over one renewal interval. Thus,

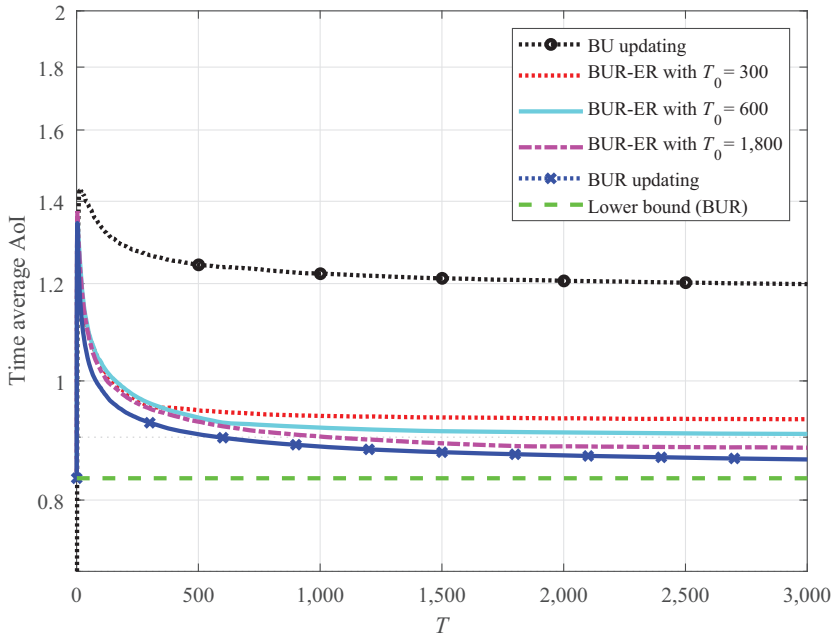


Figure 8.10 AoI under variant BU-ER updating policies, BU updating policy, and greedy updating policy versus time T. The lower bound is also plotted for comparison

we will focus on the first renewal interval and show that the expected average AoI converges to the lower bound in (8.26). The renewal interval consists of two stages. The first stage starts at time zero and ends until the battery becomes empty for the first time, or until time T_0^+, denoted as T_1. We note that all scheduled updating points over $(0, T_1)$ are feasible. The second stage starts at T_1 and ends when the battery level after a successful update becomes higher than or equal to 1 for the first time after T_1, denoted as $T_1 + T_2$, where T_2 is the duration of the second stage.

Applying renewal theory, we obtain the result. We refer the reader to [36] for more details. □

By Theorem 8.6, it is readily shown that the BUR updating policy achieves the lower bound in (8.26) and is therefore optimal, since the BUR-ER$_{T_0}$ updating policy becomes a BUR updating policy as $T_0 \to \infty$.

The performances of the BUR-ER$_{T_0}$ updating policy are numerically evaluated in Figure 8.10 for $q = 0.4$ and for various values of T_0. As a comparison, the time average AoI under the BU updating and the BUR updating policies is also plotted in the figure. It is observed that the AoI under BUR-ER$_{T_0}$ gradually decreases and approaches that under the BUR updating policy as T_0 increases. The performance gap between the BU updating and the BUR updating indicates that exploiting updating feedback can significantly reduce time average AoI in the system.

8.3.2 The case $B = 1$

We now focus on the special finite battery case of $B = 1$, in which one update completely depletes the battery. Similar to the finite battery case analysis of Section 8.2, it will be shown that an *erasure-dependent threshold policy* is optimal for the case without feedback. For the case with perfect feedback, the focus will be on a class of policies denoted *threshold-greedy policies*. In both cases, for technical correctness, we will focus on the class of *uniformly bounded policies*, for which inter-update attempt times have a bounded second moment. Focusing on this class of policies, we have the following structural result.

Theorem 8.7 ([47,48]). *The optimal status updating policy that solves problem (8.24) for $B = 1$, and for both cases with and without updating feedback, is a renewal policy in the sense that the* actual *update times sequence $\{s_i\}$ forms a renewal process, and the* actual *inter-update times y_is are i.i.d.*

The renewal structure in the previous theorem greatly reduces the complexity of the problem. Now we only need to look at one renewal interval and optimize the updating policy over it. This is done differently for the case without feedback and that with perfect feedback, as we discuss next. We will use the term *epoch* to denote the time in between two successful updates.

8.3.2.1 Updating without feedback

Considering one renewal interval (one epoch), let τ denote the time until the first energy arrival occurs. Then, problem (8.24) reduces to

$$\begin{aligned} \rho_q^{\text{noFB}}(1) = \min_{x(\cdot)} \quad & \frac{\mathbb{E}[R]}{\mathbb{E}[y]} \\ \text{s.t.} \quad & x(\tau) \geq \tau, \quad \forall \tau, \end{aligned} \tag{8.27}$$

where the random variable R denotes the area under the age curve in the epoch, and y is the epoch's length. We now introduce the auxiliary problem:

$$\begin{aligned} p_{q,1}^{\text{noFB}}(\lambda) \triangleq \min_{x(\cdot)} \quad & \mathbb{E}[R] - \lambda \mathbb{E}[y] \\ \text{s.t.} \quad & x(\tau) \geq \tau, \quad \forall \tau, \end{aligned} \tag{8.28}$$

for some $\lambda \geq 0$ and focus on finding λ^* that solves $p_{q,1}^{\text{noFB}}(\lambda^*) = 0$, which is equal to $\rho_q^{\text{noFB}}(1)$ [57]. The next theorem characterizes the solution of problem (8.28).

Theorem 8.8 ([47]). *The optimal solution of problem (8.28) depends on q. If $q < \frac{1}{2}$, then it is a λ'-threshold policy, in which:*

$$x(t) = \begin{cases} \lambda', & t < \lambda' \\ t, & t \geq \lambda' \end{cases}, \tag{8.29}$$

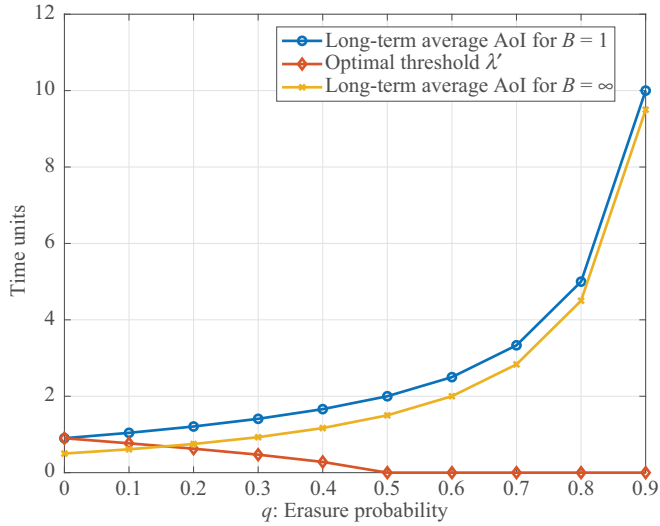


Figure 8.11 Comparison of optimal AoI for $B = 1$ and $B = \infty$, along with λ', versus q, for the case without updating feedback

where λ' is the unique solution of

$$\frac{1+q}{1-q}\lambda' + \frac{2q}{1-q}e^{-\lambda'} = \lambda. \tag{8.30}$$

Otherwise, if $q \geq \frac{1}{2}$, then the optimal solution is greedy, in the sense that $x(t) = t\ \forall t$.

The previous theorem generalizes the result for the $B = 1$ case without erasures, whose solution was a λ-threshold policy, to an *erasure-dependent* λ'-threshold policy in the case of erasures without updating feedback. The value of q determines the threshold λ'. It can be shown that λ' is actually nonincreasing in q, which is quite intuitive, since the sensor should be more eager to send new updates if the erasure probability is high, so that when the update is eventually received successfully the AoI would not be large.

In Figure 8.11, we present some numerical evaluations of $\rho_q^{\text{noFB}}(1)$ versus q. We compare with $\rho_q^{\text{noFB}}(\infty)$ as a baseline. We also show the nonincreasing nature of the threshold λ'.

8.3.2.2 Updating with perfect feedback

In the case with perfect updating feedback available, the optimal policy in a single epoch consists of a possibly infinite sequence $\{x_1, x_2, x_3, \dots\} \triangleq \boldsymbol{x}$, where x_i now denotes the time elapsed from the *beginning of the epoch* until the ith update attempt.[‖]

[‖]We slightly deviate from the original definition of x_i as the ith inter-update attempt time and assume without loss of generality that the epoch starts at time 0.

Let τ_1 denote the time until the first energy arrival in the epoch, and τ_i, $i \geq 2$ denote the time until energy arrives *after* the ith update attempt, i.e., after time x_i. This way, problem (8.24) reduces to:

$$\begin{aligned}\rho_q^{\text{wFB}}(1) \triangleq \min_{\boldsymbol{x}} \quad & \frac{\mathbb{E}\left[R(\boldsymbol{x})\right]}{\mathbb{E}\left[L(\boldsymbol{x})\right]} \\ \text{s.t.} \quad & x_1(\tau_1) \geq \tau_1 \\ & x_2(\tau_2 + x_1(\tau_1)) \geq \tau_2 + x_1(\tau_1) \\ & x_3(\tau_3 + x_2(\tau_2 + x_1(\tau_1))) \geq \tau_3 + x_2(\tau_2 + x_1(\tau_1)) \\ & \vdots, \end{aligned} \tag{8.31}$$

where the inequalities represent the energy causality constraints. As before, we introduce the following auxiliary problem for some $\lambda \geq 0$:

$$\begin{aligned} p_{q,1}^{\text{wFB}}(\lambda) \triangleq \min_{\boldsymbol{x}} \quad & \mathbb{E}\left[R(\boldsymbol{x})\right] - \lambda \mathbb{E}\left[L(\boldsymbol{x})\right] \\ \text{s.t.} \quad & \text{problem (8.31)'s constraints} \end{aligned} \tag{8.32}$$

and focus on finding λ^* that solves $p_{q,1}^{\text{wFB}}(\lambda^*) = 0$, which is equal to $\rho_q^{\text{wFB}}(1)$ [57].

We focus on an intuitive class of policies in which the first update attempt x_1 has a threshold structure, and the following attempts, if the first is not successful, $\{x_2, x_3, \dots\}$ follow a greedy structure. This class is intuitive because if the first update is unsuccessful, then the AoI has already grown to a relatively high value, which urges the sensor to transmit its following updates as soon as energy is available. This class of policies is denoted *threshold-greedy policies* in [48]. We now have the following lemma.

Lemma 8.3 ([48]). *For problem (8.32), if x_i, $i \geq 2$ are all greedy policies, then the optimal x_1 is a γ-threshold policy with $\gamma = \left[\lambda - \frac{q}{1-q}\right]^+$. Conversely, if the optimal x_1 is a γ-threshold policy, then the optimal x_i, $i \geq 2$ are all greedy policies.*

The previous lemma shows that threshold-greedy policies are not just intuitive but are actually representing a fixed-point solution of the problem. The next theorem characterizes the optimal threshold-greedy policy.

Theorem 8.9 ([48]). *The optimal threshold-greedy policy that solves problem (8.32) is such that $\gamma^* = \lambda^* - \frac{q}{1-q} > 0$, with λ^* being the unique solution of:*

$$e^{-\left(\lambda^* - \frac{q}{1-q}\right)} + \frac{2q - q^2}{2(1-q)^2} = \frac{1}{2}(\lambda^*)^2 . \tag{8.33}$$

In Figure 8.12, we present some numerical evaluations of $\rho_q^{\text{wFB}}(1)$ versus q. We compare with $\rho_q^{\text{wFB}}(\infty)$ as a baseline. We also plot the optimal threshold $\gamma^* = \lambda^* - \frac{q}{1-q}$.

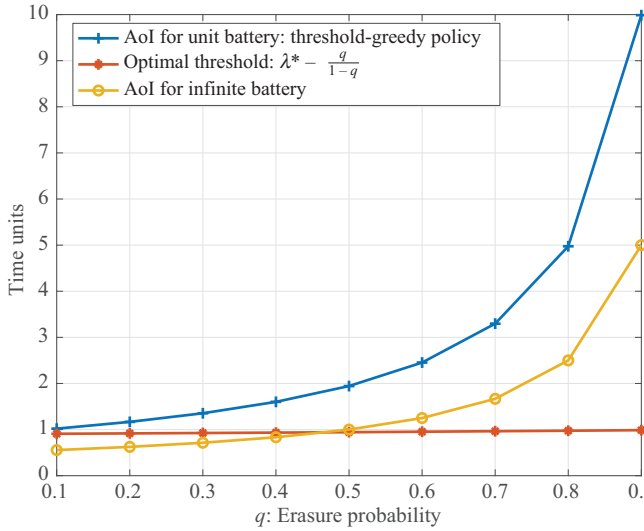


Figure 8.12 Comparison of optimal AoI for $B = 1$ and $B = \infty$, along with γ^* versus q, for the case with perfect updating feedback

8.4 Conclusion and outlook

In this chapter, the notion of AoI has been introduced as the main performance metric for energy harvesting communication systems. AoI measures latency through a different lens when compared to other conventional metrics such as throughput and transmission delay, namely, through measuring the time difference since the generation of the latest data that has reached the destination. Hence, the optimal transmission policies drastically change under its assessment. One common theme among the different transmission policies presented in this chapter has been that *greedy is not always optimal:* to minimize the AoI, on average, transmitters need to evenly spread out their updates over time, up to the extent allowed by the energy availability. This is achieved by *best effort* policies for the infinite battery cases, and by *threshold* policies in the finite battery ones.

While several system models have been discussed in detail, there are various open problems that could be tackled in this line of research. Two of which are highlighted next.

1. *Generalizing the Poisson energy arrivals model.* So far, the Poisson energy arrival model has been crucial in determining the structure of the optimal policy, particularly through the memoryless property of the exponential distribution that governs the inter-arrival times. While numerically it has been shown in some of the discussed works in this chapter that the optimal policies work relatively well under other arrival models, such as Markovian, extending the results *analytically*

is definitely of interest. One possible approach toward that end is to show that the derived best effort or threshold-type policies perform within some confidence interval away from optimal solutions (or from some lower bounds on them), since these policies are relatively simple to implement in practice.

2. *Timely networking for energy harvesting nodes.* Another interesting aspect is to extend the point-to-point models into multiterminal communications, i.e., broadcast, multiple-access, and interference channel models. There, instead of characterizing an AoI-optimal policy, the goal would be to characterize an AoI-optimal *region* that represents how the different transmitters/receivers of the system interact.

References

[1] S. K. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE Int. Conf. Comput. Commun.*, March 2012.

[2] R. D. Yates and S. K. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2012.

[3] C. Kam, S. Kompella, and A. Ephremides, "Age of information under random updates," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2013.

[4] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, 2016.

[5] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[6] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[7] Y. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[8] R. D. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[9] R. D. Yates, P. Ciblat, A. Yener, and M. A. Wigger, "Age-optimal constrained cache updating," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[10] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, 2020.

[11] J. Zhong, E. Soljanin, and R. D. Yates, "Status updates through multicast networks," in *Proc. Allerton Conf. Commun. Control Comput.*, October 2017.

[12] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[13] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Found. Trends Netw.*, vol. 12, no. 3, pp. 162–259, 2017.

[14] E. Najm and E. Telatar, "Status updates in a multi-stream M/G/1/1 preemptive queue," in *Proc. IEEE Int. Conf. Comput. Commun.*, April 2018.
[15] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. Int. Symp. Theory, Alg. Found., Protocol Design for Mobile Netw. Mobile Comput.*, June 2018.
[16] B. Zhou and W. Saad, "Optimal sampling and updating for minimizing age of information in the internet of things," in *Proc. IEEE Global Commun. Conf.*, December 2018.
[17] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.
[18] M. Zhang, A. Arafa, J. Huang, and H. V. Poor, "How to price fresh data," in *Proc. Int. Symp. Modeling Opt. Mobile Ad Hoc Wireless Netw.*, June 2019.
[19] M. Bastopcu and S. Ulukus, "Minimizing age of information with soft updates," *J. Commun. Netw.*, vol. 21, no. 3, pp. 233–243, 2019.
[20] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in multihop multicast networks," *J. Commun. Netw.*, vol. 21, no. 3, pp. 256–267, 2019.
[21] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," [Online]. Available from: arXiv:1902.03552.
[22] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "On timely channel coding with hybrid ARQ," in *Proc. IEEE Global Commun. Conf.*, December 2019.
[23] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization," [Online]. Available from: arXiv:1905.13743.
[24] B. Wang, S. Feng, and J. Yang, "When to preempt? age of information minimization under link capacity constraint," *J. Commun. Netw.*, vol. 21, no. 3, pp. 220–232, 2019.
[25] P. Zou, O. Ozel, and S. Subramaniam, "Trading off computation with transmission in status update systems," in Proc. *IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun.*, September 2019.
[26] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2015.
[27] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. Inf. Theory Appl. Workshop*, February 2015.
[28] W. Liu, X. Zhou, S. Durrani, H. Mehrpouyan, and S. D. Blostein, "Energy harvesting wireless sensor networks: Delay analysis considering energy costs of sensing and transmission," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4635–4650, 2016.
[29] A. Arafa and S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *Proc. IEEE Global Commun. Conf.*, December 2017.
[30] A. Arafa and S. Ulukus, "Timely updates in energy harvesting two-hop networks: Offline and online policies," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4017–4030, 2019.

[31] A. Arafa and S. Ulukus, "Age minimization in energy harvesting communications: Energy-controlled delays," in *Proc. Asilomar Conf. Signals Syst. Comput.*, October 2017.

[32] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017.

[33] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 193–204, 2018.

[34] S. Feng and J. Yang, "Optimal status updating for an energy harvesting sensor with a noisy channel," in *Proc. IEEE Int. Conf. Comput. Commun.*, April 2018.

[35] S. Feng and J. Yang, "Minimizing age of information for an energy harvesting source with updating failures," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2018.

[36] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: Without and with feedback," [Online]. Available from: arXiv:1808.05141.

[37] A. Baknina and S. Ulukus, "Coded status updates in an energy harvesting erasure channel," in *Proc. Conf. Inf. Sci. Syst.*, March 2018.

[38] S. Farazi, A. G. Klein, and D. R. Brown. III, "Average age of information for status update systems with an energy harvesting server," in *Proc. IEEE Int. Conf. Comput. Commun.*, April 2018.

[39] S. Farazi, A. G. Klein, and D. R. Brown. III, "Age of information in energy harvesting status update systems: When to preempt in service?," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2018.

[40] X. Zheng, S. Zhou, Z. Jiang, and Z. Niu, "Closed-form analysis of non-linear age-of-information in status updates with an energy harvesting transmitter," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4129–4142, 2019.

[41] A. Baknina, O. Ozel, J. Yang, S. Ulukus, and A. Yener, "Sending information through status updates," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2018.

[42] A. Arafa, J. Yang, and S. Ulukus, "Age-minimal online policies for energy harvesting sensors with random battery recharges," in *Proc. IEEE Int. Conf. Commun.*, May 2018.

[43] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal online policies for energy harvesting sensors with incremental battery recharges," in *Proc. Inf. Theory Appl. Workshop*, February 2018.

[44] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 534–556, 2020.

[45] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu, "Achieving the age-energy tradeoff with a finite-battery energy harvesting source," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2018.

[46] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu, "Optimal status updating with a finite-battery energy harvesting source," *J. Commun. Netw.*, vol. 21, no. 3, pp. 280–294, 2019.

[47] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Online timely status updates with erasures for energy harvesting sensors," in *Proc. Allerton Conf. Commun. Control Comput.*, October 2018.

[48] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Using erasure feedback for online timely updating with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2019.

[49] I. Krikidis, "Average age of information in wireless powered sensor networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 628–631, 2019.

[50] Z. Chen, N. Pappas, E. Bjornson, and E. G. Larsson, "Age of information in a multiple access channel with heterogeneous traffic and an energy harvesting node," in *Proc. IEEE Int. Conf. Comput. Commun.*, May 2019.

[51] S. Leng and A. Yener, "Age of information minimization for an energy harvesting cognitive radio," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 2, pp. 427–439, 2019.

[52] G. Stamatakis, N. Pappas, and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms," in *Proc. IEEE Global Commun. Conf.*, December 2019.

[53] P. Rafiee and O. Ozel, "Active status update packet drop control in an energy harvesting node," in *Proc. IEEE Intl. Workshop Signal Proc. Advances Wireless Commun.*, May 2020.

[54] O. Ozel, "Timely status updating through intermittent sensing and transmission," in *Proc. IEEE Intl. Symp. Inf. Theory*, June 2020.

[55] Y. Dong, P. Fan, and K. B. Letaief, "Energy harvesting powered sensing in IoT: Timeliness versus distortion," *IEEE Internet of Things J.*, to appear.

[56] S. M. Ross, *Stochastic Processes*. Wiley, 1996.

[57] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, 1967.

[58] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

Part III

Energy-efficient techniques and concepts for future networks

This page intentionally left blank

Chapter 9

Fundamental limits of energy efficiency in 5G multiple antenna systems

Andrea Pizzo[1], Luca Sanguinetti[2] and Emil Björnson[3]

9.1 A primer on energy efficiency

Owing to the prolific spread of Internet-enabled mobile devices and the ever-growing volume of mobile data communications, the biggest challenge in the wireless industry today is to meet the soaring demand for wireless broadband required to ensure consistent quality of service in a network. Rising to this challenge means increasing the network capacity by a 1,000-fold over the next decade. If the current technology is used to achieve this, the resulting power consumption and energy-related pollution are expected to give rise to major societal, economic, and environmental issues that would render this growth unsustainable. Therefore, the information and communications technology industry is faced with a formidable mission: the cellular technology must improve so that the network capacity is increased significantly in order to accommodate higher data rates in the years to come, but this task must be accomplished under an extremely tight energy budget.

The power consumption of communication networks has been neglected for decades by the International Telecommunication Union (ITU) that provides the minimum performance requirements for every cellular generation. For example, 4G was designed to satisfy the IMT-Advanced requirements on spectral efficiency (SE), bandwidth, latency, and mobility with no mention of energy consumption [1]. As a first step towards this direction, the ITU has recently included the energy efficiency (EE) metric in its latest release of requirements [2], which are supposed to be satisfied by the new 5G standard for being an IMT-2020 radio interface. However, no concrete measurable energy efficiency targets are provided in the list of requirements. A basic definition of the EE is:

$$\mathrm{EE}\,(\mathrm{bit/J}) = \frac{\text{data rate}\,(\mathrm{bit/s})}{\text{power consumption}\,(\mathrm{J/s})}. \tag{9.1}$$

This can be seen as a benefit–cost ratio, where the benefit (data rate) is compared with the associated cost (power consumption). Hence, it is a measure of the network's

[1]Department of Electrical and Computer Engineering, New York University, New York, USA
[2]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
[3]Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden

bit-delivery EE in the same way that the data rate over a given bandwidth measures the bit-delivery SE.

Being able to compute an upper bound on the theoretically achievable EE is very important since it provides a benchmark against which we can compare the goodness of a particular communication technology. It can also be used by the ITU to set concrete EE targets in the future. To this end, there are five key properties to bear in mind:

1. The bandwidth is licensed and allocated by international regulatory bodies [3];
2. The mobility is determined by the application under consideration;
3. The SE is upper bounded by the Shannon capacity [4];
4. The latency is fundamentally limited by the speed of light;
5. The fundamental limit on the EE is not known yet.

The aim of this chapter is to take a closer look at the fundamental limits of EE in massive MIMO, which is one of the key physical-layer technologies in 5G cellular networks and, more generally, to meet the ever-growing demand for wireless broadband connectivity.

9.1.1 Organization

In Section 9.2, we recall the fundamental characteristics that made massive MIMO the key physical-layer technology for 5G networks. The most important features are exemplified using a tractable network model, which is then used to compute a lower bounds on the achievable SE for the line-of-sight (LoS) and non-line-of-sight (NLoS) propagation environments and different linear combining schemes. Section 9.3 introduces a mathematical framework for analyzing the EE of a 5G network. A power consumption model capturing the key characteristics of a multiple-antenna technology is progressively introduced, while paying attention to exemplify how each network parameter impacts the overall network EE. The impact of base station (BS) cooperation is investigated at the beginning of Section 9.4, where a Network MIMO system is studied. Particularly, the channel capacity is first computed and then used to upper bound the achievable SE of a massive MIMO network. Numerical results are used throughout this chapter to both corroborate the analytical results and provide a support tool for the reader. Finally, in the second part of Section 9.4, the interested reader may find a survey of recent research activities wherein more complex system models are used and the implications of the initial simplifying assumptions on the EE analysis are highlighted.

9.1.2 Notation

Upper (lower) bold face letters are used for matrices (column vectors). $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{0}$ is the zero vector. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ are the transpose and conjugate transpose operators, respectively. We use $\mathrm{tr}(\cdot)$ to denote the matrix trace operator and $\|\cdot\|$ for the Euclidean norm vector operator. $\mathbb{E}_{\mathbf{n}}\{\cdot\}$ denotes the expectation operator with respect to the random vector $\mathbf{n}$, whereas $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{\mathbf{n}})$ is the shorthand for a circularly symmetric Gaussian distribution with covariance matrix $\mathbf{R}_{\mathbf{n}}$. We use $\mathbb{R}^n$ and $\mathbb{C}^n$ to denote the n-dimensional real- and complex-valued vector spaces.

9.2 Massive MIMO

9.2.1 What is massive MIMO?

Massive MIMO refers to a multicellular technology where the BSs are equipped with a very large number M of antennas to serve *simultaneously*, i.e., on the same time–frequency coherence block,* a multitude K of single-antenna user equipments (UEs) by spatial multiplexing [5]. Compared to classical multiuser MIMO (MU-MIMO) [6–9], the operating regime of massive MIMO is $M \gg K$ and imperfect channel state information (CSI) is available at the BSs [10]. The use of large-antenna arrays at the BSs not only increases the capability of collecting more power—which is known to bring only a logarithmic improvement of the SE—but allows us also to serve multiple UEs simultaneously—with a linear increase of the SE. The former gain is called *array gain*, while the latter is known as *multiplexing gain*. Both can be exploited to design communications systems that achieve higher data rate and, at the same time, simultaneously provide good service to multiple UEs [10]. Another key difference with MU-MIMO lies in the concept of *scalability*. It is well known that an MU-MIMO system approaches the Shannon channel capacity by recurring to nonlinear processing such as successive interference cancellation in the uplink (UL) [11] and dirty-paper coding in the downlink (DL) [12]. Despite optimal, the high complexity of these techniques makes them unscalable, thus limiting their applicability in networks that are expected to serve an exponentially-increasing number of UEs [13]. On the contrary, due to the operating regime $M \gg K$, massive MIMO needs only linear processing schemes with low complexity such as the zero-forcing (ZF) combining schemes to achieve nearly optimal performance. This makes massive MIMO a *scalable* technology and represents the main reason why it has quickly made its way into the 5G standard [14].

The reason why the operating regime $M \gg K$ is so promising resides in the physical properties of the wireless propagation channel. In particular, as the BS antenna array aperture grows large with M (e.g., with fixed half-wavelength antenna spacing) two fundamental phenomena happen, which are known as *favorable propagation* and *channel hardening* [15]. Favorable propagation implies that any arbitrary pair of channel directions between the BS and two different UEs becomes *asymptotically orthogonal* as $M \to \infty$ [16]. In other words, the angular resolution of the BSs becomes infinitely good, which makes it possible to separate the UEs spatially without incurring in any residual interference. This property of the wireless channel has also been confirmed by practical measurements in different scenarios [17–19]. The channel hardening arises because as M increases the statistical variations of a fading channel's power becomes increasingly concentrated around its mean value; that is, the channel behaves almost deterministically. This property is very beneficial for designing power control algorithms since they do not depend on the particular channel fading realization but rather on its statistical mean value.

*A coherence block of a wireless channel consists of a number of subcarriers and time samples over which the channel response can be considered as constant and flat-fading. Typically, it is given by the product between the coherence bandwidth and coherence time.

Notice that channel hardening requires co-located antennas [20]. Since practical BS antenna arrays have a spatially-constrained topology, the maximum number of antennas is thus practically limited. A possible alternative is to distribute the antennas over a very large geographical area by implementing a distributed massive MIMO system, also known as cell-free massive MIMO [21,22]. A more radical approach is to deploy a massive (possibly uncountably infinite) number of antennas in a compact space. A system of this sort is called Holographic MIMO since it can be regarded as the ultimate form of a spatially-constrained MIMO array [23,24].

9.2.2 A simple network model

Consider the UL of the two-cell network as illustrated in Figure 9.1. Each cell has a BS equipped with M antennas, which receives signals simultaneously from K single-antenna UEs. We assume that perfect CSI is available at the BSs. The impact of channel estimation will be discussed in Section 9.4.2. The transmission occurs over a bandwidth of B Hz. This is a tractable model due to the small number of system parameters and it is sufficient to obtain fundamental insights into the design and analysis of energy-efficient massive MIMO networks, without resorting to any cumbersome mathematical analysis.

We call $\mathbf{h}_{0i}^{0} \in \mathbb{C}^{M}$ the channel vector from UE i in cell 0 to BS 0, while $\mathbf{h}_{1i}^{0} \in \mathbb{C}^{M}$ denotes the channel vector of the interfering UE i in cell 1 to BS 0. We consider both LoS and NLoS propagation channels. Practical channels can contain a mix of a deterministic LoS component and a random NLoS component, but, by studying the differences between the two extreme cases, we can predict what will happen in the mixed cases as well.

In the LoS case, we consider a uniform linear array (ULA) with antenna spacing d (normalized to the wavelength λ). This leads to the following deterministic channel response:

$$\mathbf{h}_{\ell i}^{0} = \sqrt{\beta_{\ell i}^{0}} \left[1, e^{i2\pi d \sin(\phi_{0\ell})}, \ldots, e^{i2\pi(M-1)d \sin(\phi_{0\ell})}\right]^{\mathrm{T}}, \tag{9.2}$$

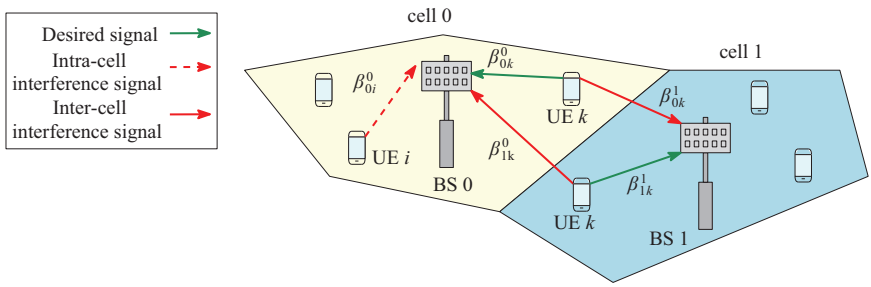


Figure 9.1 Illustration of a two-cell massive MIMO network. The UL transmission of the desired UE k in cell 0 creates an intercell interference to the reception of incoming signals at the BS in cell 1 and vice-versa. Similarly, an arbitrary UE i in cell 0 creates an intracell interference at the BS in the same cell 0

where $\beta_{\ell i}^0 \geq 0$ is the large-scale fading attenuation (i.e., the channel gain) between UE i in cell 0 and BS 0, and $\phi_{0i} \in [0, 2\pi)$ is the azimuth angle to the UE and the boresight of the BS antenna array. In the NLoS case, we assume that the channel vector is generated from an uncorrelated Rayleigh-fading distribution:

$$\mathbf{h}_{\ell i}^0 \sim \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_M, \sqrt{\beta_{\ell i}^0} \mathbf{I}_M \right). \tag{9.3}$$

Its validity in the presence of rich scattering has been demonstrated by actual experiments [25]. For an ULA, this statistical model is valid whenever we consider an isotropic propagation scenario, i.e., waves coming from all the directions and carrying equal power [24]. The performance of a cellular network is typically expressed in terms of the relative signal strength of the interference as compared to the desired signal. We assume to have equal *average* channel gains intracell (i.e., $\beta_{0,i}^0 = \beta_{1,i}^1 = \beta_0^0$) and intercell (i.e., $\beta_{0,i}^1 = \beta_{1,i}^0 = \beta_0^1$) where the average is to be considered among all the UEs–BS distances within a cell. As a consequence, every UE i in cell ℓ transmits with a power $p_{\ell i} = p$ for $\ell = 0, 1$ and $i = 1, \ldots, K$. We denote with $0 \leq \bar{\beta} \leq 1$ the ratio between inter- and intracell channel gains as

$$\bar{\beta} = \frac{\beta_0^1}{\beta_0^0} = \frac{\beta_1^0}{\beta_0^0} = \frac{\beta_1^0}{\beta_1^1} = \frac{\beta_0^1}{\beta_1^1} \tag{9.4}$$

This is typically expressed in logarithmic scale; reasonable values are $-30 \leq \bar{\beta} \leq 30$ dB. The earlier-mentioned model is an instance of the Wyner model [26]. It has been used extensively to study the fundamental information-theoretic properties of cellular networks since it embeds the basic properties of cellular communications while being realistic enough to infer general qualitative conclusions.

9.2.3 Spectral efficiency

We consider the cell 0 only since it is representative of the entire cellular network under the Wyner model settings introduced above. The received signal $\mathbf{y}_0 \in \mathbb{C}^M$ at BS in cell 0 is given by

$$\mathbf{y}_0 = \sum_{i=1}^{K} \mathbf{h}_{0i}^0 s_{0i} + \sum_{i=1}^{K} \mathbf{h}_{1i}^0 s_{1i} + \mathbf{n}_0, \tag{9.5}$$

where $s_{\ell i} \sim \mathcal{N}_{\mathbb{C}}(0, p)$ with $\ell = 0, 1$ and $i = 1, \ldots, K$ is the UL signal transmitted by UE i in cell ℓ, and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, BN_0\mathbf{I}_M)$ is the thermal noise. A typical value for the noise power spectral density is $N_0 = -174$ dBm/Hz. The received signal is multiplied by the receive combining vector $\mathbf{v}_{0k} \in \mathbb{C}^M$ to obtain:

$$y_{0k} = \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{y}_0 = \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{h}_{0k}^0 s_{0k} + \sum_{i=1, i \neq k}^{K} \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{h}_{0i}^0 s_{0i} + \sum_{i=1}^{K} \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{h}_{1i}^0 s_{1i} + n_{0k} \tag{9.6}$$

where $n_{0k} = \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{n}_0 \in \mathbb{C}$. The channel capacity for multiple access channels such as (9.5) is defined as a region since there are K different communication links simultaneously active. In many cases, a single performance metric is preferable. Among the

many possibilities, the sum capacity is a good choice. However, finding a closed-form expression of the sum capacity is generally hard [27], particularly for fading channels. An achievable sum SE is often used that lower bounds the sum capacity. The sum SE naturally depends on the strength of the received desired signal, represented by the average signal-to-noise ratio (SNR). Using the Wyner model described earlier, the average SNR of an UE in cell $\ell = 0, 1$ is

$$\mathrm{SNR} = \beta_0^0 \frac{p}{BN_0}, \tag{9.7}$$

where p denotes the UE's transmit power and BN_0 is the noise power. The following result is thus obtained [15, Corollary 1.3].

Theorem 9.1 (achievable sum SE). *An achievable UL sum SE (bit/s/Hz/cell) in cell 0 is*

$$\mathrm{SE} = \sum_{k=1}^{K} \mathbb{E}\left\{ \log_2\left(1 + \frac{|\mathbf{v}_{0k}^{\mathrm{H}} \mathbf{h}_{00k}|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} |(\mathbf{v}_{0k})^{\mathrm{H}} \mathbf{h}_{00i}|^2 + \sum_{i=1}^{K} |(\mathbf{v}_{0k})^{\mathrm{H}} \mathbf{h}_{01i}|^2 + \frac{1}{\mathrm{SNR}} \|\mathbf{v}_{0k}\|^2} \right) \right\}, \tag{9.8}$$

where the expectation is taken with respect to the channel realizations.

A popular design goal for the combiner vector $\mathbf{v}_{0k}$ consists in maximizing the received power $|\mathbf{v}_{0k}^{\mathrm{H}} \mathbf{h}_{0k}^0|^2 / \|\mathbf{v}_{0k}\|^2$ of the desired UE k in cell 0 while neglecting the interference. This leads to the MR combining scheme:

$$\mathbf{v}_{0k} = \mathbf{h}_{0k}^0. \tag{9.9}$$

Another option is to exploit the knowledge of all channel vectors $\{\mathbf{h}_{\ell i}^0\}$ between every UE i in cell $\ell = 0, 1$ and BS in cell 0 so that to minimize the mean-squared-error (MSE) $\mathbb{E}_{s_{0k}}\{s_{0k} - \mathbf{v}_{0k}^{\mathrm{H}} \mathbf{y}_0 | \{\mathbf{h}_{\ell i}^0\}\}$ between the desired signal s_{0k} and the received combined signal $\mathbf{v}_{0k}^{\mathrm{H}} \mathbf{y}_0$, given the channel realizations. This yields the so-called multi-cell minimum mean-squared-error (M-MMSE) [15,28]:

$$\mathbf{v}_{0k} = \left(\sum_{i=1}^{K} \mathbf{h}_{01}^0 (\mathbf{h}_{0k}^0)^{\mathrm{H}} + \sum_{i=1}^{K} \mathbf{h}_{1k}^0 (\mathbf{h}_{1k}^0)^{\mathrm{H}} + \frac{1}{\mathrm{SNR}} \right)^{-1} \mathbf{h}_{0k}^0. \tag{9.10}$$

It is worth observing that MR combining was primarily adopted in the early works on massive MIMO. The M-MMSE was applied in [28,29], where the authors proved that this method is robust to the so-called *pilot contamination* issue [10,30], which was believed to impose a fundamental limitation on the sum SE in massive MIMO. In other words, [28,29] showed that the SE increases theoretically unboundedly, even under pilot contamination, in the regime where the number of antennas goes to infinity,

i.e., $M \to \infty$, while both the numbers of UEs K and cells are fixed. We refer the interested reader to [29] for a detailed treatment of the subject.

When MR combining is used, analytical closed-form expressions can be derived from Theorem 9.1 [15, Lemma 1.7]. The M-MMSE case is treated numerically.

Lemma 9.1 (LoS propagation). *When MR is used at the BS in cell 0, an achievable UL sum SE (bit/s/Hz/cell) in the LoS case is*

$$\mathrm{SE} = \sum_{k=1}^{K} \log_2 \left(1 + \frac{M}{\sum_{\substack{i=1 \\ i \neq k}}^{K} g(\phi_{00k}, \phi_{00i}) + \bar{\beta} \sum_{i=1}^{K} g(\phi_{00k}, \phi_{01i}) + \frac{1}{\mathrm{SNR}}} \right), \tag{9.11}$$

where $\mathrm{SNR} = \frac{p\beta_0^0}{N_0 B}$, ϕ_{00i} *with* $i = 1, \ldots, K$ *is the azimuth angle from the BS 0 to the UE i in cell 0, and the function* $g(\phi, \xi) : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}_+$ *defined as*

$$g(\phi, \xi) = \begin{cases} \frac{\sin^2(\pi d M(\sin(\phi) - \sin(\xi)))}{M \sin^2(\pi d(\sin(\phi) - \sin(\xi)))}, & \text{if } \sin(\phi) \neq \sin(\xi) \\ M, & \text{if } \sin(\phi) = \sin(\xi) \end{cases} \tag{9.12}$$

represents the ratio between the interference power and the desired signal power.

Lemma 9.2 (NLoS propagation). *When MR is used at the BS in cell 0, an achievable UL sum SE (bit/s/Hz/cell) in the NLoS case is*

$$\mathrm{SE} = \sum_{k=1}^{K} \mathbb{E} \left\{ \log_2 \left(1 + \frac{\|\mathbf{h}_{00k}\|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \frac{|(\mathbf{h}_{00k})^{\mathrm{H}} \mathbf{h}_{00i}|^2}{\|\mathbf{h}_{00k}\|^2} + \sum_{i=1}^{K} \frac{|(\mathbf{h}_{00k})^{\mathrm{H}} \mathbf{h}_{01i}|^2}{\|\mathbf{h}_{00k}\|^2} + \frac{1}{\mathrm{SNR}}} \right) \right\} \tag{9.13}$$

$$\geq K \log_2 \left(1 + \frac{M - 1}{(K - 1) + K\bar{\beta} + \frac{1}{\mathrm{SNR}}} \right). \tag{9.14}$$

To quantitatively evaluate the sum SE, Figure 9.2 shows it as a function of the SNR in dB for $M = 50$, $K = 10$, $\bar{\beta} = -15$ dB, $N_0 = -174$ dBm/Hz. Both the LoS and NLoS propagation scenarios are considered. A ULA with half-wavelength antenna spacing is considered in both cases. The SE with MR is evaluated using the analytical expressions in Lemmas 9.1 and 9.2, whereas it is evaluated numerically with M-MMSE using the SE expression in Theorem 9.1. As seen, M-MMSE achieves the highest SE in both scenarios. In the noise-limited regime (i.e., small SNR), the gap with MR is relatively small. On the other hand, it is substantial in the interference-limited regime for which we have a higher SNR. This is due to the capability of M-MMSE to reject (intra- and inter-cell) interference. A performance gap is observed between LoS and NLoS scenarios, especially in the interference-limited regime.

The results in Figure 9.2 validate the accuracy of the lower bound in (9.14), which is thus used in the remainder of this chapter. To obtain an analytically tractable expression for the SE in the LoS scenario, we assume that the (normalized) total interference power in (9.11) does not depend on the particular UEs pair (k, i) i.e.,

$$g(\phi_{00k}, \phi_{00i}) \leq g_{00} \qquad \text{for} \quad k, i = 1, \ldots, K, \tag{9.15}$$

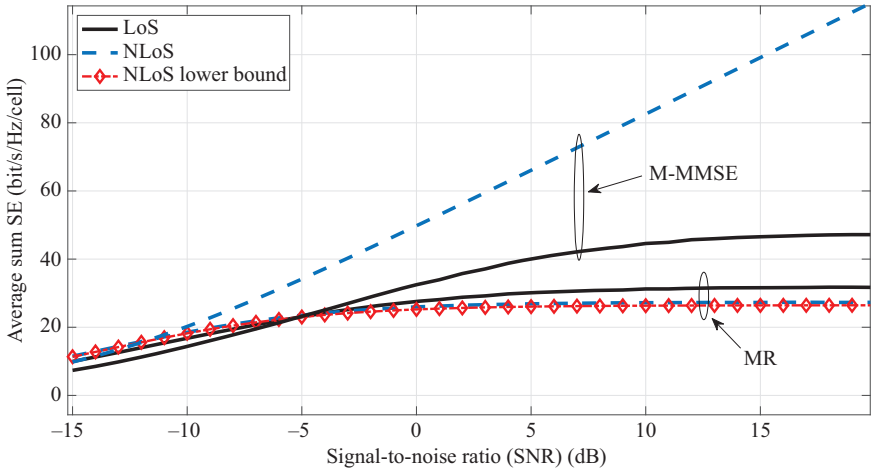


Figure 9.2 Average sum SE as a function of the SNR in dB as defined in (9.7) when $M = 50$, $K = 10$, $\bar{\beta} = -15$ dB, $N_0 = -174$ dBm/Hz. Both MR and M-MMSE are considered with either LoS or NLoS channels

where g_{00} may be obtained by considering, e.g., the maximum interference among all UEs. Also, when different UEs have the same azimuth angle with respect to the serving BS from (9.12) we have that $g(\phi_{00k}, \phi_{00k}) = M$, i.e., the normalized interference power increases proportionally to the number of antennas M. This is because the BS is coherently combining the interference signal since it comes from the same direction of the desired signal. Typically, this case never happens in practice or with very low probability due to the use of scheduling algorithms and the fact that UEs may have different elevation angles. Hence, it is not considered in our analysis. From Lemmas 9.1 and 9.2, we observe that the SE in the LoS and NLoS cases takes this general form:

$$\text{SE} \geq K \log_2 \left(1 + \frac{a}{b + \frac{1}{\text{SNR}}} \right), \tag{9.16}$$

where the SNR is given in (9.7):

$$a = \begin{cases} M & \text{LoS} \\ M - 1 & \text{NLoS} \end{cases} \tag{9.17}$$

and b accounts for the interference power

$$b = \begin{cases} (K-1)(1+\bar{\beta})g_{00} & \text{LoS} \\ (K-1) + K\bar{\beta} & \text{NLoS}, \end{cases} \tag{9.18}$$

which is obtained by using (9.15) into (9.11) for the LoS scenario.

9.3 Energy efficiency analysis

The EE of a wireless communications system is defined in (9.1) and measured in units of bit/J. The UL data rate is computed as the product between the UL sum SE and system bandwidth B:

$$\text{data rate} = B\,\text{SE} \ \ (\text{bit/s}). \tag{9.19}$$

Unlike the SE that can be computed numerically and/or analytically, the network power consumption depends on several factors among which the circuit hardware (e.g., analog-to-digital and digital-to-analog conversions, local-oscillators, and power amplifiers), digital signal processing (e.g., channel estimation, combining, and precoding), and load-dependent contributions (e.g., coding, decoding, and back-hauling) [31,32]. All these factors contribute to the circuit power (CP) consumption, which, together with the total transmit power Kp, determines the network power consumption, i.e.:

$$\text{Power consumption} = Kp + \text{CP} \ \ (\text{W}). \tag{9.20}$$

The expression of the CP consumption is often driven by analytical tractability rather than practical considerations. Next, we discuss the implications of using an over-simplistic model and show how this may bring to misleading conclusions for the fundamental limits of EE in multiple antenna communication systems. To overcome this, more practical CP consumption models are progressively introduced and studied.

9.3.1 Zero circuit power

We start by studying the case where the transmit power p is the only contribution to the power consumption. This amounts to setting $\text{CP} = 0$ in (9.20). This case never occurs in practice but can be thought of as a limiting case that resembles an extremely efficient network capable of performing a huge amount of logic operations while using a negligible computing power [33]. Under these circumstances, the EE formula in (9.1) becomes

$$\text{EE} = B\frac{\text{SE}}{Kp}, \tag{9.21}$$

where the SE is obtained in (9.16). Hence, (9.21) depends on M, K, and the ratio p/B. Particularly, (9.21) is a monotonically nondecreasing function with respect to $B/p \geq 0$. Hence, it is maximized as $p/B \to 0$, which can be achieved by taking the transmit power $p \to 0$, or by taking the bandwidth $B \to \infty$, or a combination thereof. This consideration allows us to focus on an MR combiner only since we do not expect high differences in terms of SE with respect to the M-MMSE combiner (see Figure 9.2). The limit can be easily computed [33]:

$$\lim_{p/B \to 0} \frac{KB}{Kp} \log_2\left(1 + \frac{a}{b + \frac{1}{\text{SNR}}}\right) = \frac{\log_2(e)}{N_0}\, a\beta_0^0. \tag{9.22}$$

We recognize (9.22) as the reciprocal of the minimum energy-per-bit for AWGN channels [34], with the only difference that a deterministic channel gain $a\beta_0^0$ is included.

From (9.17), it thus follows that we can arbitrarily increase (9.22) by letting $M \to \infty$, since a is proportional to M as defined in (9.17). Notice also that (9.22) does not depend on the interference power level b. This is because by taking the limit $p/B \to 0$, every UE transmits with a very low power per unit of bandwidth such that SNR $\to 0$ and the network operates in the noise-limited regime.

To quantify the EE limit in practical scenarios, Figure 9.3 illustrates (9.22) as a function of β_0^0 in dB for $M = \{10, 50, 100\}$ and $K = 10$. As seen from this figure, the EE upper limit is a monotonically non-decreasing function of β_0^0 and M and achieves values in the range of 10^4–10^5 Gbit/J for $\beta_0^0 = -80$ dB. In free-space propagation at 3 GHz with lossless isotropic antennas, $\beta_0^0 = -80$ dB corresponds to 80 m distance; distances are expected to be much shorter in future networks. The above results ignore the fact that the CP increases with M in practice. This shortcoming is addressed in what follows by considering a progressively more realistic power consumption model [31,32].

The analysis above shows that it does not matter if $p \to 0$ or $B \to \infty$ in terms of network EE. However, if we look at the achievable network data rates, this makes a huge difference since for the two cases we have that:

$$\text{data rate} \to \begin{cases} 0 & p \to 0 \\ \frac{Kp\beta_0^0}{N_0} a \log_2(e) & B \to \infty. \end{cases} \tag{9.23}$$

For example, we get 0 bit/s if $p \to 0$ or roughly 200 Tbit/s if $B \to \infty$ with $p = 20$ dBm, $\beta_0^0 = -80$ dB, $K = 10$, $M = 50$, and $N_0 = -174$ dBm/Hz. A communication system with zero data rate is practically worthless, even if it is energy efficient.

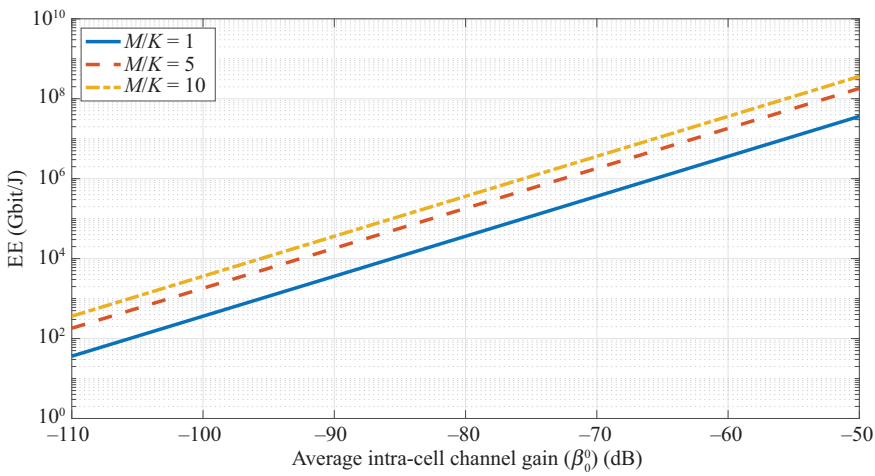


Figure 9.3 EE limit in (9.22) as a function of the average intracell channel gain β_0^0 in dB when $K = 10$, $M = 10$, 50, and 100. The circuit power is not accounted for

The reason for this weird result comes from the considered power consumption model, which accounts only for the transmit power. This will be generalized below.

Figure 9.4 shows how the EE approaches its limit as $B \to \infty$ when $p = 20$ dBm, $K = 10$, $M = 50$, and $N_0 = -174$ dBm/Hz. Both MR and M-MMSE are employed for combining. Different values of β_0^0 are considered, and these determine how quickly we approach the EE limit. For the edge-cell case of $\beta_0^0 = -110$ dB, the limit is reached at $B = 10$ GHz at 100 Gbit/J, while we need 100× more bandwidth every time β_0^0 is increased by 20 dB. For a UE close to its serving BS, i.e., $\beta_0^0 = -50$ dB, the EE upper limit is reached over $B = 1$ THz and is around 10^5 Gbit/J. M-MMSE combining is more energy efficient than MR for any value of B. This is because for the same transmit power, it achieves higher data rates. However, we notice that M-MMSE has higher complexity than MR. This inevitably increases the power consumption. This will be taken into account later on.

9.3.2 *Constant but nonzero circuit power*

Assume now that fixed CP $= \mu$ with $\mu \geq 0$ is included in the model (9.20). When communicating over long distances, it is common to have $Kp + \mu \approx Kp$, but in future small cells, it is possible that $\mu > Kp$. The EE in (9.1) becomes

$$\text{EE} = \frac{B\,\text{SE}}{Kp + \mu}, \tag{9.24}$$

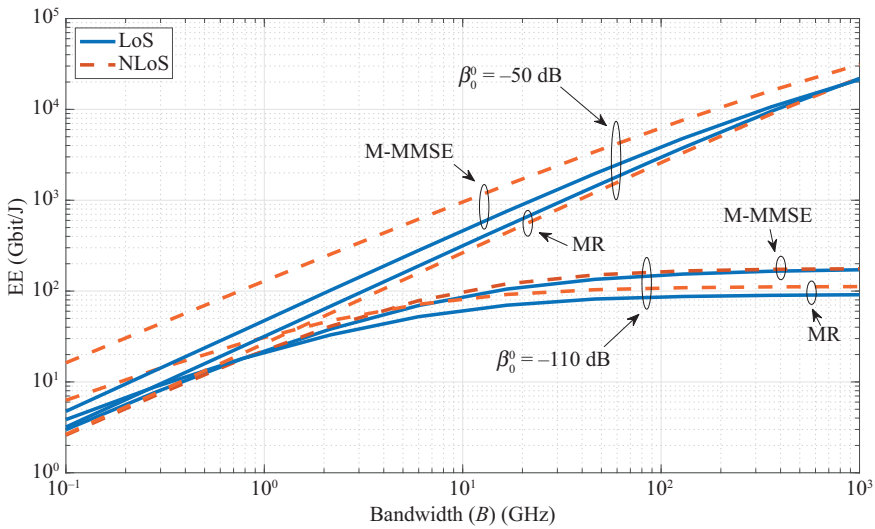


Figure 9.4 EE as a function of the system bandwidth B for different intracell channel gains $\beta_0^0 \in \{-110, -50\}$ dB. Both MR and M-MMSE are considered. We fixed $p = 20$ dBm and $M = 50$. The other parameters are in Table 9.1

which depends on B, the ratio p/B, K, and M. Unlike (9.21), its upper limit cannot be achieved by taking the limit $p/B \to 0$. In fact, the EE in (9.24) is a monotonic non-decreasing function with respect to B only, and it is thus maximized as $B \to \infty$. We can exploit this information by studying the noise-limited regime, which is well described by the MR combiner (see Figure 9.2). Similarly to (9.22), we obtain

$$\lim_{B\to\infty} \frac{BK}{Kp+\mu} \log_2\left(1+\frac{a}{b+\frac{1}{\mathrm{SNR}}}\right) = \frac{\log_2(e)a\beta_0^0}{N_0} \frac{1}{1+\frac{\mu}{Kp}}. \tag{9.25}$$

The resulting EE limit is in the same form of (9.22) with the only difference that a scaling factor $0 < \frac{1}{1+\mu/Kp} \leq 1$ is included. By setting $\mu = 0$, we obtain (9.22). The same result can be achieved by observing that (9.25) is monotonically increasing with p, and it can thus be maximized by letting $p \to \infty$. This yields

$$\lim_{p\to\infty} \frac{\log_2(e)a\beta_0^0}{N_0} \frac{1}{1+\frac{\mu}{Kp}} = \frac{\log_2(e)a\beta_0^0}{N_0}. \tag{9.26}$$

The same result can be achieved by letting P and B going jointly to infinity, but with B having a higher convergence speed than P. Notice that, by taking the limit $p \to \infty$, we implicitly assume that μ becomes negligible. This is the reason why the same EE limit is obtained if $\mu \to 0$ or $p \to \infty$.

If $p < \infty$ and $\mu > 0$, a loss in the EE upper limit is experienced compared to (9.22). This loss increases as the ratio μ/p grows large. To quantify this, let us consider a practical scenario with $p = 20$ dBm, $K = 10$ (i.e., $Kp = 1$), and $\mu = \{5, 10\}$ W [15]. If $\mu = 5$ W with $\mu/p = 50$, the EE limit is 83% lower. A 91% smaller EE is achieved when $\mu = 10$ W and thus $\mu/p = 100$.

So far, we have not taken into account the CP increases with M in practice. This is addressed in what follows.

9.3.3 Impact of BS antennas

We now consider the CP model:

$$\mathrm{CP} = \mu + \eta M, \tag{9.27}$$

with $\eta \geq 0$ accounting for the power consumption of multiantenna processing schemes and transceiver architectures [15, Chapter 5]. This choice is motivated by the fact that (i) a BS antenna is connected to a radio frequency chain, including hardware components (e.g., digital-to-analog converter, local oscillator, power amplifier); (ii) the total number of complex-valued samples to be processed at each BS (per symbol time) is proportional to M; and (iii) the complexity of digital signal processing (e.g., beamforming, channel estimation) depends on M as well. We notice that the power consumption due to hardware components typically scales linearly with M, whereas digital signal processing may lead to higher-order dependencies [15]. For simplicity,

we neglect this fact and consider only the linear model in (9.27). Hence, the EE formula takes the following form:

$$\mathrm{EE} = \frac{B\,\mathrm{SE}}{Kp + \mu + \xi M}. \tag{9.28}$$

The above expression is still monotonically increasing with respect to B and its limit as $B \to \infty$ for the representative MR case yields:

$$\lim_{B\to\infty} \frac{BK}{Kp + \mu + \xi M} \log_2\left(1 + \frac{a}{b + \frac{1}{\mathrm{SNR}}}\right) = \frac{\log_2(e) a \beta_0^0}{N_0} \frac{1}{1 + \frac{\mu}{Kp} + \frac{\xi M}{Kp}}. \tag{9.29}$$

By setting $\mu = \xi = 0$, we clearly obtain (9.22). As before, the same result can be achieved by letting $p \to \infty$. When $p < \infty$ and $\mu, \xi > 0$, we have a non-negligible loss compared to (9.22). This loss depends jointly on p, K, and M/K. With $p = 20$ dBm, $K = 10$, $\mu = 10$ W, and $\xi = 0.5$ W [15], a loss of 94% and 98% in the EE limit is achieved for $M = 10$ and 100, respectively.

In massive MIMO, a regime of interest is when $M \to \infty$ and K is kept fixed. In this case, we obtain:

$$\lim_{M\to\infty} \frac{BK}{Kp + \mu + \xi M} \log_2\left(1 + \frac{a}{b + \frac{1}{\mathrm{SNR}}}\right) = \frac{\log_2(e) \beta_0^0}{N_0} \frac{Kp}{\xi}, \tag{9.30}$$

which does not depend on M. Unlike (9.25), where we could unboundedly increase the EE limit by letting $M \to \infty$, the EE limit in (9.30) is quantified as approximately 7 Tbit/J when $p = 20$ dBm, $K = 10$, $\beta_0^0 = -80$ dB, $N_0 = -174$ dBm/Hz, and $\xi = 0.5$ W.

9.3.4 Varying circuit power

So far, we have modeled the CP consumption as independent of B. This implies that we expect no changes in the network power consumption as the system sampling rate (which is proportional to B) increases unboundedly. This is obviously not true in practice since the CP highly depends on the number of samples to be processed per second, i.e., the system sampling rate. Also, the amount of consumed power strongly depends on the daily network load variations due to coding, decoding, and backhauling of signals. This is also proportional to the data rate and, in turn, to the bandwidth B. Hence, a more realistic model is

$$\mathrm{CP} = \mu + \eta BM + \nu B + \gamma B\,\mathrm{SE}, \tag{9.31}$$

where $\eta, \nu, \gamma \geq 0$ are hardware characterization constants that are specified in Table 9.1, along with some example numbers. The exact numbers are implementation specific, while the structure of the power consumption in (9.31) is general. To obtain a fundamental limit of the EE, these values may be obtained by recurring to the futuristic limit on computational complexity, i.e., the so-called Landauer limit of approximately 10^{18} flops/J [35,36]. Using (9.31), the EE takes the following form:

$$\mathrm{EE} = \frac{B\,\mathrm{SE}}{Kp + \mu + \eta BM + \nu B + \gamma B\,\mathrm{SE}}, \tag{9.32}$$

Table 9.1 Hardware characterization constants

Parameter	Symbol	Value (J)
Radio frequency and digital signal processing	η	10^{-14}
Baseband processing	ν	10^{-14}
Load-dependent power consumption	γ	10^{-15}

Note: y is measured in units of J/bit, while the other two are measured in units of J.

which is a non-linear fractional program [37] of the network parameters (p, B, M, K). Its global optimal is in general hardly available in closed form. To gain some insights into the fundamental limits of EE, we assume the following:

1. The network operates in the noise-limited regime;
2. The fixed CP is neglected, i.e., $\mu = 0$.

To quantitatively find a condition under which the first assumption holds, one can use (9.16). The noise-limited regime requires that:

$$\frac{1}{\text{SNR}} \geq \begin{cases} (K-1)(1+\bar{\beta})g_{00} & \text{LoS} \\ (K-1)+K\bar{\beta} & \text{NLoS} \end{cases} \tag{9.33}$$

from which, by using (9.7), one obtain:

$$\frac{p}{B} \leq \begin{cases} \frac{N_0}{\beta_0^0}\frac{1}{(K-1)(1+\bar{\beta})g_{00}} = \tau_{\text{LoS}} & \text{LoS} \\ \frac{N_0}{\beta_0^0}\frac{1}{(K-1)+K\bar{\beta}} = \tau_{\text{NLoS}} & \text{NLoS.} \end{cases} \tag{9.34}$$

Interestingly, the ratio p/B under which the network operates in the noise-limited regime is independent of M. This is because both the receive noise power and the interference power do not depend on M. While this is always the case for the receive noise power, the interference power may scale linearly with M. In practice, this occurs when multiple UEs share the same pilot sequence for channel estimation (e.g., see pilot contamination effect in [10]). It happens also when there are closely spaced UEs as it follows from (9.18) for MR with LoS propagation. The case $p/B \geq \tau_{\text{LoS/NLoS}}$ leads to the interference-limited region, which is treated later on.

Figure 9.5 illustrates the pair (p, B) for which the interference power equals the noise power for LoS and NLoS scenarios. Different curves are obtained for different values of β_0^0. Each of those curves must be understood in the sense that in the region below the curves the network is in the noise-limited regime, i.e., the noise power is higher than interference. Reasonably, the farther the UEs from the serving BS, the larger is the noise-limited regime for a given B, in the sense of it appearing for higher average transmit powers. For example, every time β_0^0 is increased by 30 dB, we need a 1,000× transmit power per UE. The minimum power required to operate in the noise-limited regime changes as a function of B. Quantitatively speaking, e.g., for the NLoS scenario, when $\beta_0^0 = -80$ dB, $\bar{\beta} = -15$ dB, $K = 10$, and $N_0 = -174$ dBm/Hz, a transmit power of $p \leq 90$ μW is sufficient when $B = 2$ GHz. On the one hand, $p \leq 4$ μW is needed when $B = 100$ MHz.

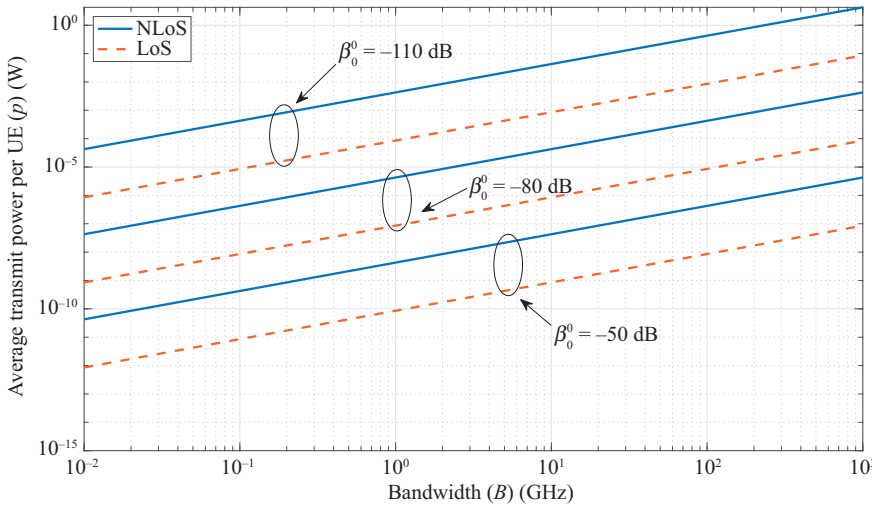


Figure 9.5 Average per-UE transmit power p needed for the noise power to equal the interference power as a function of the bandwidth B (GHz). We plot the analytical formula in (9.35) for different average channel gains β_0^0 in dB for LoS and NLoS. Also, $\bar{\beta}$ and K are listed in Table 9.1

In the noise-limited regime defined earlier and $\mu = 0$, the EE reduces to

$$\mathrm{EE} = \frac{K \log_2\left(1 + a\frac{\beta_0^0}{N_0}\frac{p}{B}\right)}{K\frac{p}{B} + \eta M + \nu + \gamma K \log_2\left(1 + a\frac{\beta_0^0}{N_0}\frac{p}{B}\right)}, \tag{9.35}$$

which is a unimodal function of p/B [32, Lemma 3] and is maximized for

$$\left(\frac{p}{B}\right)^{\star} = \frac{e^{W\left(\frac{1}{e}\left(\frac{a\beta_0^0}{N_0}\frac{(\eta M+\nu)}{K}-1\right)\right)+1} - 1}{\frac{a\beta_0^0}{N_0}}, \tag{9.36}$$

where $W(\cdot)$ denotes the Lambert function [38]. Hence, a CP model that depends on B pushes the optimal p/B away from the origin to a nonzero finite value as derived in (9.36). Interestingly, the optimal p/B does not depend on γ, which accounts for the load-dependent power consumption. This implies that every UE can adjust its own power irrespective of the traffic conditions.

A simplification of (9.36) can be obtained by assuming $M \gg K$ and by using the general fact that $a\beta_0^0/N_0 \gg 1$. Since $e^{W(z)+1} \approx z$ with $z \gg 1$ [38], (9.36) can be approximated as

$$\left(\frac{p}{B}\right)^{\star} \approx \frac{\eta M + \nu}{eK}, \tag{9.37}$$

where we have neglected the -1 term in the numerator of (9.36). This shows that, in the noise-limited regime, at the EE optimum the transmit power $p^{\star}$ scales linearly

with M/K. Quantitatively, by using the same setting of Figure 9.5 with $M = 50$ and $\eta = 10^{-14}$ J as indicated in Table 9.1, we obtain $p^\star = 2$ μW when $B = 100$ MHz and $p^\star = 40$ μW when $B = 2$ GHz. Both are within the required range of power as given by (9.34).

By inserting (9.37) back into (9.35) yields:

$$\mathrm{EE} \approx \frac{\log_2\left(1 + a\frac{\beta_0^0}{N_0}\frac{\eta M+\nu}{eK}\right)}{\frac{M}{K}\eta\left(\frac{1}{e}+1\right) + \frac{\nu}{eK} + \gamma\log_2\left(1 + a\frac{\beta_0^0}{N_0}\frac{\eta M+\nu}{eK}\right)}, \tag{9.38}$$

which is a function of K, M/K, β_0^0/N_0, and the hardware constants ν, η, and γ as specified in Table 9.1. Note that the power constraint in (9.34) maps to a constraint $M/K \leq (e\tau - \nu/K)/\eta$ required to be in the noise-limited regime, which is always satisfied for any practical network parameter values. The EE limit in (9.38) is illustrated in Figure 9.6 for the same setup of Figure 9.3. Similar to the zero CP case, the EE is a monotonically non-decreasing function of β_0^0. However, distinct operating regimes can be identified as the ratio M/K increases. Here, two regions exist: when adding BS antennas per UE might or might not be beneficial in terms of increasing the EE. Particularly, when $\beta_0^0 \in [-110, -80]$ dB (edge cell UEs), there is a gain in increasing

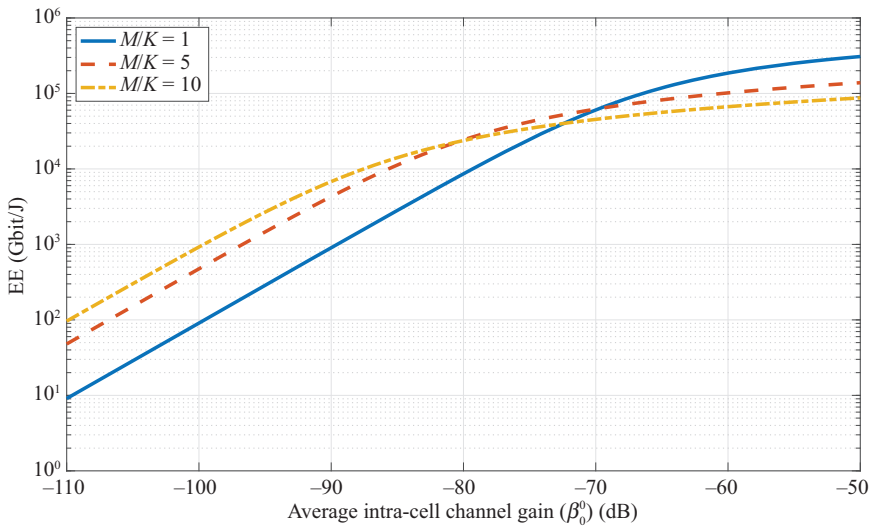


Figure 9.6 EE (Gbit/J) upper limit in (9.38) with varying circuit power $CP = \mu + \eta BM + \nu B + \gamma B$ SE (W). The EE is evaluated for a network operating in the noise-limited regime as a function of the average intracell channel gain β_0^0 in dB for different BS antenna-UE ratios M/K. The optimal p/B in (9.37) is used. The other parameters are in Table 9.1

M/K. On the other hand, a loss is experienced for $\beta_0^0 \in [-70, -50]$ dB (UEs close to the BS). The region $\beta_0^0 \in [-90, -70]$ dB is a transitory stage between the two. Also, while for the zero-CP case, we could push the EE upper limit at $\beta_0^0 = -80$ dB up to 2×10^5 Gbit/J with $M/K = 10$, by considering a more practical model we obtain an EE of at most 2×10^4 Gbit/J, with a loss of an order of magnitude.

9.3.5 *Impact of interference*

Following the discussion on the EE upper limits in the presence of a more practical varying CP model, we now take a closer look at the implications of operating in an interference-limited (noise-free) regime. As mentioned earlier, we neglect the fixed contribution to the CP, i.e., $\mu \approx 0$, while considering p/B to be high enough so that the network operates in an interference-limited regime; (see Figure 9.5). This choice is generally suboptimal since the EE may not be a monotonically increasing function of p/B over its entire domain.

Unlike the noise-limited regime, which exhibits very limited differences to system changes, the EE evaluation in the interference-limited region must be tailored specifically to each network configuration. Similarly, the curve $p(B)$ that separates the two regions may be computed by comparing the interference power and receives noise power appearing in the SE formula of Theorem 9.1. A rule of thumb suggests that $p/B \geq 10\tau$ is enough for the interference to be the dominant effect. In this case, the EE simplifies to

$$\mathrm{EE} \approx \frac{\sum_{k=1}^{K} \mathbb{E}\left\{\log_2\left(1+\frac{|\mathbf{v}_{00k}^{\mathrm{H}}\mathbf{h}_{00k}|^2}{\sum_{\substack{i=1\\ i\neq k}}^{K}|(\mathbf{v}_{00k})^{\mathrm{H}}\mathbf{h}_{00i}|^2+\sum_{i=1}^{K}|(\mathbf{v}_{00k})^{\mathrm{H}}\mathbf{h}_{01i}|^2}\right)\right\}}{\frac{Kp}{B}+\eta M+\nu+\gamma\ \sum_{k=1}^{K}\mathbb{E}\left\{\log_2\left(1+\frac{|\mathbf{v}_{00k}^{\mathrm{H}}\mathbf{h}_{00k}|^2}{\sum_{\substack{i=1\\ i\neq k}}^{K}|(\mathbf{v}_{00k})^{\mathrm{H}}\mathbf{h}_{00i}|^2+\sum_{i=1}^{K}|(\mathbf{v}_{00k})^{\mathrm{H}}\mathbf{h}_{01i}|^2}\right)\right\}}. \tag{9.39}$$

The MR combiner $\mathbf{v}_{00k}$ in (9.9) does not depend on p/B. Also, as p/B grows large, the SNR increases proportionally and the M-MMSE combiner in (9.10) become less sensitive to the level of transmit power. Hence, the above expression becomes a monotonic non-increasing function of p/B, since this appears at the denominator only. Thus, the optimal p/B is obtained as the infimum of the subset under which (9.39) is valid:

$$\left(\frac{p}{B}\right)^{\star} = 10\tau = \frac{N_0}{\beta_0^0}\frac{10}{(K-1)+K\bar{\beta}}, \tag{9.40}$$

where we substituted the value of τ in (9.34). The EE in the interference-limited regime is evaluated numerically by means of Monte Carlo simulations in Figures 9.7 and 9.8 for M-MMSE and MR, respectively. The EE upper limit is obtained by plugging (9.40) into (9.39) and is plotted as a function of $\beta_0^0 \in [-110, -50]$ dB, for $M/K = \{1, 5, 10\}$, $K = 10$, and using the hardware parameters in Table 9.1.

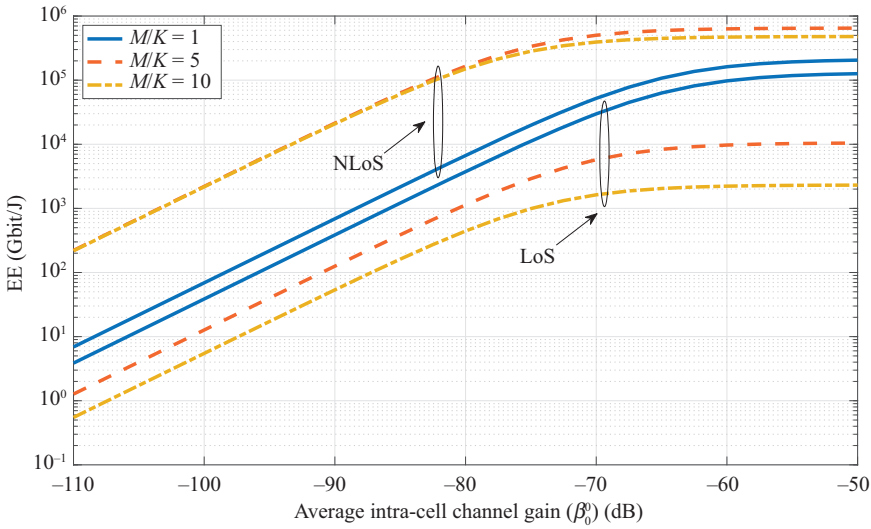


Figure 9.7 *EE (Gbit/J) upper limit in (9.39) with varying circuit power and an M-MMSE combiner. The EE is evaluated for a network operating in the inference-limited regime as a function of the average intracell channel gain β_0^0 in dB for different BS antenna-UE ratios M/K. The optimal p/B in (9.40) is used. The optimal p/B in (9.40) is used. The other parameters are in Table 9.1*

Similarly to Figures 9.3 and 9.6, the EE limit is a monotonically non-decreasing function of β_0^0. M-MMSE achieves higher EE than MR in the case of NLoS propagation, as seen from the substantial SE gap in Figure 9.2, which for $\beta_0^0 = -80$ dB, $K = 10$, and $M = 50$ can be quantified as an EE gain of an order of magnitude. The opposite is true in the LoS case, where MR achieves a smaller but significant EE gain compared to M-MMSE. This gap is due to the fact that the SE gain of M-MMSE is counteracted by the increasing computational complexity required for performing multicell processing, as opposite to single cell processing required by MR combining.

As expected from the SE in Figure 9.2, an EE gap is observed between the NLoS and LoS scenarios with M-MMSE, whereas this gap is limited with MR. The EE gap with M-MMSE becomes larger as M/K increases. For example, with $\beta_0^0 \in -80$ dB this is quantified as a factor of 2 with $M/K = 1$, whereas as a few order of magnitudes with $M/K = 10$. Different behaviors arise for the LoS and NLoS cases as a function of M/K. While, in the LoS propagation, equipping BSs with more and more antennas per UE always reduces the EE, distinct operating regimes could be identified in the NLoS case as a function of M/K (as what we have seen in the noise-limited regime).

For completeness, in Figure 9.9 we show the behavior of the EE as a function of SNR over its entire range of values. As expected, the EE is a unimodal function of the SNR wherein the maximum is achieved at SNR $= -10$ dB. The curve is

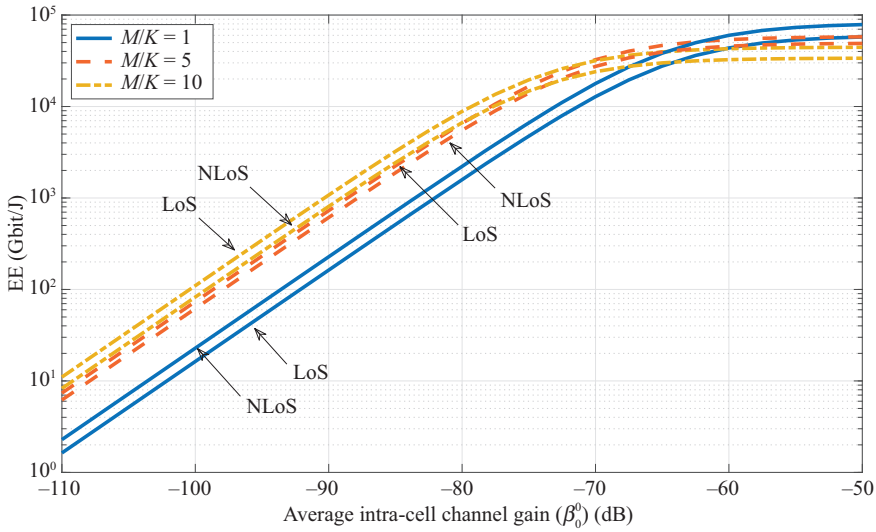


Figure 9.8 *EE (Gbit/J) upper limit in (9.39) with varying circuit power and an MR combiner. The EE is evaluated for a network operating in the inference-limited regime as a function of the average intracell channel gain β_0^0 in dB for different BS antenna-UE ratios M/K. The optimal p/B in (9.40) is used. The other parameters are in Table 9.1*

quite smooth around the maximum EE point that shows to be robust against channel condition β_0^0/N_0 and system changes p/B. We recall that the SE in Figure 9.2 is a monotonic non-decreasing function of the SNR and achieves a floor due to the interference, irrespective of the combiner. Thus, one may want to transmit with as large p/B as possible to achieve the fundamental SE upper limit. The lack of further SE benefits is negatively weighted by the EE metric, which, unlike the SE, includes the power consumption cost needed to run the cellular network. Interestingly, there is an operating region wherein we can jointly increase the SNR (and so the SE) and EE. Quantitatively speaking, when $\beta_0^0 = -80$ dB and $N_0 = -174$ dBm/Hz, this yields optimal average transmit power $p^\star = 4$ μW ($B = 100$ MHz) and $p^\star = 80$ μW ($B = 2$ GHz), which according to (9.34) may be considered as a noise-limited regime.

In summary, the adoption of an M-MMSE combiner at the BS is preferable when the network operates in an interference-limited regime—especially with NLoS propagation scenarios where its advantages in terms of SE and EE are more pronounced—whereas MR is the way to go when considering a noise-limited regime. Despite the increased computational complexity, there exist benefits from the use of M-MMSE rather than MR when the incoming signals reach the BS through multiple paths. This is due to the capability of M-MMSE to suppress the intra- and inter-cell interference coming from several directions.

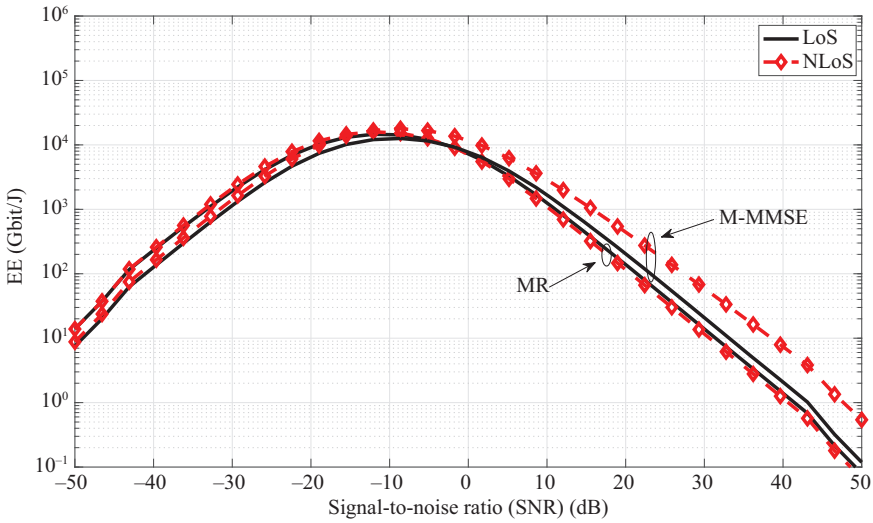


Figure 9.9 *EE (Gbit/J) upper limit in (9.39) with varying circuit power and an M-MMSE combiner. The EE is evaluated for a network operating in the inference-limited regime as a function of the average intracell channel gain β_0^0 in dB for different BS antenna-UE ratios M/K. The optimal p/B in (9.40) is used. There are $K = 10$ active UEs and the relative interference is $\bar{\beta} = -15$ (dB)*

9.3.6 Summary of Section 9.3

- The EE is the number of bits that can be reliably transmitted per unit of energy (bit/J). This number is defined as the ratio between the data rate and associated power consumed for running the cellular network. A common model for the network power consumption accounts for the total transmit power and CP;
- The CP is generally obtained as a sum of different power contributions: digital signal processing (e.g., due to channel estimation and baseband processing), radio frequency (e.g., to run the transceiver hardware components), and load-dependent processing (e.g., due to coding, decoding and backhauling);
- The adoption of a simplified CP consumption model to analyze the EE of a communication system may lead to misleading conclusions. For example, for an ideal zero CP (i.e., considering the transmit power only), the EE upper limit equals the minimum energy-per-bit of an AWGN channel, i.e., $\frac{\log_2(e)\beta_0^0 M}{N_0}$ (bit/J), which grows unboundedly with the number M of antennas;
- Adding a constant CP term to the network power consumption does not change the EE behavior, but rather diminishes the EE upper limit by roughly a factor of $\times 2$;
- The dependence of the CP on M naturally appears due to the additional hardware deployed. This increases both the signal processing and transceiver power consumption. Including M into our CP model makes the EE upper limit bounded

as $M \to \infty$. For $M/K \gg 1$ (i.e., massive MIMO regime), this limit is equal to $\frac{\log_2(e)\beta_0^0}{N_0}\frac{Kp}{\xi}$ (bit/J), which does not depend on M;

- A more realistic CP model must account for the bandwidth B due to the increase in the number of samples processed by the network. This leads to a realistic network power consumption model that is generally hard to treat mathematically. The EE maximization problem becomes a multivariate fractional programing that is a function of several network design parameters;
- To establish general conclusions that do not rely heavily on specific design parameters, we study the EE behavior in two different regimes: noise- and interference-limited regimes. These two regions are found by requiring the ratio p/B to be lower or higher than a certain threshold, respectively;
- In the noise-limited regime, i.e., $p/B < \tau$, when $M/K \gg 1$, the optimal transmit power $p^\star$ scales linearly with the ratio M/K and does not depend on the traffic load conditions. The EE behaves the same way in LoS and NLoS scenarios. Interestingly, distinct network operating regimes can be identified as M/K increases. While edge-cell UEs (i.e., $\beta_0^0 \in [-110, -80]$ dB) experiences an EE gain, UEs close to their BS (i.e., $\beta_0^0 \in [-70, -50]$ dB) incur in an EE loss;

  The use of the MR combiner at the BS is convenient, with respect to a multicell processing (e.g., M-MMSE), when the network operates in a noise-limited regime. Under this settings, we do not really need interference suppression capabilities and the small SE gain due to the use of multicell processing is counteracted by its increasing complexity. Interestingly, this result is true irrespective of the propagation conditions;
- In the interference-limited regime, i.e., $p/B > \tau$, different behaviors are found depending on whether we are in LoS or NLoS scenario. In NLoS propagation, the M-MMSE achieves the highest EE with respect to MR, whereas the opposite is true in the LoS case, where MR achieves a smaller but significant EE gain compared to M-MMSE. The change in the EE behavior is due to the reduced SE benefits of M-MMSE on MR with respect to the noise-limited regime in the LoS case;
- Unlike the noise-limited regime, for which the EE has distinct operating regimes as M/K increases, in the interference-limited regime the EE is an increasing function of M/K both for MR (LoS and NLoS scenarios) and M-MMSE (NLoS scenario only). Therefore, the adoption of an M-MMSE combiner at the BS pays off in terms of EE when the network operates in an interference-limited regime and in presence of NLoS propagation.

9.4 State of the art on energy efficiency analysis

9.4.1 Impact of cooperation

The analysis in the previous section showed that interference plays a major role in the fundamental limits of EE. This motivated the use of interference-suppressing processing like M-MMSE. Further benefits may be obtained by enabling cooperation among different BSs to jointly serve the active UEs [39]. This scenario is depicted in Figure 9.10 in which the UEs are simultaneously served by *multiple* BSs, connected

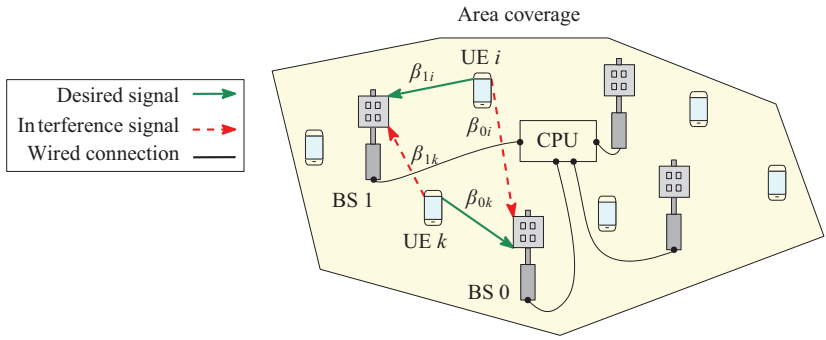


Figure 9.10 Illustration of a Network MIMO scenario with two BSs. There are no cells and the BSs are connected via infinite-capacity backhaul links to a CPU. The UL transmission of the desired UE k in cell 0 is received from both the BSs and sent to the CPU that performs global receive combining

via backhaul links to a centralized processing unit (CPU). The received signal at each BS is sent along with the channel estimates to the CPU that is responsible of combining and decoding.

Different levels of cooperation can be envisaged. In what follows, we consider the case of full cooperation among BSs and perfect CSI. This implies that the backhaul links connecting the BSs are modeled as zero-latency, error-free, and having unlimited capacity. This network setup constitutes the fundamental limit of any multiantenna communications system and serves as a benchmark to evaluate the performance gap in terms of EE with a classical massive MIMO system in which BSs do not cooperate. Following the analysis mentioned earlier, we consider both LoS and NLoS propagation scenarios.

Consider a Network MIMO system with two BSs distributed, each one of which has M antennas and serves simultaneously $2K$ UEs as in Figure 9.10. During UL transmission, the complex-valued M-dimensional received signal at BS ℓ is

$$\mathbf{y}_\ell = \sum_{i=1}^{2K} \mathbf{h}_{\ell i} s_i + \mathbf{n}_\ell, \tag{9.41}$$

where $\mathbf{h}_{\ell i}$ is the channel response between UE i and BS ℓ having an average channel gain $\beta_{\ell i}$, $s_i \sim \mathcal{N}_{\mathbb{C}}(0, p_i)$ with $i = 1, \ldots, 2K$ are the transmitted signals, and $\mathbf{n}_\ell \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, BN_0\mathbf{I}_M)$ is the noise vector. Under the assumption of perfect backhaul links, the receive data $\mathbf{y}_\ell$ at the two BSs are passed without any loss of information to the CPU. This aggregate signal can be equivalently rewritten in column form $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}]^{\mathrm{T}}$ as

$$\mathbf{y} = \sum_{i=1}^{2K} \mathbf{h}_i s_i + \mathbf{n}, \tag{9.42}$$

where $\mathbf{h}_i = [\mathbf{h}_{1i}^{\mathrm{T}}, \mathbf{h}_{2i}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{n} = [\mathbf{n}_1^{\mathrm{T}}, \mathbf{n}_2^{\mathrm{T}}]^{\mathrm{T}}$ such that $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{2M}, BN_0\mathbf{I}_{2M})$.

The sum channel capacity in a Network MIMO system is given by [27]

$$C_{\text{sum}} = \log_2 \det\left(\mathbf{I}_{2M} + \frac{1}{BN_0}\mathbf{HPH}^{\text{H}}\right), \tag{9.43}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{2K}]$ and $\mathbf{P} = \text{diag}(p_1, \ldots, p_{2K})$. The expression in (9.43) coincides with the capacity of a single-user MIMO system[†] [40] with $2K$ transmit antennas and $2M$ receive antennas under transmission of statistically independent signals (i.e., diagonal transmit correlation matrix) with individual per-antenna power constraints. Unlike in the MIMO model, the single-antenna UEs in Network MIMO communicate with multiple-antenna BSs. The lack of cooperation between antennas at the transmit side does not decrease the channel capacity. Also, there are different average channel gains between each pair of transmit and receive antennas, thus offering a macroscopic diversity gain with respect to the MIMO case; that is, every BS may experience a very different channel condition to the desired UE. Notice that the capacity in (9.43) is achieved by using nonlinear processing [27].

In the case where every UE transmits with the same power $p_k = p$, the sum channel capacity for the LoS and NLoS propagation scenarios are reported below.

Lemma 9.3 (LoS propagation). *The sum channel capacity (bit/s/Hz) in the LoS case is*

$$C_{\text{sum}} = \sum_{j=1}^{\min(2M,2K)} \log_2\left(1 + \frac{p}{BN_0}\sigma_j^2(\mathbf{H})\right), \tag{9.44}$$

where $\sigma_j(\mathbf{H})$ are the singular values associated with the global channel matrix $\mathbf{H}$ of $\text{rank}(\mathbf{H}) \leq \min(2M, 2K)$.

Lemma 9.4 (NLoS propagation). *The sum channel capacity (bit/s/Hz) in the NLoS case is*

$$C_{\text{sum}} = \sum_{j=1}^{\min(2M,2K)} \mathbb{E}\left\{\log_2\left(1 + \frac{p}{BN_0}\sigma_j^2(\mathbf{H})\right)\right\}. \tag{9.45}$$

where the expectation is taken with respect to the channel realizations.

In Figure 9.11, the sum SE achieved with Network MIMO is compared to that provided by a massive MIMO system with the optimal M-MMSE as a function of the SNR in dB. We assume $M = 50$, $K = 10$, $\bar{\beta} = -15$ dB, $N_0 = -174$ dBm/Hz, and half-wavelength antenna spacing. Network MIMO largely outperforms massive MIMO for the entire range of SNR values and provides roughly the same performance under LoS and NLoS scenarios. This is due to the combined effects of cooperation among the BSs and the optimal nonlinear signal processing at the CPU. Compared to massive MIMO, the SE with Network MIMO is increased by $5\times$ at SNR $= 0$ dB and by a factor of 4 at SNR $= 15$ dB.

[†]A single multiple-antenna UE communicates with a single multiple-antenna BS. Cooperation takes place both at transmit and receive sides.

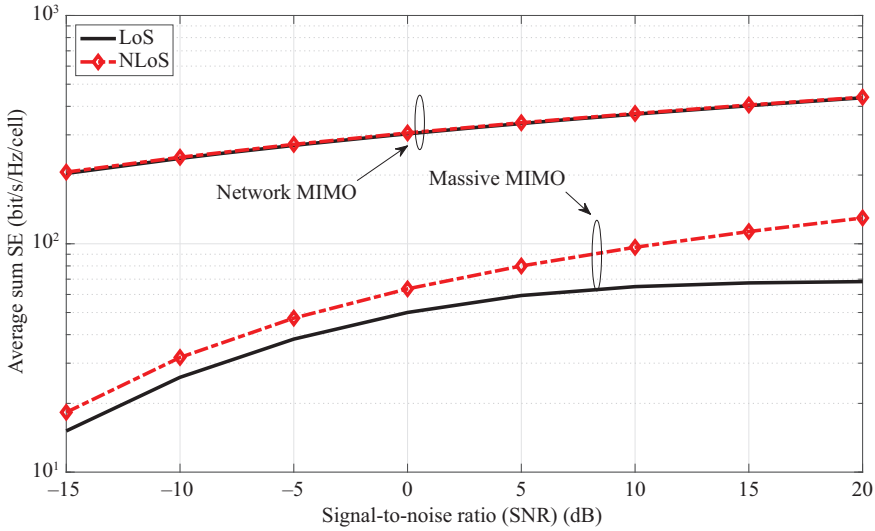


Figure 9.11 Average UL sum SE as a function of the SNR_0 in dB. There are $M = 50$ BS antennas, $K = 10$ active UEs per cell and the average relative interference between the two cells is $\bar{\beta} = -15$ dB. The (optimal) nonlinear SIC precoding scheme is compared to the (linear) M-MMSE scheme at BS under LoS and NLoS propagation

The EE of the two communication systems is numerically evaluated in Figure 9.12 as a function of SNR in dB, using the power consumption model in (9.31). To take into account the increased complexity of the non-linear processing, the values of the hardware characterization constants are increased by a factor 10 with respect to the ones listed in Table 9.1. In both cases, the EE is a unimodal function of the SNR and achieve the same maximum at roughly SNR $= -10$ dB.

In brief, there are operating regions in which, we can increase the network EE by allowing cooperation among the serving BSs. In a practical setting, however, where SNR $\in \{-20, 3\}$ dB, massive MIMO is surprisingly better than Network MIMO. This result, which is evaluated numerically only, depends highly on the network power consumption parameters and needs a more deep analysis. Thus, despite the advantage of introducing cooperation among different BSs in terms of SE is clear, it is not well understood whether the EE benefits as well. Certainly, the EE of Network MIMO is more flat around its maximum than massive MIMO, as showed in Fig. 9.12. In other words, the EE-optimal operating point is more robust to system changes due to varying network parameters.

9.4.2 Impact of imperfect channel knowledge

So far, we have assumed that the network operates with perfect CSI. In practice, this condition is never met and the BSs must estimate the channels regularly in order

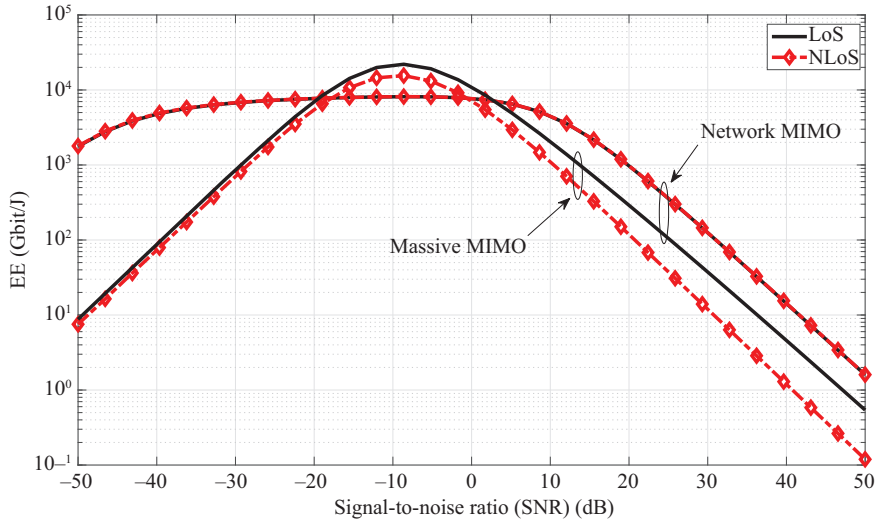


Figure 9.12 *EE as a function of the SNR_0 in dB. There are $M = 50$ BS antennas, $K = 10$ active UEs per cell, and the average relative interference between the two cells is $\bar{\beta} = -15$ dB. The (optimal) nonlinear SIC precoding scheme is compared to the (linear) M-MMSE scheme at BS under LoS and NLoS propagation*

to combine the received signals. The channel responses are approximately constant within each of its time–frequency coherence block, which typically spans a time interval (τ_c) in the order of a few milliseconds and a frequency interval (B_c) of a few hundred of Hz [10,15]; they are, respectively, called coherence time and coherence bandwidth.

The main method for CSI acquisition in massive MIMO is UL pilot signaling, where a predefined signal $\boldsymbol{\phi}_{0i} \in \mathbb{C}^{\tau_p}$ such that $\|\boldsymbol{\phi}_{0i}\|^2 = \tau_p$—chosen from an orthogonal set of pilot sequences $\boldsymbol{\Phi} \in \mathbb{C}^{\tau_p \times \tau_p}$—is sent from each UE i in cell 0. The BS 0 collects these pilot signals over a certain time interval $\tau_p \ll \tau_c$ to obtain $\mathbf{Y}_0^{\mathrm{p}} \in \mathbb{C}^{M \times \tau_p}$ and then correlates the observed signal to every pilot signals in order to obtain a sufficient statistic $\mathbf{y}_{00k}^{\mathrm{p}} = \mathbf{Y}_0^{\mathrm{p}} \boldsymbol{\phi}_{0k}^*$ for the desired channel associated, e.g., with UE k [10,15] we obtain

$$\mathbf{y}_{00k}^{\mathrm{p}} = \sqrt{p_{0k}}\tau_p \mathbf{h}_{0k}^0 + \sum_{(\ell,i)\in\mathscr{P}_{0k}} \sqrt{p_{\ell i}}\tau_p \mathbf{h}_{\ell i}^0 + \mathbf{n}_0^{\mathrm{p}}, \tag{9.46}$$

where $\mathscr{P}_{0k} = \{(\ell, i) : \boldsymbol{\phi}_{\ell i} = \boldsymbol{\phi}_{0k}\}$ is the set of indices of all UEs that utilize the same pilot signal as UE k in cell 0 and $\mathbf{n}_0^{\mathrm{p}} = \mathbf{N}_0^{\mathrm{p}} \boldsymbol{\phi}_{0k}^* \sim \mathscr{N}_{\mathbb{C}}(\mathbf{0}_M, \sigma^2 \tau_p \mathbf{I}_M)$ with σ^2 being the variance of each entry of $\mathbf{N}_0^{\mathrm{p}}$. The sufficient statistics are then used to acquire channel knowledge. Different methods can be used, each with a different trade-off between complexity and estimation accuracy. Typically, the channels are assumed to be Gaussian distributed [41], and Bayesian estimators are thus preferable. The MMSE

estimator minimizes the MSE between the current realization and its estimate. For an i.i.d. Rayleigh fading channel between UE k in cell 0 and its serving BS 0, this is given by

$$\hat{\mathbf{h}}_{0k}^{0} = \frac{\sqrt{p_{0k}}\beta_{0k}^{0}}{\sum_{(\ell,i)\in\mathscr{P}_{0k}} p_{\ell i}\tau_p\beta_{\ell i}^{0} + \sigma^2}, \tag{9.47}$$

which requires the knowledge of the second-order statistics (i.e., pathloss) of the channels over the entire network.

Fundamentally, the performance of a massive MIMO network as $M \to \infty$ is limited by the so-called coherent-interference. This name is due to the fact that its average power scales proportionally to M, just as the desired signal power [10]. Coherent-interference arises when the same pilot sequence is assigned to different UEs within a limited area coverage, which occurs in practice since the channel estimation phase duration must be much shorter than the regular period over which the channel has to be estimated, i.e., $\tau_p \ll \tau_c$. Hence, to limit the number of orthonormal pilot sequences τ_p, BSs must share them among multiple UEs across the network. This lack of orthogonality among different UE signatures leads to channel estimates indistinguishable at the BS up to a proportionality factor in the case of uncorrelated Rayleigh fading. Nevertheless, when there is a certain degree of spatial correlation among these channels, the use of a smart combiner at the BS that exploits this intrinsic structure of the channel statistics allows the rejection of coherent interference [28]. In fact, asymptotically as $M \to \infty$, the M-MMSE receiving scheme is found to achieve unlimited capacity when the number K of UE and number of cell are fixed.

9.4.3 Impact of spatial correlation

Throughout this book chapter we used i.i.d. Rayleigh fading for modeling an NLoS propagation scenario where the channel is described as in (9.3) and given by a collection of zero-mean, circularly-symmetric, complex Gaussian random variables. Despite its simplicity, this model is still the basis of most theoretical research in multiple antenna systems. For example, in massive MIMO it leads to neat, understandable closed-form expressions for the SE [10]. Practical channels among different antenna pairs, however, are spatially correlated [42]. This is because the physical propagation environment makes some propagation directions more probable to carry strong signals than others, and the antenna patterns are generally not isotropic either. The impact of spatial correlation in massive MIMO is studied in detail in [15] in terms of SE and EE. A tractable way to model spatially-correlated channels with NLoS path is the correlated Rayleigh fading model [29]: $\mathbf{h}_{0k}^{0} \sim \mathscr{N}_{\mathbb{C}}(\mathbf{0}_M, \mathbf{R}_{0k}^{0})$, where $\mathbf{R}_{0k}^{0} \in \mathbb{C}^{M\times M}$ is the spatial correlation matrix. Different models exist for the generation of $\mathbf{R}_{0k}^{0}$ based on the geometry of the propagation environment (e.g., the model in [15, Sec. 7.3]). The MMSE estimator for a correlated Rayleigh fading channel between UE k in cell 0 and its serving BS 0 is obtained as

$$\hat{\mathbf{h}}_{0k}^{0} = \sqrt{p_{0k}}\mathbf{R}_{0k}^{0}\boldsymbol{\Psi}_{0k}^{0}\mathbf{y}_{00k}^{\mathrm{p}}, \tag{9.48}$$

where $\boldsymbol{\Psi}_{0k}^{0} = (\sum_{(\ell,i)\in\mathscr{P}_{0k}} p_{\ell i}\tau_p\mathbf{R}_{\ell i}^{0} + \sigma^2\mathbf{I}_M)^{-1}$. The above formula for the MSE estimator depends jointly on the channel second-order statistics (i.e., the spatial

correlation matrices) of all the intra- and inter-cell channels and the sufficient statistic [10,15]. Differently from (9.47) that relies on the pathloss of every channel, (9.48) requires the knowledge of the spatial correlation matrices of the channels over the entire network.

Previous works dealing with the EE in multiple antenna networks using the MMSE channel estimator can be found in [43–45], where it is shown that the maximal EE is achieved by a massive MIMO setup. If full knowledge of the spatial correlation matrices is not available, one can alternatively use the element-wise MMSE (EW-MMSE) estimator, which requires only knowledge of the diagonal elements of the spatial correlation matrices. Moreover, the least squares (LS) estimator does not require any statistical knowledge and provides the lowest complexity at the cost of reduced estimation accuracy. In [15, Section 4], the impact on SE of the reduced channel estimation quality is numerically evaluated. Results show that the loss in SE incurred by using a suboptimal channel estimator is less than 10% for most linear combining schemes. Sophisticated combining schemes, such as M-MMSE, are those that benefit much from using a more complex channel estimator.

The estimation accuracy of the channel vector, including all the channels between every BS antenna and each of its served UE, increases as they become more spatially correlated. This is mainly due to the fact that the second-order statistics of the channel (i.e., the spatial correlation matrix) becomes more and more structured as spatial correlation increases, and this may be exploited during channel estimation. Also, the larger the number M of BS antennas, the lower the estimation under spatial correlation. This means that the use of large antenna arrays at the BSs increases the accuracy of estimation, despite the larger amount of parameters to be estimated. All these effects add coherently together in the SE expression, which benefits from the higher channel estimation accuracy. On one hand, the use of more complex channel estimators and a larger number of BS antennas leads to a higher SE. On the other hand, it tends to increase the digital signal processing to be performed by the network and, in turn, the CP. As expected, dealing with spatially correlated channels rather than uncorrelated channels leverage an EE trade-off since the second-order statistics of the channels now include all the cross-antenna terms, which are set to zero in the other case, with a consequent increased digital signal-processing effort for the network.

In summary, the behavior of the EE for different channel estimators is not totally understood and it is not clear whether choosing one estimator with respect to another will be beneficial or not. Therefore, the impact of the spatial correlation on the network EE requires further investigation.

9.4.4 Impact of densification

For simplicity, in this chapter we have modeled the BSs and UEs using the celebrated Wiener model [26] as if they were located at fixed spatial positions. An intermediate step was taken in [46] by considering a more complex spatial deployment over a two-dimensional hexagonal lattice. The performance of a wireless network depends critically on its spatial configuration, because received signal power and interference are critically impacted by the average channel gain based on the distances between a BS and all the UEs. The desire for tractability has led to over-simplified models, e.g., the

Wiener model, where all interfering links contribute equally to the total interference. Thus, a more realistic approach to the analysis of wireless networks includes the impact of *spatial configuration* [47] that accounts for the fact that UEs located at different distances from the BS contribute differently to the total interference power.

The exponential increase in the number of wirelessly-connected devices has leveraged the emerging of distributed network solutions to provide high data rate links to several UEs (e.g., small cells, femtocells, relays), which give rise to a more irregularly deployed and dense network infrastructure. Due to the dependence on many factors such as BS and UE locations, and buildings infrastructure, among others, a statistical approach to the modeling of network spatial configuration has been pursued and led to advanced modeling tools based on the mathematical framework of *stochastic geometry* (e.g., [48–50]). Within the stochastic geometry framework, Poisson point process (PPP) has become very popular among the wireless research community. In a PPP spatial model, the BSs are independently and uniformly distributed in a given compact area, and their locations form a point process whose cardinality is a Poisson-distributed random variable that is independent among different disjoint sets [48]. The UEs are either considered uniformly distributed across the network or drawn from a PPP considered statistically independent from the BS process. The simplest scenario is when the BSs are spatially distributed according to a homogeneous PPP in which the average number of BSs within a compact sets is proportional to its geographical area up to a proportionality constant λ (BSs/km^2), which is the intensity of the point process.

Earlier works on the EE analysis of a cellular network under a stochastic geometry framework can be found in [51,52]. Both works focus on the impact of network densification—achieved through the deployment of more and more BSs in the same geographical area—on the EE. In [51], the use of sleeping strategies together with the deployment of small cells proved to be a promising solution for increasing the EE. In [52], the EE was found to be a monotonic nondecreasing function of the cell density λ. Thus, network densification leads to a higher EE as λ increases, but this gain saturates quickly. Further benefits can be achieved by using BS network densification jointly with the massive MIMO technology.

The majority of works in the research literature use a standard pathloss model where received power decays exponentially as $d^{-\alpha}$ with the distance d between transmitter and receiver and constant decadence term α. This pathloss model is quite idealized, and practically α is itself a function of distance, typically an increasing one [53]. Despite this argument, single slope pathloss models are preferred in general by communication theorists since they made theoretical analysis much simpler than using multislope (distance dependent) pathloss models. The use of a general multislope pathloss model leads to rather different conclusions in terms of throughput and coverage probability [53,54] and EE [55]. For example, distinct operating regimes could be identified in a practical environment for which an increase, saturation, or decrease of the throughput is observed as the network densifies [54]. In the extreme case, ultra-BS-densification may even lead to zero throughput. Differently from [52], under a generic multi-slope pathloss model, the EE in [55] is found to be a unimodal function of λ, which means the cellular network achieves its maximum EE for a finite

number of BSs. This result is independent of the combining scheme implemented at the BS.

References

[1] ITU. Requirements related to technical performance for IMT-advanced radio interface(s); 2008.

[2] ITU. Minimum requirements related to technical performance for IMT-2020 radio interface(s); 2017.

[3] Rappaport TS, Sun S, Mayzus R, *et al.* Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! IEEE Access. 2013;1:335–349.

[4] Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal. 1948;27(3):379–423.

[5] Marzetta TL. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. IEEE Transactions on Wireless Communications. 2010;9(11):3590–3600.

[6] Caire G and Shamai S. On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel. IEEE Transactions on Information Theory. 2003;49(7):1691–1706.

[7] Vishwanath S, Jindal N, and Goldsmith A. Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels. IEEE Transactions on Information Theory. 2003;49(10):2658–2668.

[8] Viswanath P and Tse DNC. Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality. IEEE Transactions on Information Theory. 2003;49(8):1912–1921.

[9] Wei Yu and Cioffi JM. Sum Capacity of Gaussian Vector Broadcast Channels. IEEE Transactions on Information Theory. 2004;50(9):1875–1892.

[10] Marzetta TL, Larsson EG, Yang H, *et al.* Fundamentals of Massive MIMO. Cambridge University Press; 2016.

[11] Patel P and Holtzman J. Analysis of a Simple Successive Interference Cancellation Scheme in a DS/CDMA System. IEEE Journal on Selected Areas in Communications. 1994;12(5):796–807.

[12] Costa M. Writing on Dirty Paper. IEEE Transactions on Information Theory. 1983;29(3):439–441.

[13] Qualcomm. Rising to meet the 1000x mobile data challenge; 2012.

[14] Björnson E, Sanguinetti L, Wymeersch H, *et al.* Massive MIMO is a Reality – What is Next? Five Promising Research Directions for Antenna Arrays. Digital Signal Processing. 2019;94:3–20.

[15] Björnson E, Hoydis J, and Sanguinetti L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. Foundations and Trends® in Signal Processing. 2017;11(3–4):154–655.

[16] Ngo HQ, Larsson EG, and Marzetta TL. Aspects of favorable propagation in Massive MIMO. In: 2014 22nd European Signal Processing Conference (EUSIPCO); 2014. pp. 76–80.

[17] Gao X, Edfors O, Rusek F, *et al.* Linear pre-coding performance in measured very-large MIMO channels. In: 2011 IEEE Vehicular Technology Conference (VTC Fall); 2011. pp. 1–5.

[18] Hoydis J, Hoek C, Wild T, *et al.* Channel measurements for large antenna arrays. In: 2012 International Symposium on Wireless Communication Systems (ISWCS); 2012. pp. 811–815.

[19] Payami S and Tufvesson F. Channel measurements and analysis for very large array systems at 2.6 GHz. In: 2012 6th European Conference on Antennas and Propagation (EUCAP); 2012. pp. 433–437.

[20] Chen Z and Björnson E. Can we rely on channel hardening in cell-free massive MIMO? In: 2017 IEEE Globecom Workshops (GC Wkshps); 2017. pp. 1–6.

[21] Ngo HQ, Ashikhmin A, Yang H, *et al.* Cell-Free Massive MIMO Versus Small Cells. IEEE Transactions on Wireless Communications. 2017;16(3): 1834–1850.

[22] Ngo HQ, Tran L, Duong TQ, *et al.* On the Total Energy Efficiency of Cell-Free Massive MIMO. IEEE Transactions on Green Communications and Networking. 2018;2(1):25–39.

[23] Marzetta TL. Spatially-stationary propagating random field model for massive MIMO small-scale fading. In: 2018 IEEE International Symposium on Information Theory (ISIT); 2018. pp. 391–395.

[24] Pizzo A, Marzetta TL, and Sanguinetti L. Spatially-Stationary Model for Holographic MIMO Small-Scale Fading. IEEE Journal on Selected Areas in Communications. 2020;38(9):1964–1979.

[25] Gao X, Edfors O, Rusek F, *et al.* Massive MIMO Performance Evaluation Based on Measured Propagation Data. IEEE Transactions on Wireless Communications. 2015;14(7):3899–3911.

[26] Wyner AD. Shannon-Theoretic Approach to a Gaussian Cellular Multiple-Access Channel. IEEE Transactions on Information Theory. 1994;40(6):1713–1727.

[27] Cover TM and Thomas JA. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). New York, NY, USA: Wiley-Interscience; 2006.

[28] Björnson E, Hoydis J, and Sanguinetti L. Massive MIMO Has Unlimited Capacity. IEEE Transactions on Wireless Communications. 2018;17(1): 574–590.

[29] Sanguinetti L, Björnson E, and Hoydis J. Towards Massive MIMO 2.0: Understanding Spatial Correlation, Interference Suppression, and Pilot Contamination. IEEE Transactions on Communications. 2019;68(1):232–257.

[30] Ashikhmin A and Marzetta T. Pilot contamination precoding in multi-cell large scale antenna systems. In: 2012 IEEE International Symposium on Information Theory Proceedings; 2012. pp. 1137–1141.

[31] Auer G, Giannini V, Desset C, *et al.* How Much Energy is Needed to Run a Wireless Network? IEEE Wireless Communications. 2011;18(5): 40–49.

[32] Björnson E, Sanguinetti L, Hoydis J, *et al.* Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer? IEEE Transactions on Wireless Communications. 2015;14(6):3059–3075.

[33] Björnson E and Larsson EG. How Energy-Efficient Can a Wireless Communication System Become? In: 2018 52nd Asilomar Conference on Signals, Systems, and Computers; 2018. pp. 1252–1256.

[34] Verdu S. On Channel Capacity Per Unit Cost. IEEE Transactions on Information Theory. 1990;36(5):1019–1030.

[35] Christian S and Christopher BW, editors. Energy limits of message-passing error control decoders. In: Proceedings of IZS; 2014, Feb 26–28; Zurich, Switzerland. Zurich; 2014.

[36] Mammela A and Anttonen A. Why Will Computing Power Need Particular Attention in Future Wireless Devices? IEEE Circuits and Systems Magazine. 2017 First quarter;17(1):12–26.

[37] Zappone A and Jorswieck E. Energy Efficiency in Wireless Networks via Fractional Programming Theory. Foundations and Trends® in Communications and Information Theory. 2015;11(3–4):185–396.

[38] Hoorfar A and Hassani M. Inequalities on the Lambert W Function and Hyperpower Function. Journal of Inequalities in Pure and Applied Mathematics. 2008;9(2):1–5.

[39] Gesbert D, Hanly S, Huang H, *et al.* Multi-Cell MIMO Cooperative Networks: A New Look at Interference. IEEE Journal on Selected Areas in Communications. 2010;28(9):1380–1408.

[40] Telatar E. Capacity of Multi-antenna Gaussian Channels. European Transactions on Telecommunications. 1999;10(6):585–595.

[41] Moustakas AL, Baranger HU, Balents L, *et al.* Communication Through a Diffusive Medium: Coherence and Capacity. Science. 2000;287(5451): 287–290.

[42] Paulraj A, Nabar R, and Gore D. Introduction to Space-Time Wireless Communications. Cambridge, UK: Cambridge University Press; 2003.

[43] Yang H and Marzetta TL. Total energy efficiency of cellular large scale antenna system multiple access mobile networks. In: 2013 IEEE Online Conference on Green Communications (OnlineGreenComm); 2013. pp. 27–32.

[44] Ngo HQ, Larsson EG, and Marzetta TL. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. IEEE Transactions on Communications. 2013;61(4):1436–1449.

[45] Bjornson E, Sanguinetti L, Hoydis J, *et al.* Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer? IEEE Transactions on Wireless Communications. 2015;14(6):3059–3075.

[46] Somekh O and Shamai S. Shannon-Theoretic Approach to a Gaussian Cellular Multiple-Access Channel With Fading. IEEE Transactions on Information Theory. 2000;46(4):1401–1425.

[47] Andrews JG, Ganti RK, Haenggi M, *et al.* A Primer on Spatial Modeling and Analysis in Wireless Networks. IEEE Communications Magazine. 2010;48(11):156–163.

[48] Baccelli F and Blaszczyszyn B. Stochastic Geometry and Wireless Networks: Volume I Theory. Foundations and Trends in Networking. 2008;3(3–4): 249–449.

[49] Haenggi M, Andrews JG, Baccelli F, *et al.* Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks. IEEE Journal on Selected Areas in Communications. 2009;27(7):1029–1046.

[50] Andrews JG, Baccelli F, and Ganti RK. A Tractable Approach to Coverage and Rate in Cellular Networks. IEEE Transactions on Communications. 2011;59(11):3122–3134.

[51] Soh YS, Quek TQS, Kountouris M, *et al.* Energy Efficient Heterogeneous Cellular Networks. IEEE Journal on Selected Areas in Communications. 2013;31(5):840–850.

[52] Björnson E, Sanguinetti L, and Kountouris M. Deploying Dense Networks for Maximal Energy Efficiency: Small Cells Meet Massive MIMO. IEEE Journal on Selected Areas in Communications. 2016;34(4):832–847.

[53] Zhang X and Andrews JG. Downlink Cellular Network Analysis With Multi-Slope Path Loss Models. IEEE Transactions on Communications. 2015;63(5):1881–1894.

[54] Nguyen VM and Kountouris M. Performance Limits of Network Densification. IEEE Journal on Selected Areas in Communications. 2017;35(6):1294–1308.

[55] Pizzo A, Verenzuela D, Sanguinetti L, *et al.* Network Deployment for Maximal Energy Efficiency in Uplink With Multislope Path Loss. IEEE Transactions on Green Communications and Networking. 2018;2(3):735–750.

Chapter 10

Energy-efficient design for doubly massive MIMO millimeter wave wireless systems

Stefano Buzzi[1,2] and Carmen D'Andrea[1,2]

10.1 Introduction

Future wireless networks are expected to provide huge performance improvements compared to currently available systems [1]. Fifth-generation (5G) and beyond wireless systems will support three different service classes, namely, enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultrareliable and low-latency communications (URLLC) [2]. The eMBB class will support stable connections with very high peak data rates; the mMTC will support a massive number of Internet of Things devices, which are only sporadically active and send small data payloads, while, finally, URLLC will support low-latency transmissions of small payloads with very high reliability from a limited set of terminals, which are active according to patterns typically specified by outside events. Among the main factors that are at the foundation of 5G and beyond networks we find (a) the reduction in the size of the radio cells, so that a larger area spectral efficiency (SE) can be achieved; (b) the use of large-scale antenna arrays at the base stations (BSs), i.e., massive multiple-input–multiple-output (MIMO) systems [3]; and (c) the use of carrier frequencies in the range 10–100 GHz, also known as millimeter waves (mm-waves)* [4]. Factor (a) is actually a trend that has been observed for some decades, since the size of the radio cells has been progressively reduced over time from one generation of cellular networks to the next one. Factor (b) has been considered starting from the fourth generation of wireless networks, and indeed the latest Third-Generation Partnership Project LTE releases already include the possibility of equipping BSs with antenna arrays of up to 64 elements; this trend continues in the 5G New Radio standard, where fully digital (FD) beamforming structures are adopted. Finally, factor (c) is instead a more recent technology, since for decades mm-waves have been considered as being unsuited for cellular wireless networks.

[1]Department of Electric and Information Engineering (DIEI), University of Cassino and Southern Latium, Cassino, Italy
[2]Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Parma, Italy
*Even though mm-waves is a term that historically refers to the range 30–300 GHz, in the recent literature about future wireless networks, the term is used to refer to frequency above 6 GHz.

10.1.1 State of the art

Focusing on the massive MIMO technology, most of the research and experimental work has traditionally considered its use at conventional cellular frequencies (e.g., sub-6 GHz). Such a range of frequencies is generally denoted as microwave. In the recent past, the combination of the massive MIMO concept with the use of mm-wave frequency bands has started being considered [5,6]. For conventional sub-6 GHz cellular systems, it has been shown that equipping a BS with a very large (>100) number of antennas [3,7,8] significantly increases the network capacity, mainly due to the capability of serving several users on the same frequency slot with nearly orthogonal vector channels. In the traditional massive MIMO literature, while the number of antennas at the BS grows large, the user device is usually assumed to have only one or very few antennas. When moving to mm-wave, however, the wavelength gets reduced, and, at least in principle, a large number of antennas can be mounted not only on the BS but also on the user device. As an example, at a carrier frequency of 30 GHz, the wavelength is 1 cm, and for a planar antenna array with half-wavelength spacing, more than 180 antennas can be placed in an area as large as a standard credit card. This leads to the concept of *doubly massive MIMO system* [6,9] that is defined as a wireless communication system where the number of antennas grows large at both the transmitter and the receiver.

While there are certainly a number of serious practical constraints—e.g., large power consumption, low efficiency of power amplifiers (PAs), hardware complexity, analog-to-digital conversion and beamformer implementation—that currently prevent the feasibility of the use of large-scale antenna arrays at both sides of the communication links, it is, on the other hand, believed that these are just technological issues that will be solved or worked around in the near future. Since in massive MIMO systems, hardware complexity and energy consumption issues may be problematic, FD beamforming, which requires one radio-frequency (RF) chain for each antenna element, is very challenging; as a consequence, recent research efforts have been devoted toward devising suboptimal, lower complexity, beamforming structures [10]. Hybrid (HY) beamforming structures have been proposed, with a limited number (much smaller than the number of antenna elements) of RF chains. The paper [11], as an instance, analyzes the achievable rate for an MU-MIMO system with HY precoding and limited feedback; it is therein shown that, for the case of single-path (i.e., rank-one) channels, HY precoding structures achieve an SE very close to that of an FD beamformer. In [12], it is shown that an HY beamformer with twice as many RF chains as transmitted data streams may exactly mimic an FD beamformer; the analysis neglects energy efficiency (EE) issues, assumes perfect channel state information (CSI), and is limited to either a single-user MIMO system or an MU-MIMO system with single-antenna receivers. The paper [13] proposes a new low-complexity postcoding structure, based on switches rather than on analog phase shifters; the performance of this new structure is evaluated in a rather simple scenario, i.e., single-user MIMO system with a limited number of transmit and receive antennas. The paper [14] focuses on sub-6 GHz frequencies and introduces a novel postcoding structure made of fixed (rather than tunable) phase shifters and of switches, under the assumption that the receiver is

equipped with a large array, while the transmitters have only one antenna. In [15], the authors consider five different low-complexity decoding structures, all based on the use of phase shifters and switches, and provide an analysis of the achievable SE along with estimates of the energy consumption of the proposed structures. The paper [16] considers the issue of EE maximization in a downlink massive MIMO mm-wave system by deriving an energy-efficient HY beamformer; however, the paper considers the case in which the user terminals are equipped with just one antenna, and this is a key assumption that is exploited to solve the considered optimization problems. Recently, paper [17] shows that the FD beamforming with zero-forcing (ZF) beamforming at the BS outperforms low-complexity structures both in terms of spectral and energy efficiencies. This trend is also confirmed in [18–20] where the FD beamforming is again considered as a possible candidate technology in order to achieve high EE at mm-wave frequencies.

10.1.2 Chapter organization

This chapter focuses on the analysis of doubly massive MIMO mm-wave systems, proposing several beamforming structures and comparing them from the point of view of the EE. In particular, a detailed description of doubly massive MIMO systems at mm-waves is given highlighting the differences with respect to massive MIMO systems operating at microwave frequencies. Some relevant use cases where the use of doubly massive MIMO systems may turn out to be extremely useful are then discussed. Afterward, building upon mm-wave channel models available in the current literature, the clustered channel model is described. Two main performance measures are defined, i.e., the achievable SE and the global EE, which are defined as the ratio between the SE and the total power consumption of the network. Beamforming structures are then detailed for FD, HY, and analog implementation, with emphasis on the power consumption of each structure. An asymptotic analysis is performed in the limit of a large number of transmit and receive antennas, and SE expressions for the FD and analog beamforming (AB) structures in this regime are derived. Using asymptotic formulas for the large number of antennas regime, low-complexity power allocation strategies aimed at the EE maximization are derived. In particular two different techniques are described, one relying on the asymptotic expressions in the interference-free case and the other relying on the interference-limited case. In both cases, the EE-maximizing power allocation is obtained by means of fractional programming techniques [21]. Numerical results are finally presented to corroborate the analysis and validate the theoretical findings.

This chapter is organized as follows. The next section contains the description of doubly massive MIMO systems and the discussion of some relevant use cases. Section 10.3 reports the system model, the transceiver processing, and the definition of the performance measures. Section 10.4 details the considered beamforming structures, while Section 10.5 derives the asymptotic SE expressions for the digital and AB structures. Section 10.6 contains the EE-maximizing power allocation strategies based on the derived asymptotic expressions, while Section 10.7 contains the discussion

about the obtained numerical results. Finally, concluding remarks are given in Section 10.8.

10.1.3 Notation

The following notation is used. The transpose, the inverse, and the conjugate transpose of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^{-1}$, and $\mathbf{A}^H$, respectively. The square root of the matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{1/2}$. The trace and the determinant of the matrix $\mathbf{A}$ are denoted as $\text{tr}(\mathbf{A})$ and $|\mathbf{A}|$, respectively. The magnitude of the complex scalar a is denoted as $|a|$. The N-dimensional identity matrix is denoted as $\mathbf{I}_N$, the $(N \times M)$-dimensional matrix with all zero entries is denoted as $\mathbf{0}_{N\times M}$. The matrix containing the rows of $\mathbf{A}$ from the ℓth to the ℓ'th and the columns from the mth to the m' is denoted as $\mathbf{A}_{(\ell:\ell',m:m')}$. The diagonal matrix obtained from scalars $a_1, \ldots, a_N$ is denoted by $\text{diag}(a_1, \ldots, a_N)$. The statistical expectation operator is denoted as $\mathbb{E}[\cdot]$; $\mathscr{CN}\left(\mu, \sigma^2\right)$ denotes a complex circularly symmetric Gaussian random variable with mean μ and variance σ^2 and $\mathscr{U}(a, b)$ denotes a random variable uniformly distributed in the range $[a, b]$.

10.2 Doubly massive MIMO systems

The idea of using large-scale antenna array was originally launched by Marzetta [7] with reference to BSs. The paper showed that in the limit of a large number of BS antennas, small-scale fading effects vanish by virtue of channel hardening and favorable propagation effects. The former effect makes a fading channel behave as deterministic and the latter makes the directions of two users channels asymptotically orthogonal. Consequently, plain channel-matched (CM) beamforming at the BS permits serving several users on the same time–frequency resource slot with (ideally) no interference, and the only left impairment is imperfect channel estimates due to the fact that orthogonal pilots are limited and they must be reused throughout the network (this is the so-called pilot contamination). Recent research has shown that pilot contamination represents a fundamental limit of massive MIMO system only if maximum ratio combining/precoding is considered [22], and that advanced signal processing techniques can be used to mitigate this effect [23]. Reference [7] considered a system where user devices were equipped with just one antenna. Further studies have extended the massive MIMO idea at microwave frequencies to the case in which the user devices have multiple antennas, but this number is obviously limited to few units. Indeed, at microwave frequencies the wavelength is in the order of several centimeters, and it is thus difficult to pack many antennas on small-sized user devices. At such frequencies, thus, massive MIMO just refers to BSs. Things are instead different at mm-waves, wherein multiple antennas are necessary first and foremost to compensate for the increased path loss with respect to conventional sub-6 GHz frequencies [24,25]. The feasibility of communicating at a high rate in line-of-sight (LOS), benefiting from the wide available bandwidth, also over long distances has been exploited using high-gain directional antennas. Instead of deploying a huge array at one side of the link, the same signal-to noise ratio can be achieved by deploying substantially smaller

arrays at both sides. The beamforming gains are multiplied together, so, considering the downlink, instead of having 1,000 antennas at the transmitter to serve single-antenna receivers, we can have 100 antennas at the transmitter and 10 antennas at the receivers. This also opens the door to explore systems with massive arrays at both sides. This consideration leads to the concept of doubly massive MIMO system, first introduced in [6,17], wherein the number of antennas grows large at both sides of the communication link.

10.2.1 Differences with massive MIMO at microwave frequencies

An understanding of the electromagnetic propagation is crucial when considering massive MIMO systems at mm-wave and microwave frequencies. The propagation channels build on the same physics, but basic phenomena such as diffraction, attenuation, and Fresnel zones are substantially different [25]. Now, while at microwave, the use of hybrid beamformer brings unavoidable performance degradation, at mm-waves something different happens in the limiting regime of a large number of antennas by virtue of the different propagation mechanisms. In particular, in [24] it has been shown that the use of large-scale antenna arrays at mm-wave has not an as beneficial impact on the system multiplexing capabilities as it has at microwave frequencies. In fact, the *clustered* structure of the mm-wave channel makes the purely analog (beam-steering) beamforming optimal in a single-user scenario. Otherwise stated, in a single-user link, the channel eigendirections associated with the largest eigenvalues are just the beam-steering vectors corresponding to the arrival and departure angles and associated with the predominant scatterers. This suggests that precoding and postcoding simply require pointing a beam toward the predominant scatterer at the transmitter and at the receiver, respectively. Additionally, the availability of doubly massive MIMO wireless links enables the generation of very narrow beams, resulting in reduced co-channel interference to other users using the same time–frequency resources. Another key difference between massive MIMO at mm-waves with respect to microwave is the fact that the computational complexity of channel estimation weakly depends on the number of antennas, especially for the case in which analog (beam-steering) beamforming strategies are used [24,25].

10.2.2 Use cases

The use of doubly massive MIMO systems at mm-wave in the traditional vision of cellular communication may be unfeasible due to practical constraints that prevent the use of large antenna arrays on mobile handheld devices. However, different and realistic use-cases for doubly massive MIMO systems can be already envisioned. First, considering currently available technology, doubly massive MIMO mm-wave systems can be used for the vehicle-to-vehicle and vehicle-to-infrastructure communications. For example, the authors of the paper [26] consider massive MIMO access points mounted on street lampposts spaced at very short intervals and they compare the performance obtained with mm-wave and microwave frequencies. The use of large antenna arrays at both sides of the communications can be applied, again with currently available technology, in order to implement the wireless backhaul in a cellular

network. Since mm-wave is also suitable for access links, an integrated access and backhaul architecture for the 5G cellular networks, in which the same infrastructure and spectral resources will be used for both access and backhaul links [27], can be realized.

Considering now emerging technologies that are not available in today's networks but have significant potential for future sixth-generation (6G) systems, the role of the mm-waves and consequently of the doubly massive MIMO systems will be crucial. The market demands of 2030 and beyond will introduce new applications, with more stringent requirements, in terms of ultrahigh reliability, capacity, EE, and low latency, which may saturate the capacity of traditional technologies for wireless systems and 6G will contribute to fill this gap [28]. Focusing on the capacity requirements, [26] confirms that mm-wave massive MIMO can deliver Gbps data rates for next-generation wireless networks. The availability of such data rates will allow the implementation of augmented reality (AR) and virtual reality (VR) applications. AR and VR over wireless will become a key application in various use cases including, but not limited to, education, training, gaming, workspace communication, and general entertainment. VR/AR applications will require a large amount of data rate and extremely low latency, so that the use of mm-wave can be crucial both for large bandwidths but also for the high gain obtained from large arrays at both the sides of the communication links. Other 6G applications that require high data rate and very low latency are *holographic telepresence* and *eHealth*: these applications can benefit from mm-wave and doubly massive MIMO systems. Another interesting application of the doubly massive MIMO systems can be in the improving the *indoor coverage*. Network infrastructures operating in the mm-wave spectrum will hardly provide indoor connectivity as high-frequency radio signals cannot easily penetrate solid material. One solution to improve the indoor coverage using mm-wave can be the use of cell-free massive MIMO systems [29,30] that can be used in order to guarantee a good coverage for indoor users; some results on cell-free mm-wave massive MIMO systems can indeed be found in [31,32].

10.3 System model

We focus on the downlink of a doubly massive MIMO system at mm-wave. The parameter N_T denotes the number of antennas at the transmitter, and N_R denotes the number of antennas at the receiver.[†]

10.3.1 The clustered channel model

The popular narrowband clustered mm-wave channel model is assumed to hold [33–36]. The considered reference scenario for the clustered channel model is reported in Figure 10.1. The baseband equivalent of the propagation channel between the BS

[†]For the sake of simplicity, all the receivers are assumed to have the same number of antennas; however, this hypothesis can be easily relaxed.

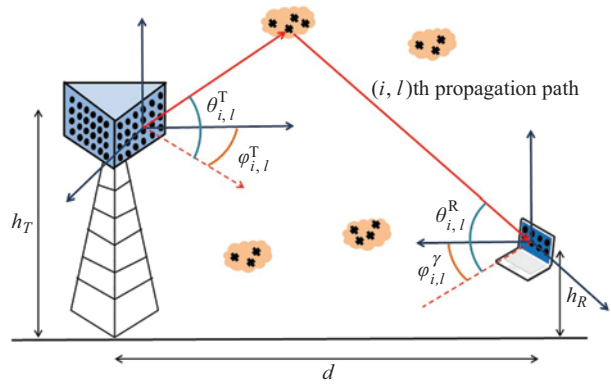


Figure 10.1 The considered reference scenario

and the generic receiver[‡] is thus represented by an $(N_{\mathrm{R}} \times N_{\mathrm{T}})$-dimensional matrix expressed as[§]

$$\mathbf{H} = \gamma \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray},i}} \alpha_{i,l} \sqrt{L(r_{i,l})} \mathbf{a}_{\mathrm{R}}(\phi^{\mathrm{R}}_{i,l}) \mathbf{a}^{H}_{\mathrm{T}}(\phi^{\mathrm{T}}_{i,l}) + \mathbf{H}_{\mathrm{LOS}}. \tag{10.1}$$

In (10.1), it is implicitly assumed that the propagation environment is made of N_{cl} scattering clusters, each of which contributes with $N_{\mathrm{ray},i}$ propagation paths, $i = 1, \ldots, N_{\mathrm{cl}}$, plus a possibly present LOS component. The parameters $\phi^{\mathrm{R}}_{i,l}$ and $\phi^{\mathrm{T}}_{i,l}$ denote the angles of arrival and departure of the lth ray in the ith scattering cluster, respectively. The quantities $\alpha_{i,l}$ and $L(r_{i,l})$ are the complex path gain and the attenuation associated to the (i,l)th propagation path. The complex gain $\alpha_{i,l} \sim \mathcal{CN}(0, \sigma^2_{\alpha,i})$, with $\sigma^2_{\alpha,i} = 1$ [33]. The vectors $\mathbf{a}_{\mathrm{R}}(\phi^{\mathrm{R}}_{i,l})$ and $\mathbf{a}_{\mathrm{T}}(\phi^{\mathrm{T}}_{i,l})$ represent the normalized receive and transmit array responses evaluated at the corresponding angles of arrival and departure; for an uniform linear array (ULA) with half-wavelength inter-element spacing it holds:

$$\mathbf{a}_{\mathrm{T}}(\phi^{\mathrm{T}}_{i,l}) = \frac{1}{\sqrt{N_{\mathrm{T}}}} \left[1 \; e^{-j\pi \sin\phi^{\mathrm{T}}_{i,l}} \; \cdots \; e^{-j\pi(N_{\mathrm{T}}-1)\sin\phi^{\mathrm{T}}_{i,l}} \right]^{T}. \tag{10.2}$$

A similar expression can also be given for $\mathbf{a}_{\mathrm{R}}(\phi^{\mathrm{R}}_{i,l})$. Finally:

$$\gamma = \sqrt{\frac{N_{\mathrm{R}} N_{\mathrm{T}}}{\sum_{i=1}^{N_{\mathrm{cl}}} N_{\mathrm{ray},i}}} \tag{10.3}$$

[‡]For ease of notation, we omit, for the moment, the subscript k to denote the BS to the k-th user channel matrix.

[§]For the sake of ease of explanation, a frequency flat channel model is considered here. A wideband channel model as in [11,37] could, however, be adopted.

is a normalization factor ensuring that the received signal power scales linearly with the product $N_{\mathrm{R}}N_{\mathrm{T}}$. With regard to the attenuation of the (i,l)-th path, we considered the results of [38] for four different use-case scenarios: Urban Microcellular (UMi) Open-Square, UMi Street-Canyon, Indoor Hotspot (InH) Office, and InH Shopping Mall. Following [38], the attenuation of the (i,l)-th path is written in logarithmic units as

$$L(r_{i,l}) = -20\log_{10}\left(\frac{4\pi}{\lambda}\right) - 10n\left[1 - b + \frac{bc}{\lambda f_0}\right]\log_{10}\left(r_{i,l}\right) - X_\sigma, \tag{10.4}$$

with n the path-loss exponent, X_σ the zero mean, σ^2-variance Gaussian-distributed shadow fading term in logarithmic units, b a system parameter, and f_0 a fixed reference frequency, the centroid of all the frequencies represented by the path-loss model. The values for all these parameters for each use-case scenario are reported in Table 10.1.

Regarding the LOS component, denoting by $\phi^{\mathrm{R}}_{\mathrm{LOS}}$, $\phi^{\mathrm{T}}_{\mathrm{LOS}}$ the arrival and departure angles corresponding to the LOS link, it is assumed that:

$$\mathbf{H}_{\mathrm{LOS}} = I_{\mathrm{LOS}}(d)\sqrt{N_{\mathrm{R}}N_{\mathrm{T}}L(d)}e^{j\theta}\mathbf{a}_{\mathrm{R}}(\phi^{\mathrm{R}}_{\mathrm{LOS}})\mathbf{a}^H_{\mathrm{T}}(\phi^{\mathrm{T}}_{\mathrm{LOS}})\ . \tag{10.5}$$

In the earlier equation, $\theta \sim \mathscr{U}(0, 2\pi)$, while $I_{\mathrm{LOS}}(d)$ is a random variate indicating if an LOS link exists between transmitter and receiver, with $\widetilde{P}$ the probability that $I_{\mathrm{LOS}}(d) = 1$. Denoting by $\widetilde{P}$ the probability that $I_{\mathrm{LOS}}(d) = 1$, i.e., an LOS link exists, we use the results in [38,39]; for the UMi scenarios, we have:

$$\widetilde{P} = \min\left(\frac{20}{d}, 1\right)\left(1 - e^{-\frac{d}{39}}\right) + e^{-\frac{d}{39}}\ , \tag{10.6}$$

while for the InH scenarios we have:

$$\widetilde{P} = \begin{cases} 1 & d \le 1.2, \\ e^{-\left(\frac{d-1.2}{4.7}\right)} & 1.2 < d \le 6.5, \\ 0.32e^{-\left(\frac{d-6.5}{32.6}\right)} & d \ge 6.5. \end{cases} \tag{10.7}$$

A detailed description of all the parameters needed for the generation of sample realizations for the channel model of (10.1) is reported in [40].

Table 10.1 Parameters for path loss model [38]

Scenario	**Model parameters**
UMi Street Canyon LOS	$n = 1.98,\ \sigma = 3.1$ dB, $b = 0$
UMi Street Canyon NLOS	$n = 3.19,\ \sigma = 8.2$ dB, $b = 0$
UMi Open Square LOS	$n = 1.85,\ \sigma = 4.2$ dB, $b = 0$
UMi Open Square NLOS	$n = 2.89,\ \sigma = 7.1$ dB, $b = 0$
InH Indoor Office LOS	$n = 1.73,\ \sigma = 3.02$ dB, $b = 0$
InH Indoor Office NLOS	$n = 3.19,\ \sigma = 8.29$ dB, $b = 0.06$, $f_0 = 24.2$ GHz
InH Shopping Mall LOS	$n = 1.73,\ \sigma = 2.01$ dB, $b = 0$
InH Shopping Mall NLOS	$n = 2.59,\ \sigma = 7.40$ dB, $b = 0.01$, $f_0 = 39.5$ GHz

10.3.2 Transmitter and receiver processing

Assume that M denotes the number of streams sent to each user in each signaling interval,[‖] and $\mathbf{x}_k$ is the M-dimensional vector of the data symbols intended for the k-th user, with $\mathbb{E}\left[\mathbf{x}_k \mathbf{x}_k^H\right] = \mathbf{I}_M$; the discrete-time signal transmitted by the BS can be expressed as the N_{T}-dimensional vector:

$$\mathbf{s} = \sum_{\ell=1}^{K} \mathbf{Q}_\ell \mathbf{P}_\ell^{1/2} \mathbf{x}_\ell \,, \tag{10.8}$$

with $\mathbf{Q}_\ell$ the $(N_{\mathrm{T}} \times M)$-dimensional precoding matrix for the ℓth user and $\mathbf{P}_k^{1/2} = \mathrm{diag}\big(\sqrt{p_{\ell,1}}, \ldots, \sqrt{p_{\ell,M}}\big)$, with $p_{\ell,q}$ is the downlink transmit power for the ℓ-th user in the q-th stream. The signal received by the generic k-th user is expressed as the following N_{R}-dimensional vector:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s} + \mathbf{n}_k, \tag{10.9}$$

with $\mathbf{H}_k$ representing the clustered channel (modeled as in (10.1)) from the BS to the k-th user and $\mathbf{n}_k$ is the N_{R}-dimensional additive white Gaussian noise with zero-mean independent and identically distributed entries with variance σ_n^2. Denoting by $\mathbf{D}_k$ the $(N_{\mathrm{R}} \times M)$-dimensional postcoding matrix at the k-th user device, the following M-dimensional vector is finally obtained:

$$\mathbf{r}_k = \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_k \mathbf{P}_k^{1/2} \mathbf{x}_k + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_\ell \mathbf{P}_\ell^{1/2} \mathbf{x}_\ell + \mathbf{D}_k^H \mathbf{w}_k. \tag{10.10}$$

10.3.3 Performance measures

Two performance measures will be considered: the SE and the EE. The SE is measured in (bit/s/Hz), while the EE is measured in (bit/J) [41]. Assuming Gaussian data symbols in (10.10), the SE is [42][¶]

$$\mathrm{SE} = \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \mathbf{R}_{\mathrm{D},k}^{-1} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_k \mathbf{P}_k \mathbf{Q}_k^H \mathbf{H}_k^H \mathbf{D}_k \right|, \tag{10.11}$$

wherein $\mathbf{R}_{\mathrm{D},k}$ is the covariance matrix of the overall disturbance seen on the downlink by the k-th user receiver, i.e.,

$$\mathbf{R}_{\mathrm{D},k} = \sigma_n^2 \mathbf{D}_k^H \mathbf{D}_k + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{D}_k^H \mathbf{H}_k \mathbf{Q}_\ell \mathbf{P}_\ell \mathbf{Q}_\ell^H \mathbf{H}_k^H \mathbf{D}_k. \tag{10.12}$$

[‖] Otherwise stated, the BS transmits in each time–frequency slot MK data symbols.

[¶] This expression represents the achievable SE only under the assumption of perfect CSI that we assume to hold in this chapter. Under the assumption of imperfect CSI, instead, it is possible to derive SE bounds, as reported in [43].

The EE is defined as

$$\mathrm{EE} = \frac{W\mathrm{SE}}{\eta \sum_{k=1}^{K} \sum_{q=1}^{M} p_{k,q} + P_{\mathrm{TX,c}} + KP_{\mathrm{RX,c}}}, \tag{10.13}$$

where W is the system bandwidth, $P_{\mathrm{TX,c}}$ is the amount of power consumed by the transmitter circuitry, $P_{\mathrm{RX,c}}$ is the amount of power consumed by the receiver circuitry, and $\eta > 1$ is a scalar coefficient modeling the PA inefficiency [44]. The cost due to signal processing at both the transmitter and the receiver is contained in the terms $P_{\mathrm{TX,c}}$ and $P_{\mathrm{RX,c}}$, which will be specified later for each beamforming structure.

10.4 Beamforming structures

In the following, some beamforming precoding and postcoding structures are described, along with details on their power consumption.

10.4.1 Channel-matched, fully digital (CM-FD) beamforming

Let $\mathbf{H}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{V}_k^H$ denote the singular-value decomposition (SVD) of the matrix $\mathbf{H}_k$, and assume, without loss of generality, that the diagonal entries of $\mathbf{\Lambda}_k$ are sorted in descending order [45]. The k-th user precoding and postcoding matrices are chosen as the columns of the matrices $\mathbf{V}_k$ and $\mathbf{U}_k$, respectively, corresponding to the M largest entries in the eigenvalue matrix $\mathbf{\Lambda}_k$, $\forall k = 1, \ldots, K$, i.e.,

$$\begin{aligned} \mathbf{Q}_k^{\text{CM-FD}} &= [\mathbf{v}_{k,1}, \mathbf{v}_{k,2}, \ldots, \mathbf{v}_{k,M}], \\ \mathbf{D}_k^{\text{CM-FD}} &= [\mathbf{u}_{k,1}, \mathbf{u}_{k,2}, \ldots, \mathbf{u}_{k,M}], \end{aligned} \tag{10.14}$$

where the column vectors $\mathbf{u}_{k,i}$ and $\mathbf{v}_{k,i}$ denote the ith column of the matrices $\mathbf{U}_k$ and $\mathbf{V}_k$, respectively. In the perfect CSI case, the CM-FD beamforming is optimal in the interference-free case and tends to be optimal in the case in which the number of antennas at the transmitter grows. The considered FD precoding architecture requires a baseband digital precoder that adapts the M data streams to the N_{T} transmit antennas; then, for each antenna there are a digital-to-analog converter (DAC), an RF chain, and a PA. At the receiver, a low-noise amplifier (LNA), an RF chain, an analog-to-digital converter (ADC) are required for each antenna, plus a baseband digital combiner that combines the N_{R} outputs of ADC to obtain the soft estimate of the M transmitted symbols. The amount of power consumed by the transmitter circuitry can thus be expressed as

$$P_{\mathrm{TX,c}} = N_{\mathrm{T}}(P_{\mathrm{RFC}} + P_{\mathrm{DAC}} + P_{\mathrm{PA}}) + P_{\mathrm{BB}}, \tag{10.15}$$

and the amount of power consumed by the receiver circuitry can be expressed as

$$P_{\mathrm{RX,c}} = N_{\mathrm{R}}(P_{\mathrm{RFC}} + P_{\mathrm{ADC}} + P_{\mathrm{LNA}}) + P_{\mathrm{BB}}. \tag{10.16}$$

In the earlier equations, $P_{\mathrm{RFC}} = 40$ mW [15] is the power consumed by the single RF chain, $P_{\mathrm{DAC}} = 110$ mW [46] is the power consumed by each DAC, $P_{\mathrm{ADC}} = 200$ mW [15] is the power consumed by each ADC, $P_{\mathrm{PA}} = 16$ mW [47] is the power consumed by each PA, $P_{\mathrm{LNA}} = 30$ mW [15] is the power consumed by each LNA, and P_{BB} is

the amount of power consumed by each baseband precoder/combiner; assuming a CMOS implementation, we have a power consumption of 243 mW [48]. The values of the power consumed by the each ADC present high variability in literature [15]. A conservative value is chosen since the literature does not refer to commercial products and these values might be too optimistic with respect to the final products. The values of the power consumption of each device considered in this section are summarized in Table 10.2.

10.4.2 Partial zero-forcing, fully digital (PZF-FD) beamforming

ZF precoding nulls interference at the receiver through the constraint that the k-th user precoding be such that the product $\mathbf{H}_\ell \mathbf{Q}_k = \mathbf{0}_{N_\mathrm{T} \times M}$ for all $\ell \neq k$. In order to avoid a too severe noise enhancement, a partial ZF approach is adopted here, namely, the columns of the precoding matrix $\mathbf{Q}_k$ are required to be orthogonal to the M (the number of transmitted data streams to each user) right eigenvectors of the channel $\mathbf{H}_\ell$ corresponding to the largest eigenvalues of $\mathbf{H}_\ell$, for all $\ell \neq k$. In this way, the precoder orthogonalizes only to an $M(K-1)$-dimensional subspace and nulls the most significant part of the interference. Formally, the precoder $\mathbf{Q}_k^{\text{PZF-FD}}$ is obtained as the projection of the CM-FD precoder $\mathbf{Q}_k^{\text{CM-FD}}$ onto the orthogonal complement of the subspace spanned by the M dominant right eigenvectors of the channel matrices $\mathbf{H}_1, \ldots, \mathbf{H}_{k-1}, \mathbf{H}_{k+1}, \ldots, \mathbf{H}_K$. Otherwise stated,

$$\mathbf{Q}_k^{\text{PZF-FD}} = \sqrt{M} \frac{\left(\mathbf{I}_{N_\mathrm{T}} - \widetilde{\mathbf{V}}_k \widetilde{\mathbf{V}}_k^H\right) \mathbf{Q}_k^{\text{CM-FD}}}{\left\| \left(\mathbf{I}_{N_\mathrm{T}} - \widetilde{\mathbf{V}}_k \widetilde{\mathbf{V}}_k^H\right) \mathbf{Q}_k^{\text{CM-FD}} \right\|_F}, \tag{10.17}$$

where $\widetilde{\mathbf{V}}_k$ is a matrix containing M dominant eigenvectors associated with the nonzero eigenvalues of the matrix

$$\overline{\mathbf{V}}_k = [\mathbf{V}_1(:, 1:M), \mathbf{V}_2(:, 1:M), \ldots, \mathbf{V}_{k-1}(:, 1:M), \mathbf{V}_{k+1}(:, 1:M), \ldots, \mathbf{V}_K(:, 1:M)],$$

and the factor $\sqrt{M}$ allows us to maintain the same Frobenius norm of the precoders. The postcoding matrix is such that $\left(\mathbf{D}_k^{\text{PZF-FD}}\right)^H = \left(\mathbf{H}_k \mathbf{Q}_k^{\text{PZF-FD}}\right)^+$. Since the PZF-FD

Table 10.2 Power consumption of each device [17]

Name	Value	Device	Reference
P_{RFC}	40 mW	RF chain	[15]
P_{DAC}	110 mW	DAC	[46]
P_{ADC}	200 mW	ADC	[15]
P_{PA}	16 mW	PA	[15]
P_{LNA}	30 mW	LNA	[15]
P_{BB}	243 mW	Baseband beamformer	[48]
P_{PS}	19.5 mW	Phase shifter	[52]
P_{element}	27 mW	Element of the phased array	[47]
η	2.66	PA inefficiency	[53]

beamforming requires an FD postcoding, its power consumption is the same as that of the CM-FD beamformer.

10.4.3 Channel-matched, hybrid (CM-HY) beamforming

In order to avoid the use of the same number of RF chains as the number of antennas, HY beamforming architectures have been proposed in the literature; in particular, denoting by $N_{\mathrm{T}}^{\mathrm{RF}}$ and $N_{\mathrm{R}}^{\mathrm{RF}}$ the number of RF chains available at the transmitter and at the receiver, respectively, the k-th user precoding and postcoding matrices are decomposed as follows:

$$\mathbf{Q}_k^{\text{CM-HY}} = \mathbf{Q}_k^{\mathrm{RF}}\mathbf{Q}_k^{\mathrm{BB}}\,, \quad \mathbf{D}_k^{\text{CM-HY}} = \mathbf{D}_k^{\mathrm{RF}}\mathbf{D}_k^{\mathrm{BB}}. \tag{10.18}$$

In the earlier-mentioned decomposition, the matrices $\mathbf{Q}_k^{\mathrm{RF}}$ and $\mathbf{D}_k^{\mathrm{RF}}$ have dimensions $(N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}})$ and $(N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}})$, respectively, and their entries are constrained to have constant norm (i.e., they are implemented through a network of phase-shifters**); the matrices $\mathbf{Q}_k^{\mathrm{BB}}$ and $\mathbf{D}_k^{\mathrm{BB}}$, instead, have dimension $(N_{\mathrm{T}}^{\mathrm{RF}} \times M)$ and $(N_{\mathrm{R}}^{\mathrm{RF}} \times M)$, respectively, and their entries are unconstrained complex numbers. A block scheme of the architecture of the HY transceiver is depicted in Figure 10.2.

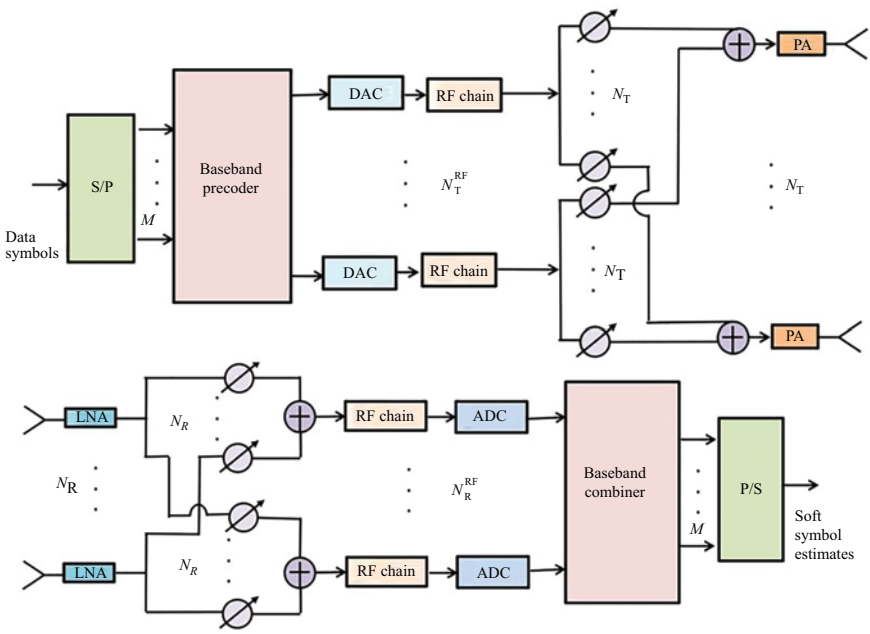


Figure 10.2 Block-scheme of a transceiver with HY digital/analog beamforming

**The case of quantized phase shifts is also considered in the literature, but it is neglected here for the sake of simplicity.

Now, designing an HY beamformer is tantamount to finding expressions for the matrices $\mathbf{Q}_k^{\mathrm{RF}}$, $\mathbf{Q}_k^{\mathrm{BB}}$, $\mathbf{D}_k^{\mathrm{RF}}$, and $\mathbf{D}_k^{\mathrm{BB}}$, so that the FD beamforming matrices are approximated with the decomposition reported in (10.18). For the CM-HY beamforming, the desired beamformers are the CM-FD matrices, and their approximation is here realized by using the block coordinate descent for subspace decomposition (BCD-SD) algorithm [49,50]. Briefly, a block coordinate descent is an optimization algorithm that successively minimizes along coordinate directions to find the minimum of a function. At each iteration, the algorithm determines a coordinate block via a coordinate selection rule then minimizes over the corresponding coordinate hyperplane while keeping fixed all other coordinates blocks. A line search along the coordinate direction can be performed at the current iteration to determine the appropriate step size [51]. The BCD-SD algorithm here used is briefly reported in Algorithm 10.1, with $\mathbf{Q}^{\mathrm{opt}} = \mathbf{Q}^{\mathrm{CM\text{-}FD}}$ and $\mathbf{D}^{\mathrm{opt}} = \mathbf{D}^{\mathrm{CM\text{-}FD}}$.

The amount of power consumed by the transmitter circuitry is [15]

$$P_{\mathrm{TX,c}} = N_{\mathrm{T}}^{\mathrm{RF}}(P_{\mathrm{RFC}} + P_{\mathrm{DAC}} + N_{\mathrm{T}}P_{\mathrm{PS}}) + N_{\mathrm{T}}P_{\mathrm{PA}} + P_{\mathrm{BB}}, \tag{10.19}$$

and the amount of power consumed by the receiver circuitry is

$$P_{\mathrm{RX,c}} = N_{\mathrm{R}}^{\mathrm{RF}}(P_{\mathrm{RFC}} + P_{\mathrm{ADC}} + N_{\mathrm{R}}P_{\mathrm{PS}}) + N_{\mathrm{R}}P_{\mathrm{LNA}} + P_{\mathrm{BB}}. \tag{10.20}$$

Numerical values for the earlier-mentioned quantities have already been given, except that for P_{PS}, the power consumed by each phase shifters, that is assumed to be 19.5 mW as in [52].

10.4.4 Partial zero-forcing, hybrid (PZF-HY) beamforming

Similarly to what has been described in the previous subsection, also the PZF beamformers may be approximated through HY architectures. In this case, expressions for the matrices $\mathbf{Q}_k^{\mathrm{RF}}$, $\mathbf{Q}_k^{\mathrm{BB}}$, $\mathbf{D}_k^{\mathrm{RF}}$, and $\mathbf{D}_k^{\mathrm{BB}}$ are to be found, so that the PZF-FD beamforming

---

Algorithm 10.1: Block coordinate descent for subspace decomposition algorithm for hybrid beamforming

---

1: Initialize $I_{\max}$ and set $i = 0$
2: Set arbitrary $\mathbf{Q}_{\mathrm{RF},0}$ and $\mathbf{D}_{\mathrm{RF},0}$
3: **repeat**
4: Update $\mathbf{Q}_{\mathrm{BB,i+1}} = \left(\mathbf{Q}_{\mathrm{RF,i}}^H\mathbf{Q}_{\mathrm{RF,i}}\right)^{-1}\mathbf{Q}_{\mathrm{RF,i}}^H\mathbf{Q}^{\mathrm{opt}}$
and $\mathbf{D}_{\mathrm{BB,i+1}} = \left(\mathbf{D}_{\mathrm{RF,i}}^H\mathbf{D}_{\mathrm{RF,i}}\right)^{-1}\mathbf{D}_{\mathrm{RF,i}}^H\mathbf{D}^{\mathrm{opt}}$
5: Set $\phi_i = \mathbf{Q}^{\mathrm{opt}}\mathbf{Q}_{\mathrm{BB,i+1}}^H\left(\mathbf{Q}_{\mathrm{BB,i+1}}\mathbf{Q}_{\mathrm{BB,i+1}}^H\right)^{-1}$
and $\psi_i = \mathbf{D}^{\mathrm{opt}}\mathbf{D}_{\mathrm{BB,i+1}}^H\left(\mathbf{D}_{\mathrm{BB,i+1}}\mathbf{D}_{\mathrm{BB,i+1}}^H\right)^{-1}$
6: Update $\mathbf{Q}_{\mathrm{RF,i}} = \frac{1}{\sqrt{N_{\mathrm{T}}}}e^{j\phi_i}$
and $\mathbf{D}_{\mathrm{RF,i}} = \frac{1}{\sqrt{N_{\mathrm{R}}}}e^{j\psi_i}$
7: Set $i = i + 1$
8: **until** convergence or $i = I_{\max}$

---

matrices are approximated as closely as possible. Also in this case, the BCD-SD algorithm can be used and the hybrid beamformers are evaluated following Algorithm 10.1, with $\mathbf{Q}^{\text{opt}} = \mathbf{Q}^{\text{PZF-FD}}$ and $\mathbf{D}^{\text{opt}} = \mathbf{D}^{\text{PZF-FD}}$.

The amount of power consumed by the transmitter circuitry of the PZF-HY beamformers is the same as that consumed by the CM-HY ones.

10.4.5 Fully analog beam-steering beamforming (AB)

Fully AB requires that the entries of the precoding and postcoding matrices have a constant norm, the block scheme of a transceiver with AB and M RF chains at both sides of the communication link is reported in Figure 10.3. Here, it is considered an even simpler structure by introducing a further constraint and assuming that the columns of matrices $\mathbf{Q}_k$ and $\mathbf{D}_k$ are unit-norm beam-steering vectors, i.e., the generic column of an N-dimensional beamformer is

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}}[1 \; e^{-jkd\sin\phi} \; \cdots \; e^{-jkd(N-1)\sin\phi}]. \tag{10.21}$$

Focusing on the generic k-th user, the columns of the matrix $\mathbf{Q}_k^{\text{AB}}$ are chosen as the array responses corresponding to the departure angles in the channel model (10.1) associated with the M-dominant paths. A similar choice is made for $\mathbf{D}_k^{\text{AB}}$, whose columns contain the array responses corresponding to the M arrival angles associated with the M-dominant paths. In order to avoid self-interference, a further constraint

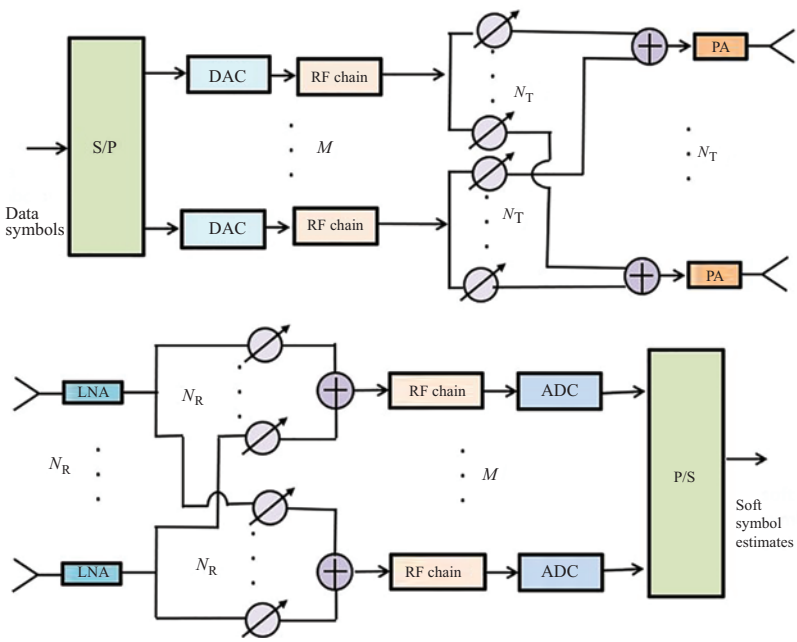


Figure 10.3 Block scheme of a transceiver with AB

is added in the choice of the dominant paths to ensure that the angles of departure (arrival) of the selected paths are spaced of at least 5 degrees. Note that for large values of N_{T} and N_{R} the array responses of the transmitter and receiver, corresponding to the departure and arrival angles associated with the dominant propagation path, become coincident with dominant right and left singular vectors of the channel. This implies that the AB beamforming structure (10.21) tends to become optimal. The amount of power consumed by the transmitter circuitry is

$$P_{\mathrm{TX,c}} = N_{\mathrm{T}}^{\mathrm{RF}}(P_{\mathrm{RFC}} + N_{\mathrm{T}}P_{\mathrm{element}} + P_{\mathrm{DAC}}), \tag{10.22}$$

and the amount of power consumed by the receiver circuitry is

$$P_{\mathrm{RX,c}} = N_{\mathrm{R}}^{\mathrm{RF}}(P_{\mathrm{RFC}} + N_{\mathrm{R}}P_{\mathrm{element}} + P_{\mathrm{ADC}}), \tag{10.23}$$

where $P_{\mathrm{element}} = 27$ mW [47] is the power consumed by each element of the phased array.

10.5 Asymptotic SE analysis

10.5.1 CM-FD beamforming

In the large number of antennas regime, making the assumption that the set of arrival and departure angles across clusters and users are different with probability 1, it readily follows from the SVD expression of the channel that $\mathbf{D}_k^H\mathbf{H}_k\mathbf{Q}_\ell \to \mathbf{\Lambda}_{k,M}\mathbf{V}_{k,M}\mathbf{Q}_\ell$, whenever $k \neq \ell$, where $\mathbf{\Lambda}_{k,M}$ is an $(M \times M)$-dimensional diagonal matrix containing the M largest eigenvalues (denoted by $\lambda_{k,1}, \ldots, \lambda_{k,M}$) of the channel matrix $\mathbf{H}_k$ and $\mathbf{V}_{k,M}$ is an $(N_{\mathrm{T}} \times M)$-dimensional matrix containing the columns of $\mathbf{V}_k$ associated with the eigenvalues in $\mathbf{\Lambda}_{k,M}$. Using the earlier-mentioned limiting values, the asymptotic SE in (10.11) can be expressed as

$$\mathrm{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \left(\sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{\Lambda}_{k,M}\mathbf{V}_{k,M}^H\mathbf{Q}_\ell\mathbf{P}_\ell\mathbf{Q}_\ell^H\mathbf{V}_{k,M}\mathbf{\Lambda}_{k,M}^H \right)^{-1} \mathbf{\Lambda}_{k,M}\mathbf{P}_k\mathbf{\Lambda}_{k,M}^H \right|. \tag{10.24}$$

10.5.2 PZF-FD beamforming

For PZF-FD beamforming, using a similar approach as in (10.24) and the definition of the PZF-FD in (10.17), we can note that the product $\mathbf{V}_{k,M}^H\mathbf{Q}_\ell$ is an all-zero matrix whenever $k \neq \ell$. As a consequence, $\mathbf{R}_{\mathrm{D},k} \sim \sigma_n^2\mathbf{I}_M$ and the asymptotic SE can be shown to be written as[††]

$$\mathrm{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \frac{1}{\sigma_n^2}\mathbf{\Lambda}_{k,M}\mathbf{P}_k\mathbf{\Lambda}_{k,M}^H \right|. \tag{10.25}$$

[††]This is an asymptotic expression since the noise enhancement effect is neglected (i.e., a decreasing function of N_{T}) induced by the nulling of the interference.

In order to explicitly show the dependence of (10.25) on the number of antennas, note that the squared moduli of the eigenvalues $\lambda_{k,i}$ depend linearly on the product $N_{\mathrm{T}}N_{\mathrm{R}}$. Otherwise stated, the following holds

$$\lambda_{k,q} = \sqrt{N_{\mathrm{T}}N_{\mathrm{R}}}\,\widetilde{\lambda}_{k,q}, \quad \forall\, k = 1,\ldots,K, \;\; q = 1,\ldots,M, \tag{10.26}$$

with $\widetilde{\lambda}_{k,q}$ normalized eigenvalues independent of the number of transmit and receive antennas. Using this definition, (10.25) can be written as

$$\mathrm{SE} \sim \sum_{k=1}^{K}\sum_{q=1}^{M} \log_2\left(1 + N_{\mathrm{T}}N_{\mathrm{R}}p_{k,q}\frac{|\widetilde{\lambda}_{k,q}|^2}{\sigma_n^2}\right). \tag{10.27}$$

10.5.3 Analog beamforming

The case of AB precoding and postcoding is now considered. As a preliminary step to the analysis, it is convenient to recall that the ULA response in (10.21) is a unit-norm vector, and that the inner product between two ULA responses of length P and corresponding to incidence angles ϕ_1 and ϕ_2 is written as

$$f_P(\phi_1,\phi_2) \triangleq \mathbf{a}^H(\phi_1)\mathbf{a}(\phi_2) = \frac{1}{P}\frac{1-e^{jkd(\sin\phi_1-\sin\phi_2)P}}{1-e^{jkd(\sin\phi_1-\sin\phi_2)}}. \tag{10.28}$$

The earlier-mentioned inner product that is denoted by $f_P(\phi_1,\phi_2)$ has a magnitude that, for large P, vanishes as $1/P$, whenever $\phi_1 \neq \phi_2$. Let us now write the channel matrix for k-th user as

$$\mathbf{H}_k = \gamma_k \sum_{i=1}^{N} \alpha_{k,i}\mathbf{a}_{\mathrm{R}}(\phi_{i,k}^{\mathrm{R}})\mathbf{a}_{\mathrm{T}}^H(\phi_{i,k}^{\mathrm{T}}) = \gamma_k \mathbf{A}_{k,\mathrm{R}}\mathbf{L}_k\mathbf{A}_{k,\mathrm{T}}^H, \tag{10.29}$$

namely, the path-loss term has been lumped into the coefficients $\alpha_{\cdot,\cdot}$, and the summation over the clusters and the rays has been compressed in just one summation, with $N = N_{\mathrm{cl}}N_{\mathrm{ray}}$. Additionally, $\mathbf{A}_{k,\mathrm{R}}$ is an $(N_{\mathrm{R}} \times N)$-dimensional matrix containing on its columns the vectors $\mathbf{a}_{\mathrm{R}}(\phi_{1,k}^{\mathrm{R}}),\ldots,\mathbf{a}_{\mathrm{R}}(\phi_{N,k}^{\mathrm{R}})$, $\mathbf{L}_k = \mathrm{diag}(\alpha_{1,k},\ldots,\alpha_{N,k})$, and $\mathbf{A}_{k,\mathrm{T}}$ is an $(N_{\mathrm{T}} \times N)$-dimensional matrix containing on its columns the vectors $\mathbf{a}_{\mathrm{T}}(\phi_{1,k}^{\mathrm{T}}),\ldots,\mathbf{a}_{\mathrm{T}}(\phi_{N,k}^{\mathrm{T}})$[‡‡]. It is also assumed, with no loss of generality, that the paths are sorted in decreasing magnitude order, i.e., $|\alpha_{1,k}| \geq |\alpha_{2,k}| \geq \cdots \geq |\alpha_{N,k}|$. In the following analysis, it is assumed that there are no collisions between arrival and departure angles across users, an assumption that is usually verified unless there are very close users.

The analog postcoding and precoding matrices are written as

$$\begin{aligned}\mathbf{D}_k &= [\mathbf{a}_{\mathrm{R}}(\phi_{1,k}^{\mathrm{R}}),\ldots,\mathbf{a}_{\mathrm{R}}(\phi_{M,k}^{\mathrm{R}})]\,,\\ \mathbf{Q}_k &= [\mathbf{a}_{\mathrm{T}}(\phi_{1,k}^{\mathrm{T}}),\ldots,\mathbf{a}_{\mathrm{T}}(\phi_{M,k}^{\mathrm{T}})]\,,\end{aligned} \tag{10.30}$$

[‡‡]In order to avoid a heavy notation, it is here dropped the dependence of the matrices $\mathbf{A}_{k,\mathrm{R}}$ and $\mathbf{A}_{k,\mathrm{T}}$ on the propagation paths arrival and departure angles, respectively.

$\forall k$, and they are actually submatrices of $\mathbf{A}_{k,\mathrm{R}}$ and $\mathbf{A}_{k,\mathrm{T}}$, respectively. Define now the following $(M \times N)$-dimensional matrices: $\mathbf{F}_{k,\ell,\mathrm{R}} \triangleq \mathbf{D}_k^H \mathbf{A}_{\ell,\mathrm{R}}$ and $\mathbf{F}_{k,\ell,\mathrm{T}} \triangleq \mathbf{Q}_k^H \mathbf{A}_{\ell,\mathrm{T}}$. Note that the (m,n)th entry of the matrix $\mathbf{F}_{k,\ell,\mathrm{T}}$ is $f_{N_\mathrm{T}}\left(\phi_{m,k}^{\mathrm{T}}, \phi_{n,\ell}^{\mathrm{T}}\right)$, while the (m,n)th entry of the matrix $\mathbf{F}_{k,\ell,\mathrm{R}}$ is $f_{N_\mathrm{R}}\left(\phi_{m,k}^{\mathrm{R}}, \phi_{n,\ell}^{\mathrm{R}}\right)$. Equipped with this notation, the SE in (10.11) can be now expressed as follows:

$$\mathrm{SE} = \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \gamma_k^2 \mathbf{R}_{\mathrm{D},k}^{-1} \mathbf{F}_{k,k,\mathrm{R}} \mathbf{L}_k \mathbf{F}_{k,k,\mathrm{T}}^H \mathbf{P}_k \mathbf{F}_{k,k,\mathrm{T}} \mathbf{L}_k^H \mathbf{F}_{k,k,\mathrm{R}}^H \right|, \tag{10.31}$$

with

$$\mathbf{R}_{\mathrm{D},k} = \sigma_n^2 \mathbf{D}_k^H \mathbf{D}_k + \gamma_k^2 \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{F}_{k,k,\mathrm{R}} \mathbf{L}_k \mathbf{F}_{\ell,k,\mathrm{T}}^H \mathbf{P}_\ell \mathbf{F}_{\ell,k,\mathrm{T}} \mathbf{L}_k^H \mathbf{F}_{k,k,\mathrm{R}}^H. \tag{10.32}$$

In order to have an asymptotic expression of (10.31) for a large number of antennas, it can be noted that the $(M \times N)$-dimensional matrix $\mathbf{F}_{k,\ell,\mathrm{R}}$ is such that (a) for $k \neq \ell$ all its entries have a norm that for large N_R vanishes as $1/N_\mathrm{R}$, while (b) for $k = \ell$ the M entries on the main diagonal are equal to 1 while all the remaining terms again vanish in norm as $1/N_\mathrm{R}$. A similar statement also applies to the matrix $\mathbf{F}_{k,\ell,\mathrm{T}}$, of course with entries vanishing as $1/N_\mathrm{T}$. Accordingly, the following asymptotic formulas can be proven.

(a) $N_\mathrm{T} \to +\infty$, **finite** N_R: in this case the system becomes free from the interference between different users and it is obtained

$$\mathrm{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \frac{\gamma_k^2}{\sigma_n^2} \left(\mathbf{D}_k^H \mathbf{D}_k \right)^{-1} \mathbf{F}_{k,k,\mathrm{R}} \mathbf{L}_k \widetilde{\mathbf{P}}_{k,N} \mathbf{L}_k^H \mathbf{F}_{k,k,\mathrm{R}}^H \right|, \tag{10.33}$$

where $\widetilde{\mathbf{P}}_{k,N}$ is a N-dimensional diagonal matrix defined as

$$\widetilde{\mathbf{P}}_{k,N} = \mathrm{diag}\left(p_{k,1}, \ldots, p_{k,M}, \underbrace{0, \ldots, 0}_{N-M} \right). \tag{10.34}$$

Note that in this case there is interference between different streams intended for the same user.

(b) $N_\mathrm{R} \to +\infty$, **finite** N_T: it holds now

$$\mathrm{SE} \sim \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \gamma_k^2 \mathbf{R}_{\mathrm{D},k}^{-1} \left(\mathbf{L}_k \mathbf{F}_{k,k,\mathrm{T}}^H \mathbf{P}_k \mathbf{F}_{k,k,\mathrm{T}} \mathbf{L}_k^H \right)_{(1:M,1:M)} \right|, \tag{10.35}$$

$$\text{with} \quad \mathbf{R}_{\mathrm{D},k} = \sigma_n^2 \mathbf{I}_M + \gamma_k^2 \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \left(\mathbf{L}_k \mathbf{F}_{\ell,k,\mathrm{T}}^H \mathbf{P}_\ell \mathbf{F}_{\ell,k,\mathrm{T}} \mathbf{L}_k^H \right)_{(1:M,1:M)}.$$

(c) $N_\mathrm{R}, N_\mathrm{T} \to \infty$: finally it holds

$$\mathrm{SE} \sim \sum_{k=1}^{K} \sum_{q=1}^{M} \log_2 \left(1 + p_{k,q} \frac{\gamma_k^2 |\alpha_{k,q}|^2}{\sigma_n^2} \right). \tag{10.36}$$

It is easily seen that the earlier-mentioned expression coincides with (10.25), i.e., the system becomes free from the interference between different users and streams.

10.6 EE maximizing power allocation

In this section, the problem of the downlink power allocation maximizing the asymptotic EE is presented. In particular, two different situations are separately analyzed: the interference-free case and the interference-limited case. The former problem is addressed in order to obtain a simple and low-complexity solution, and then the latter problem is discussed and solved using an alternating optimization technique.

10.6.1 Interference-free case

In the interference-free case, the SE of one user is not corrupted by other co-channel communications. From the asymptotic SE analysis, we can note that (10.27) and (10.36) fall in this category. We thus focus on the following optimization problem:

$$\begin{cases} \max\limits_{\mathbf{p}} \dfrac{\sum_{k=1}^{K}\sum_{q=1}^{M} W\log_2\left(1+p_{k,q}g_{k,q}\right)}{\eta\sum_{k=1}^{K}\sum_{q=1}^{M} p_{k,q}+P_{\mathrm{TX,c}}+KP_{\mathrm{RX,c}}} \\ \text{s.t.} \sum\limits_{k=1}^{K}\sum\limits_{q=1}^{K} p_{k,q} \le P_{\max} \\ \quad p_{k,q} \ge 0\ ,\forall\ k=1,\ldots,K,\ q=1,\ldots,M, \end{cases} \tag{10.37}$$

where $\mathbf{p}=\left[p_{1,1},\ldots,p_{1,M},\ldots,p_{K,1},\ldots,p_{K,M}\right]^T$. Notice that, considering (10.27), i.e., PZF-FD beamforming, $g_{k,q}=N_{\mathrm{T}}N_{\mathrm{R}}\frac{|\tilde{\lambda}_{k,q}|^2}{\sigma_n^2}$, and considering (10.36), i.e., AB, $g_{k,q}=\frac{\gamma_k^2|\alpha_{k,q}|^2}{\sigma_n^2}$. The objective function in Problem (10.37) is the ratio of a concave function (with respect to the optimization variables) over a linear one, and, thus, Dinkelbach's algorithm in [54] may be readily applied to maximize the ratio [21,55]. Denoting as $N(\mathbf{p})$ and $D(\mathbf{p})$ as

$$N(\mathbf{p})=\sum_{k=1}^{K}\sum_{q=1}^{M} W\log_2\left(1+p_{k,q}g_{k,q}\right), \tag{10.38}$$

$$D(\mathbf{p})=\eta\sum_{k=1}^{K}\sum_{q=1}^{M} p_{k,q}+P_{\mathrm{TX,c}}+KP_{\mathrm{RX,c}}, \tag{10.39}$$

the Dinkelbach's procedure can be summarized as in Algorithm 10.2.

Given the structure of the objective in Problem (10.37), we can note that standard techniques can be used to solve the concave maximization in Problem (10.40). The

Algorithm 10.2: Dinkelbach's procedure [54]

1: Set $\varepsilon = 0, \pi = 0$ and FLAG $= 0$
2: **repeat**
3: Update $\mathbf{p}$ by solving the following concave maximization:

$$\max_{\mathbf{p}} \; N(\mathbf{p}) - \pi D(\mathbf{p}) \tag{10.40}$$

4: **if** $N(\mathbf{p}) - \pi D(\mathbf{p}) < \varepsilon$ **then**
5: FLAG $= 1$
6: **else**
7: Set $\pi = \dfrac{N(\mathbf{p})}{D(\mathbf{p})}$
8: **end if**
9: **until** FLAG $= 1$

solution of the concave maximization in Problem (10.40) can be computed starting from the following KKT conditions [56]:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}p_{k,q}} \left(N(\mathbf{p}) - \pi D(\mathbf{p})\right) - \lambda + \mu_{k,q} = 0, \quad \forall k, q \\
&\sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \leq P_{\max} \\
&\lambda \left(P_{\max} - \sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \right) = 0, \\
&\lambda >= 0 \\
&p_{k,q} \geq 0, \quad \forall k, q \\
&\mu_{k,q} p_{k,q} = 0, \; \forall k, q \\
&\mu_{k,q} \geq 0, \quad \forall k, q
\end{aligned} \tag{10.41}$$

where λ and $\mu_{k,q}$ are the Lagrange multipliers associated with the constraints on the maximum power radiated by the transmitter and on the minimum power level of the transmitter to the k-th user on the q-th streams, respectively. After standard algebraic manipulations, we obtain the following waterfilling-like problem:

$$\begin{cases} p_{k,q} = \max \left\{ 0, \dfrac{W/\ln 2}{\pi \eta + \lambda} - \dfrac{1}{g_{k,q}} \right\} \\ \displaystyle\sum_{k=1}^{K} \sum_{q=1}^{K} p_{k,q} \leq P_{\max}, \end{cases} \tag{10.42}$$

and the optimal value of the nonnegative Lagrange multiplier λ can be derived by bisection search. The optimality and convergence conditions of the algorithm to solve Problem (10.37) follow from the ones of Dinkelbach's procedure in Algorithm 10.2, see [21,54], further details are omitted for the sake of brevity.

10.6.2 Interference-limited case

The interference-limited case includes the asymptotic SE expressions in (10.24), (10.33), and (10.35). We can rewrite these expressions in the following common form:

$$\begin{aligned}\mathrm{SE} &= \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_M + \left(\sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right)^{-1} \mathbf{A}_{k,k} \mathbf{P}_k \mathbf{A}_{k,k}^H \right| \\ &= \sum_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|, \end{aligned} \tag{10.43}$$

where $\mathbf{A}_{k,\ell}$, $\forall k, \ell$ are properly defined matrices. Given the SE expression in (10.43), the EE maximizing power allocation is obtained by solving the following optimization problem

$$\begin{cases} \max\limits_{\mathbf{P}_1, \ldots \mathbf{P}_K} & \dfrac{\sum_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|}{\eta \sum_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} \\ \text{s.t.} & \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) \leq P_{\max} \\ & p_{k,q} \geq 0 \ , \ \forall \ k = 1, \ldots, K, \ q = 1, \ldots, M, \end{cases} \tag{10.44}$$

Problem (10.44) has nonconcave objective function, which makes its solution challenging. Additionally, the large number of optimization variables would still pose a significant complexity challenge. To solve (10.44) the framework of alternating optimization can be used, see [57, Section 2.7] for the details. In particular, each subproblem, one for each user, can be written as

$$\begin{cases} \max\limits_{\mathbf{P}_k} & \dfrac{\sum_{k=1}^{K} \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\ell=1}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right| - \log_2 \left| \sigma_n^2 \mathbf{I}_M + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{A}_{k,\ell} \mathbf{P}_\ell \mathbf{A}_{k,\ell}^H \right|}{\eta \sum_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) + P_{\mathrm{TX,c}} + K P_{\mathrm{RX,c}}} \\ \text{s.t.} & \sum\limits_{k=1}^{K} \mathrm{tr}(\mathbf{P}_k) \leq P_{\max} \\ & p_{k,q} \geq 0 \ , \forall \ k = 1, \ldots, K, \ q = 1, \ldots, M. \end{cases} \tag{10.45}$$

Since the logarithm of the identity plus an Hermitian positive semidefinite matrix is a matrix-concave function [58], the numerator of the objective function in (10.45) is concave in $\mathbf{P}_k$. Regarding the denominator, the sum of the traces of $\mathbf{P}_k$ $\forall k$ is a linear function and consequently Problem can be solved with affordable complexity by means of Dinkelbach's procedure [21] in Algorithm 10.2.

10.7 Numerical results

In the simulation setup, a communication bandwidth of $W = 1$ GHz centered over the carrier frequency $f_c = 28$ GHz is considered. We assume a single-cell doubly massive MIMO system with $K = 10$ users. We consider two different scenarios: the former is an outdoor scenario, denoted as "scenario 1" and the latter is an indoor scenario, denoted as "scenario 2." In scenario 1, it is considered the downlink communication between one BS connected via wired backhaul to the core network, with K BSs served via wireless backhaul. In scenario 2, the downlink communication between an indoor BS with K robots in a shopping mall is assumed. A qualitative representation of the considered scenarios is reported in Figure 10.4. The parameters for the generation of the matrix channel is the ones reported in Section 10.3.1 for the "UMi Street Canyon NLOS" in scenario 1 and for "InH Shopping Mall NLOS" in scenario 2. The noise power $\sigma_n^2 = FN_0W$, with $F = 5$ dB the receiver noise figure and $N_0 = -174$ dBm/Hz. A detailed description of the simulation parameters is reported in Table 10.3. The results shown come from an average over 500 independent realizations of users' locations and propagation channels.

Figure 10.4 *In (a) scenario 1: outdoor street canyon environment, and in (b) scenario 2: indoor shopping mall environment*

Table 10.3 System parameters

Description	**Value**
BS position (scenario 1)	Horizontal: origin of the reference system. Vertical: 10 m
BS position (scenario 2)	Horizontal: origin of the reference system. Vertical: 5 m
Users distribution (scenario 1)	Horizontal: uniform, azimuth in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$, fixed distance at 100 m. Vertical: 10 m
Users distribution (scenario 2)	Horizontal: uniform, azimuth in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$ degree, distance in [5, 100] m. Vertical: 1.65 m
Carrier frequency	$f_c = 28$ GHz
Bandwidth	$W = 1$ GHz
BS antenna array	ULA with $\lambda/2$ spacing
User antennas	ULA with $\lambda/2$ spacing
N_0	-174 dBm/Hz
Noise figure	5 dB

We start considering the uniform power allocation (Uni), i.e., the M-dimensional diagonal matrix containing the power allocation is

$$\mathbf{P}_\ell = \text{diag}\left(\frac{P_T}{KM}, \ldots, \frac{P_T}{KM}\right) \quad \forall \ell = 1, \ldots, K, \tag{10.46}$$

where P_T is the total transmit power. In all the presented results, the hybrid and the analog beamformers have been realized using a number of RF chains equal to KM at the transmitter and equal to M at the receiver, i.e., $N_{\text{T}}^{\text{RF}} = KM$ and $N_{\text{R}}^{\text{RF}} = M$.

Figures 10.5 and 10.6 report the downlink system SE and EE versus the transmit power, for the case of multiplexing order $M = 4$ and uniform power allocation for the two scenarios in Figure 10.4. In scenario 1, $N_{\text{T}} = N_{\text{R}} = 64$, a symmetrical communication is considered, while in scenario 2, $N_{\text{T}} = 64$ and $N_{\text{R}} = 16$, an asymmetrical one is assumed. This choice is based on the use-cases associated with the considered scenarios: in scenario 1, the transmitter and the receivers are BSs that communicate for wireless backhaul while in scenario 2 the communication is between an indoor BS and mobile users, in particular robots, so it reasonable to assume that the number of antennas at the users is lower with respect to the one at the transmitter. Inspecting the figures, in both scenarios, it is seen that the best performing beamforming structure is the PZF-FD, both in terms of SE and of EE. This means that the increase in system SE given by the PZF-FD beamforming structure, which nulls the interference and gives good performance in terms of simultaneously transmission, is not well compensated by the reduction of the power consumption in the PZF-HY structure. In particular in Figure 10.5, we can see that the performance in terms of SE of the CM-FD, CM-HY, PZF-HY, and AB are very close, while the performances of the PZF-FD are considerably better; similar results are presented also in [17,20,59]. Focusing on the EE performance, we can see that the reduction of the power consumption in the hybrid and analog structures gives an increase in terms of EE, while the reduction of power consumption in the PZF-HY structure cannot compensate the gap in terms of SE with

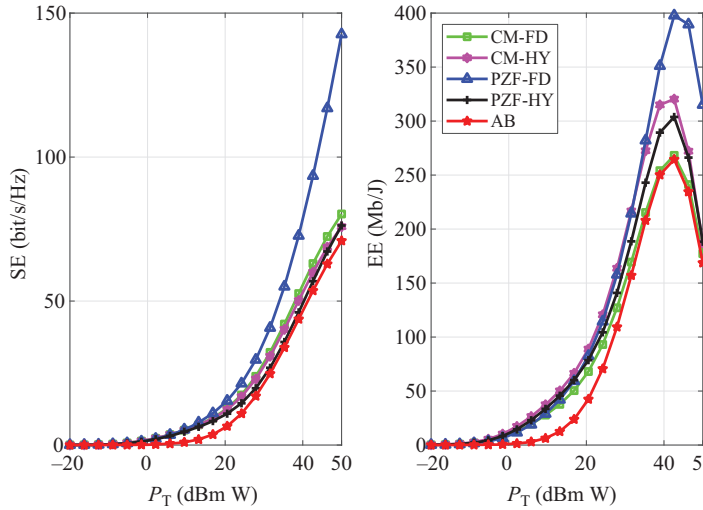


Figure 10.5 *SE and EE versus transmit power in scenario 1. Parameters: Uni, $N_T = 64$, $N_R = 64$, $K = 10$, and M*

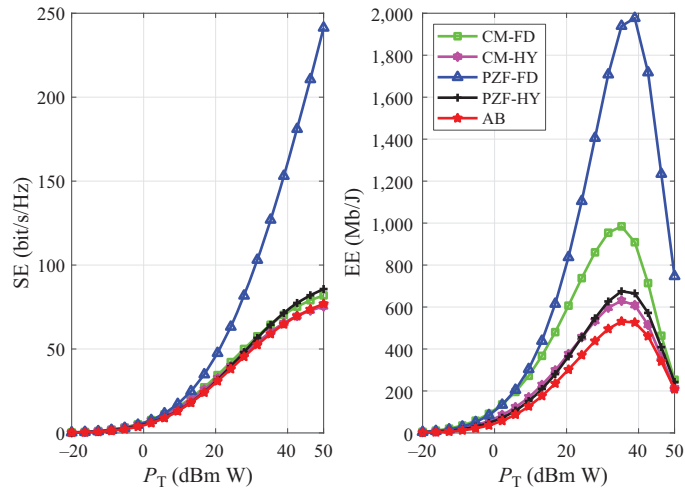


Figure 10.6 *SE and EE versus transmit power in scenario 2. Parameters: Uni, $N_T = 64$, $N_R = 16$, $K = 10$, and $M = 4$*

the PZF-FD beamforming. Results show here a trend that has already been found elsewhere (e.g., in [55]); in particular, while the SE grows with the transmit power (at least in the considered range of values), the EE exhibits instead a maximum around 40 dBm W in both scenarios. This behavior is explained by the fact that for large

values of the transmit power, the numerator in the EE grow; at a slower rate than the denominator of the EE, and so the EE itself decreases. From an energy-efficient perspective, increasing the transmit power beyond the EE-optimal point leads to moderate improvements in the system throughput at the price of a much higher increase in the consumed power. Additionally, we can see that there is a region where both SE and EE grow, similar results can also be observed elsewhere in the literature, as an example the reader can be referred to [22].

Figure 10.7 is devoted to the validation of the derived asymptotic formulas in the large number of antennas regime for scenario 1 in Figure 10.4. In particular, "CM-FD, Asy" refers to (10.24), "PZF-FD, Asy" refers to (10.27) and finally "AB, Asy" refers to (10.33) for increasing N_T and to (10.35) for increasing N_R. Similar results are also obtained in scenario 2 and they are here omitted due to the lack of space. Results fully confirm the effectiveness of the found asymptotic formulas, for both the single- and multiple stream cases. Given the effectiveness of the asymptotic formulas in terms of SE, we now consider the performance of the power allocation based on these expressions and detailed in Section 10.6.

Figures 10.8 and 10.9 report the performance of the EE-maximizing power allocation detailed in Section 10.6 for the case of PZF-FD and AB beamforming. We report the performance for the two antenna configurations in both the scenarios and we denote by "Opt" the power allocation obtained with the procedure in Section 10.6.1. Inspecting the figures, we can see that for low values of P_{max} the Uni and

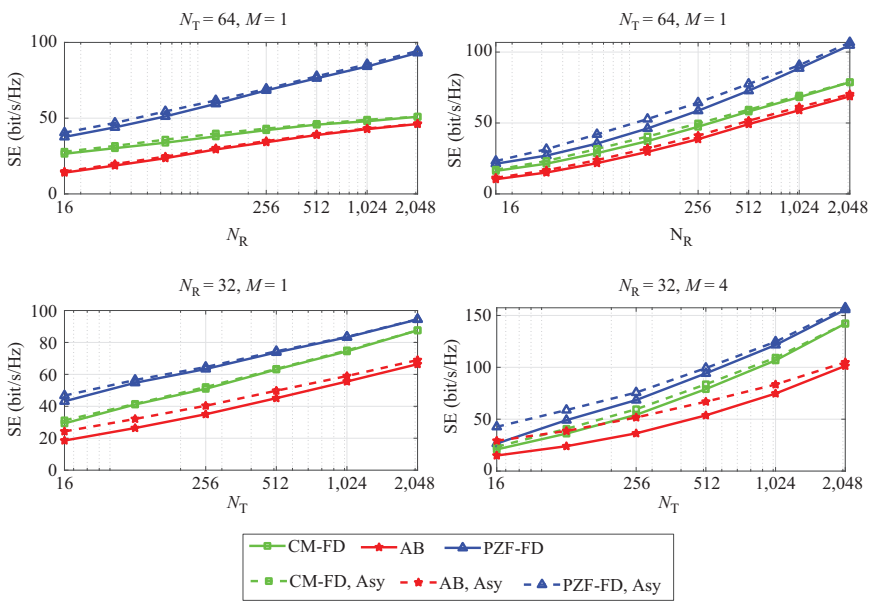


Figure 10.7 SE versus number of antennas in scenario 1, Uni. Validation of the derived asymptotic SE formulas

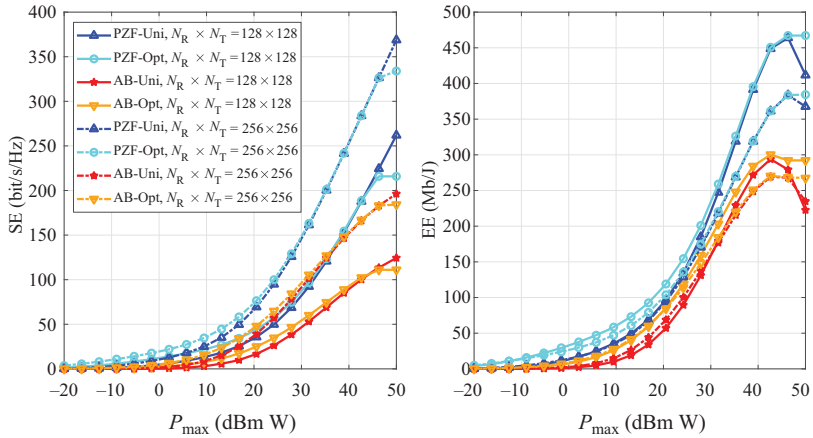


Figure 10.8 SE and EE versus P_{max}, comparison between EE-maximizing and Uni in scenario 1 with $M = 4$ and two antenna configurations

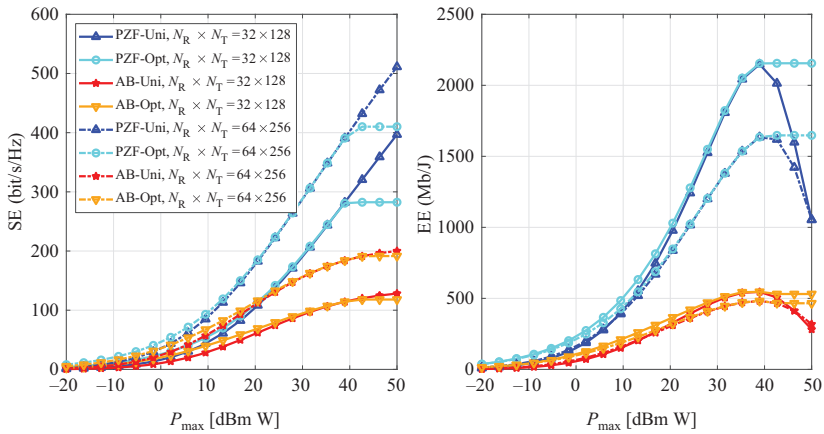


Figure 10.9 SE and EE versus P_{max}, comparison between EE-maximizing and Uni in scenario 2 with $M = 4$ and two antenna configurations

the EE-maximizing power allocation provide similar performance, both in terms of SE and EE, since the radiated power consumption and also the co-channel interference is small compared to the noise power. For larger values of $P_{\max}$, around 40 dBm W in all the results shown, instead, the two power allocations lead to different performance. In this regime, the Uni increases the SE at the price of heavy degradation in the system EE. On the other hand, the EE-maximizing power allocation exhibits a floor as $P_{\max}$ increases, since it does not use the excess available power to further increase the rate.

10.8 Conclusions

This chapter has focused on the analysis of doubly massive MIMO mm-wave systems, and it has presented some relevant use cases, focusing on EE issues. In particular, a detailed description of doubly massive MIMO system has been given highlighting the differences with respect to massive MIMO at microwave. Some examples of use-cases of doubly massive MIMO systems have been detailed. Asymptotic formulas for the large number of antennas regime and low-complexity EE-maximizing power allocation strategies have been derived. In particular, two different techniques are described, the first one used the asymptotic expressions in the interference-free case, the second one used the expressions with interference. In both cases, the EE-maximizing power allocation is obtained by means of fractional programing techniques. The obtained results have revealed that, using some of the most recent available data on the energy consumption of transceiver components, FD architectures, especially in the multiple-streams transmission, are superior not only in terms of achievable rate but also in terms of EE. In particular, among FD implementations, the PZF architecture has been shown to provide the best performance, while the AB structure can be considered for its extremely low complexity. Of course the provided results and the relative ranking among the considered structures in terms of EE are likely to change in the future as technology progresses and devices with reduced power consumption appear on the scene, even though it may be expected that in the long run FD architectures will be fully competitive, in terms of hardware complexity and energy consumption [17,20,25], with hybrid alternatives.

Acknowledgments

Their work has been supported by Ministero dell'Istruzione dell'Universita' e della Ricerca (MIUR) through the program "Dipartimenti di Eccellenza 2018-2022" and through the Progetti di Ricerca di Interesse Nazionale (PRIN) 2017 "LiquidEDGE" project.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, *et al.*, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] 3GPP, "Study on new radio (NR) access technology physical layer aspects," 3GPP, Tech. Rep. TR. 38.802, Mar. 2017.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[4] T. S. Rappaport, S. Sun, R. Mayzus, *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, 2014.

[6] S. Buzzi and C. D'Andrea, "Doubly massive mmWave MIMO systems: Using very large antenna arrays at both transmitter and receiver," in *Proc. of 2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–6.

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[8] X. Ge, R. Zi, H. Wang, J. Zhang, and M. Jo, "Multi-user massive MIMO communication systems based on irregular antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5287–5301, 2016.

[9] S. Buzzi, "Doubly-massive MIMO systems at mmWave frequencies: Opportunities and research challenges," in *IEEE WCNC'2016 Workshop on Green and Sustainable 5G Wireless Networks*, Doha, Qatar, Apr. 2016, keynote talk.

[10] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[11] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801–1818, 2016.

[12] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[13] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?," in *2015 Information Theory and Applications Workshop (ITA)*, Feb. 2015, pp. 90–97.

[14] A. Alkhateeb, Y.-H. Nam, J. Zhang, and R. W. Heath, "Massive MIMO combining with switches," *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 232–235, 2016.

[15] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[16] R. Zi, X. Ge, J. Thompson, C. X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 758–771, 2016.

[17] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmWave systems," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 385–396, 2018.

[18] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmWave," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 756–770, 2020.

[19] O. Orhan, H. Nikopour, J. Nam, N. Naderializadeh, and S. Talwar, "A power efficient fully digital beamforming architecture for mmWave communications," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, Apr. 2019, pp. 1–6.

[20] A. Pizzo and L. Sanguinetti, "Optimal design of energy-efficient millimeter wave hybrid transceivers for wireless backhaul," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017, pp. 1–8.

[21] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3–4, pp. 185–396, 2015.

[22] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017.

[23] L. Sanguinetti, E. Bjornson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, 2020.

[24] S. Buzzi and C. D'Andrea, "Massive MIMO 5G cellular networks: mm-wave vs. μ-wave frequencies," *ZTE Communications*, vol. 15, no. S1, pp. 41–49, 2017, invited paper.

[25] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, 2019.

[26] S. A. Busari, M. A. Khan, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Millimetre-wave massive MIMO for cellular vehicle-to-infrastructure communication," *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 983–990, 2019.

[27] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in millimeter wave integrated access and backhaul for 5G," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8195–8210, 2018.

[28] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[29] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.

[30] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706–709, 2017.

[31] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 651–663, 2019.

[32] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44596–44612, 2019.

[33] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[34] S. Haghighatshoar and G. Caire, "Enhancing the estimation of mm-Wave large array channels by exploiting spatio-temporal correlation and sparse scattering," in *Proc. of WSA 2016; 20th International ITG Workshop on Smart Antennas*, Mar. 2016, pp. 1–7.

[35] S. Buzzi, C. D'Andrea, T. Foggi, A. Ugolini, and G. Colavolpe, "Single-carrier modulation versus OFDM for millimeter-wave wireless MIMO," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1335–1348, 2018.

[36] J. Lee, G. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 3326–3331.

[37] K. Venugopal, N. Gonzalez-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 330–333, 2017.

[38] Aalto University, BUPT, CMCC, Nokia, NTT DOCOMO, New York University, Ericsson, Qualcomm, Huawei, Samsung, Intel, University of Bristol, KT Corporation, University of Southern California, "5G channel model for bands up to 100 GHz," Dec. 2015, Tech. Rep. [Online]. Available: http://www.5gworkshops.com/5GCM.html.

[39] K. Haneda, J. Zhang, L. Tan, *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–7.

[40] S. Buzzi and C. D'Andrea, "On clustered statistical MIMO millimeter wave channel simulation," *arXiv preprint arXiv:1604.00648*, May 2016.

[41] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, 2016.

[42] F. Negro, S. P. Shenoy, I. Ghauri, and D. T. Slock, "On the MIMO interference channel," in *Proc. of Information Theory and Applications Workshop 2010*, pp. 1–9, Feb. 2010.

[43] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3258–3268, 2018.

[44] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, 2015.

[45] G. H. Golub and C. F. Van Loan, *Matrix Computations*. JHU Press, 2012, vol. 3.

[46] A. V. den Bosch, M. A. F. Borremans, M. S. J. Steyaert, and W. Sansen, "A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 315–323, 2001.

[47] L. Kong, "Energy-efficient 60GHz phased-array design for multi-Gb/s communication systems," Ph.D. dissertation, EEECS Department University of California, Berkeley, 2014.

[48] Y.-Y. Lee, C.-H. Wang, and Y.-H. Huang, "A hybrid RF/Baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 305–317, 2014.

[49] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, 2016.

[50] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimetre-wave MIMO systems," in *Proc. of 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, Jun. 2015, pp. 395–399.

[51] S. J. Wright, "Coordinate descent algorithms," *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, 2015.

[52] Y. Yu, P. G. M. Baltus, A. de Graauw, E. van der Heijden, C. S. Vaucher, and A. H. M. van Roermund, "A 60 GHz phase shifter integrated with LNA and PA in 65 nm CMOS for phased array systems," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1697–1709, 2010.

[53] D. Y. C. Lie, J. C. Mayeda, Y. Li, and J. Lopez, "A review of 5G power amplifier design at cm-wave and mm-wave frequencies," *Wireless Communications and Mobile Computing*, vol. 2018, 16 pages, 2018.

[54] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.

[55] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 1–14, 2015.

[56] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, 2004.

[57] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[58] E. Jorswieck and H. Boche, "Majorization and matrix-monotone functions in wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 6, pp. 553–701, 2007.

[59] S. Buzzi and C. D'Andrea, "Are mmWave low-complexity beamforming structures energy-efficient? analysis of the downlink MU-MIMO," in *Proc. of 2016 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.

Chapter 11

Energy-efficient methods for cloud radio access networks

Kien-Giang Nguyen[1,2], Quang-Doanh Vu[2], Le-Nam Tran[3] and Markku Juntti[2]

Cloud radio access network (C-RAN) is an evolutionary radio network architecture in which a cloud-computing-based baseband (BB) signal-processing unit is shared among distributed low-cost wireless access points. This architecture offers a number of significant improvements over the traditional RANs, including better network scalability, spectral, and energy efficiency. As such C-RAN has been identified as one of the enabling technologies for the next-generation mobile networks. This chapter focuses on examining the energy-efficient transmission strategies of the C-RAN for cellular systems. In particular, we present optimization algorithms for the problem of transmit beamforming designs maximizing the network energy efficiency. In general, the energy efficiency maximization in C-RANs inherits the difficulty of resource allocation optimizations in interference-limited networks, i.e., it is an intractable non-convex optimization problem. We first introduce a globally optimal method based on monotonic optimization (MO) to illustrate the optimal energy efficiency performance of the considered system. While the global optimization method requires extremely high computational effort and, thus, is not suitable for practical implementation, efficient optimization techniques achieving near-optimal performance are desirable in practice. To fulfill this gap, we present three low-complexity approaches based on the state-of-the-art local optimization framework, namely, successive convex approximation (SCA).

11.1 Introduction

Recent years witness the rapid evolution of wireless technologies toward the fifth-generation (5G) mobile networks to adapt the ever-growing demand of mobile data volumes [1,2]. From the perspective of RAN, novel network architecture, namely, *C-RAN*, has been emerging as a powerful candidate to be implemented in 5G and

[1]Nokia, Oulu, Finland
[2]Centre for Wireless Communications, University of Oulu, Oulu, Finland
[3]School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland

beyond [3–5]. In fact, C-RAN has been named by 5G-PPP initiative as a key technology for the 5G network architecture [6]. Moreover, the first C-RAN solutions for commercial 5G have already been developed and released by several companies, such as Nokia (with Nokia AirScale C-RAN) [7] or Ericsson [8]. Unlike the conventional RAN where both BB signal-processing and radio frequency (RF) functionalities are installed at base stations (BSs), BB units (BBUs) in the C-RAN are reintegrated at a central cloud-computing-based platform, referred to as BBU pool. The BSs only need to carry the radio interfaces that can be replaced by low-cost low-power access points called remote radio heads (RRHs). The BBU pools are connected to RRHs by high-speed fronthaul links that are used to exchange the BB and control signals. The data is conveyed through the fronthaul links mainly following either data-sharing or compression-based strategies. For the former, users' messages are directly forwarded from the BBU to a set of cooperative RRHs, where they are precoded/beamformed before being transmitted [9,10]. Alternatively, the users' messages are centrally precoded/beamformed at the BBU in the latter strategy; the processed signals are then compressed before being forwarded to RRHs [11,12].

The innovative architecture of the C-RAN brings a number of advantages over the conventional distributed RAN architecture. The centralized signal-processing mechanism at the cloud across the connected RRHs can provide an efficient allocation of radio and computing resources. This feature also leverages the advanced interference management techniques, such as multicell cooperative transmission, to improve the system capacity. In addition to spectrum efficiency, the C-RAN also provides substantial gain in the energy efficiency of mobile networks. In particular, RRHs with simplified architecture require much reduced amount of power consumption compared to the traditional BSs. Moreover, centralized resource management allows BBU pool controlling the connected RRHs for power-saving purposes, i.e., some RRHs can be switched off if needed to save power consumption, while others remain active to guarantee required users' quality-of-service (QoS). However, one of the critical bottlenecks for the deployment of the C-RAN in practice is the requirement of high-speed fronthaul links. Regardless of the recent development in fronthauling technologies, the capacity of fronthaul links is physically limited [13].

With the mentioned advantages, the C-RAN is currently identified as one of the key enablers to the deployment of essential technologies in 5G to deal with the increasing capacity demand, such as small-cell and/or heterogeneous dense networks [14]. In fact the use of low-power RRHs reduces the size of the cells. This allows to deploy more cells in the same covering geographic area of the traditional BSs. In addition, intercell interference, which constitutes the main limiting factor to the dense deployment, can be efficiently managed at the cloud by means of the centralized signal processing. Nevertheless, such scenarios raise a concern over the huge amount of power consumption required to operate a large number of electronic circuit elements involved. As a result, developing energy-efficient transmission strategies for the C-RAN is deemed important. This chapter presents optimization algorithms for the energy efficiency maximization problems in the C-RANs.

This chapter is organized as follows. In Section 11.2, we provide mathematical preliminaries of optimization techniques that lay the foundations for the development

of solutions for energy efficiency optimization in C-RANs. In Section 11.3, the system model of a C-RAN and the formulation of the energy efficiency maximization problem are described, followed by the application of the introduced optimization approaches in Section 11.2 to the design problem.

11.2 Energy efficiency optimization: mathematical preliminaries

A general energy efficiency maximization problem for wireless networks with data rate as the desired QoS metric can be expressed as

$$\underset{\mathbf{x}}{\text{maximize}} \quad \frac{f(\mathbf{x})}{g(\mathbf{x})} \tag{11.1a}$$

$$\text{subject to} \quad h_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m, \tag{11.1b}$$

where $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ represents the data rate of the system, $g(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is the corresponding total power consumption, and $h_i(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ defines the design constrains, i.e., it can be a function of transmit power of a transmitter or an antenna, user-specific QoS, fronthaul capacity, etc. Function $h_i(\mathbf{x})$ can be either convex or nonconvex. In either case, (11.1) is a *nonconvex program* due to the fractional structure of the objective [15]. In the special case where $f(\mathbf{x})$, $g(\mathbf{x})$, and $h_i(\mathbf{x})$ $(i = 1, \ldots, m)$ are convex, (11.1) is quasi-concave program of which globally optimum solutions can be derived by linearly (or even superlinearly) convergent optimization methods such as Dinkelbach's algorithms or Charnes–Cooper's transformations [15]. In wireless communications, this likely corresponds to the noise-limited scenarios [16]. However, this is not the case of the C-RANs where multiple RRHs serve multiple users using the same sources of time and frequency, creating interference-limited channels. To be specific, the rate function in the C-RAN is nonconvex with respect to power or beamforming variables. In such scenarios, globally optimal solutions of nonconvex program (11.1) can be found by the non-polynomial time global optimization frameworks such as MO [16]. Nevertheless, such methods are suitable for benchmarking purposes only as it often requires prohibitively high computational complexity. In practice, efficient optimization approaches that tend to yield near-optimal solutions using lower computational resources are more desirable. In the following, we discuss global and local optimization frameworks that can be used to solve (11.1).

11.2.1 Global optimization method: monotonic optimization

Numerous nonconvex optimization problems in wireless communications fall to the class of MO [17]. These can be solved globally by a powerful algorithm in the framework of MO, namely, branch-reduce-and-bound (BRnB). In this section, we provide some backgrounds of the MO and describe the BRnB method. To proceed, some basic concepts of MO are first introduced [18].

Definition 11.1. *For any two vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, writing* $\mathbf{y} \geq \mathbf{x}$ *(*$\mathbf{y} > \mathbf{x}$*, resp.) means* $y_i \geq x_i$ *(*$y_i > x_i$*, resp.) for every* $i = 1, 2, \ldots, n$*. If* $\mathbf{a} \leq \mathbf{b}$*, then the box* $[\mathbf{a}, \mathbf{b}]$ *((*$\mathbf{a}, \mathbf{b}$*], resp.) is the set of all* $\mathbf{x} \in \mathbb{R}^n$ *satisfying* $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$ *(*$\mathbf{a} < \mathbf{x} \leq \mathbf{b}$*, resp.).*

Definition 11.2. (Increasing function) *A function* $f : \mathbb{R}^n_+ \to \mathbb{R}$ *is said to be increasing on* $\mathbb{R}^n_+$ *if* $f(\mathbf{x}') \geq f(\mathbf{x})$ *when* $\mathbf{x}' \geq \mathbf{x}$*.*

Definition 11.3. (Normal set) *A set* $\mathscr{S} \subset [\mathbf{a}, \mathbf{b}]$ *is said to be normal in* $[\mathbf{a}, \mathbf{b}]$ *(or briefly, normal) if* $\mathbf{x} \in \mathscr{S} \Rightarrow [\mathbf{a}, \mathbf{x}] \subset \mathscr{S}$*. For a closed normal set* $\mathscr{S}$ *in* $[\mathbf{a}, \mathbf{b}]$*, a point* $\bar{\mathbf{x}}$ *is called an upper boundary point if the cone* $\boldsymbol{K}_{\bar{\mathbf{x}}} \triangleq \{\mathbf{x} \mid \mathbf{x} > \bar{\mathbf{x}}\}$ *contains no point of* $\mathbf{x} \in \mathscr{S}$*.*

The standard form of an MO problem is generally written as [18]

$$\underset{\mathbf{x}}{\text{maximize}} \quad f(\mathbf{x}) \tag{11.2a}$$

$$\text{subject to} \quad h_i(\mathbf{x}) - \tilde{h}_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m \tag{11.2b}$$

$$\mathbf{x} \subseteq B \triangleq [\mathbf{a}; \mathbf{b}], \tag{11.2c}$$

where $f(\mathbf{x})$, $h_i(\mathbf{x})$, and $\tilde{h}_i(\mathbf{x})$ $(i = 1, \ldots, m)$ are increasing functions of variable $\mathbf{x}$, and B is the box containing the feasible set of (11.2) i.e., $\{\mathbf{x} | h_i(\mathbf{x}) - \tilde{h}_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m\} \subseteq B$. Vectors $\mathbf{a}$ and $\mathbf{b}$ are the lower and upper vertices of B, respectively. We consider here a mixed-variable vector $\mathbf{x}$ of size N containing binary and continuous variables, i.e., $x_j \in \{0, 1\}$ for $j = 1, \ldots, s$ and $x_j \in \mathbb{R}$ for $j = s + 1, \ldots N$. The MO problem can be solved by the BRnB method introduced in [18,19]. The following presents the main steps of the BRnB.

Branch–reduce-and-bound algorithm

BRnB algorithm is an iterative procedure executing three basic operations at each iteration: *branching*, *reduction*, and *bounding*. Starting from initial box $[\mathbf{a}; \mathbf{b}]$, the algorithm (i) iteratively divides it into smaller and smaller ones; (ii) removes boxes that do not contain an optimal solution; and (iii) searches the remaining boxes for an improved solution until an error tolerance is met. An example of the BRnB procedure is illustrated in Figure 11.1.

It is worth mentioning that the BRnB algorithm was first developed for continuous MO problems [19]. As a result, when applying to problems containing discrete variables/feasible set as in (11.2), the algorithm may not return exact solutions of the discrete variables, but approximated ones [18]. To avoid such drawback, a modification of the BRnB was later developed for the discrete optimization programs in which an adjustment step is additionally applied to the discrete variables to map them into the corresponding discrete set. The details of BRnB method solving (11.2) are provided next.

For the ease of exposition, we introduce the following denotations that are used throughout this section: $\mathbf{e}_j$ is a unit vector such that $[\mathbf{e}_j] = 1$ and $[\mathbf{e}_i] = 0$ if $i \neq j$; n counts the iteration index; $\mathscr{B}_n$ denotes the set of candidate boxes at the nth iteration and $V = [\mathbf{p}; \mathbf{q}] \in \mathscr{B}_n$ denotes an arbitrary box in $\mathscr{B}_n$; $f_{\text{UpB}}(V)$ denotes the upper bound of V, i.e., $f_{\text{UpB}}(V) \geq \max\{f(\mathbf{x}) \mid h_i(\mathbf{x}) - \tilde{h}_i(\mathbf{x}) \leq 0, i = 1, \ldots, m,\ \mathbf{x} \in [\mathbf{p}; \mathbf{q}]\}$; the *current*

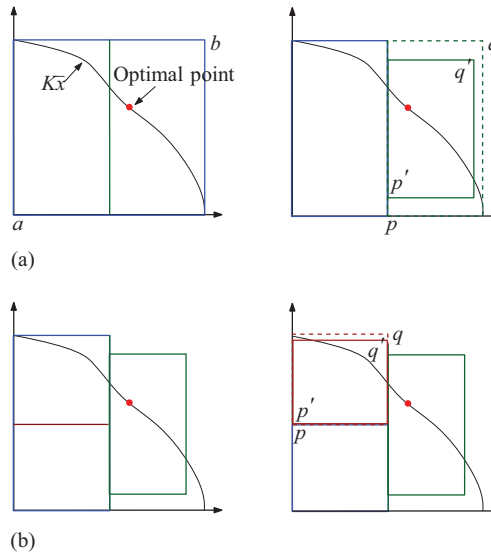


Figure 11.1 Example of the BRnB procedure for branching and reduction operations over two iterations: (a) Iteration 1 and (b) Iteration 2

best objective (CBO) stands for the known point in the feasible set that offers the best objective value at the current iteration. Note that at the very first iteration, we have $\mathscr{B}_1 \equiv B$, $[\mathbf{p}; \mathbf{q}] \equiv [\mathbf{a}; \mathbf{b}]$, $f_{\text{UpB}}(V) = f(\mathbf{b})$.

Branching

At iteration n, one box in $\mathscr{B}_n$ is picked to be branched into two new boxes. In particular, the candidate box, denoted as V_{c}, is bisected along the longest edge, which is determined as $l = \arg\max_{1 \le j \le N}(q_j - p_j)$. This results in two new smaller boxes of equal size given by

$$V^1 = [\mathbf{p}; \mathbf{q}'] \text{ where } q'_j = \begin{cases} q_j & \text{if } j \neq l, \\ 0 & \text{if } j = l \le s, \\ q_j - (q_j - p_j)/2 & \text{if } j = l > s, \end{cases} \tag{11.3}$$

$$V^2 = [\mathbf{p}'; \mathbf{q}] \text{ where } p'_j = \begin{cases} p_j & \text{if } j \neq l, \\ 1 & \text{if } j = l \le s, \\ p_j + (q_j - p_j)/2 & \text{if } j = l > s. \end{cases} \tag{11.4}$$

We note that V_{c} in general should be selected such that it has the largest upper bound among boxes in $\mathscr{B}_n$, i.e., $V_{\text{c}} = \arg\max_{V \in \mathscr{B}_n} f_{\text{UpB}}(V)$ [19]. This is to ensure the monotonic decrease (increase, respectively) of the upper bound (lower bound, respectively) of the resulting boxes.

Reduction

The newly created boxes may contain some portions of no interest, i.e., those have no feasible solution, or if they do, their feasible points give objective values smaller than the CBO. Thus, efficiently cutting such regions out of the box shall limit the feasible space, thereby accelerating the convergence. The reduction step is developed for this purpose. In particular, given a box $V = [\mathbf{p}; \mathbf{q}]$, we use reduction cut to find a smaller box, denoted as $\mathrm{r}(V) \triangleq [\mathbf{p}'; \mathbf{q}'] \subset V$, such that an optimal solution (if it exists in V) must be contained in $\mathrm{r}(V)$. In other words, we eliminate the portions $[\mathbf{p}; \mathbf{p}')$ and $(\mathbf{q}'; \mathbf{q}]$ that are checked to not contain an optimal solution. To do so, we replace $\mathbf{p}$ by $\mathbf{p}' \geq \mathbf{p}$ where $\mathbf{p}' = \mathbf{q} - \sum_{j=1}^{N} \rho_j^{(1)}(q_j - p_j)\mathbf{e}_j$ and

$$\begin{aligned}\rho_j^{(1)} = \sup\{\rho^{(1)} | 0 \leq \rho^{(1)} \leq 1,\ h_i(\mathbf{p}) - \tilde{h}_i(\mathbf{q} - \rho^{(1)}(q_j - p_j)\mathbf{e}_j) \leq 0,\\ f(\mathbf{q} - \rho^{(1)}(q_j - p_j)\mathbf{e}_j) \geq \mathrm{CBO}\},\ \text{for}\, j = 1, \ldots, N.\end{aligned} \tag{11.5}$$

To find $\mathbf{q}'$, we replace $\mathbf{q}$ by $\mathbf{q}' \leq \mathbf{q}$ where $\mathbf{q}' = \mathbf{p}' + \sum_{j=1}^{N} \rho_j^{(2)}(q_j - p_j')\mathbf{e}_j$ and

$$\rho_j^{(2)} = \sup\{\rho^{(2)} | 0 \leq \rho^{(2)} \leq 1,\ h(\mathbf{p}' + \rho^{(2)}(q_j - p_j')\mathbf{e}_j) - \tilde{h}_i(\mathbf{q}) \leq 0\}. \tag{11.6}$$

The values of $\rho_j^{(1)}$ and $\rho_j^{(2)}$ can be found easily using the bisection method.

Remark that for $j = 1, \ldots, s$ (i.e., *binary variables*), we have $q_j - p_j = 1$, and thus, p_j' can be quickly obtained by

$$p_j' = \begin{cases} 0 & \text{if } h_i(\mathbf{p}) - \tilde{h}_i(\mathbf{q} - \mathbf{e}_j) \leq 0 \\ 1 & \text{otherwise.} \end{cases}$$

If $p_j' = 0$, then $q_j - p_j' = 1$, and thus q_j' is computed as

$$q_j' = \begin{cases} 1 & \text{if } h(\mathbf{p}' + \mathbf{e}_j) - \tilde{h}_i(\mathbf{q}) \leq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Bounding

Bounding is the most crucial operation to ensure the convergence of the branch-and-bound-type methods in general and the BRnB method in particular. In this step, we update the upper and lower bounds of the boxes after reduction (i.e., $\mathrm{r}(V) \triangleq [\mathbf{p}'; \mathbf{q}']$). Due to the monotonicity, i.e., $\mathbf{p}' \leq \mathbf{x} \leq \mathbf{q}'$ for any $\mathbf{x} \in \mathrm{r}(V)$, the upper and lower bounds of box $\mathrm{r}(V)$ can be simply found as $f_{\mathrm{LoB}}(V) = f(\mathbf{p}')$ and $f_{\mathrm{UpB}}(V) = f(\mathbf{q}')$, respectively. The convergence is declared when $f_{\mathrm{UpB}}(V) - f_{\mathrm{LoB}}(V) \leq \varepsilon$ for small threshold $\varepsilon > 0$.

It is worth mentioning that if the newly updated upper bound of a box is smaller than the CBO, we can remove that box to save the computational complexity. Thus, together with bounding, efficiently updating the CBO is beneficial for improving the algorithm's efficiency.

Algorithm 11.1: The SCA procedure solving (11.7)

1: **Initialization:** Set $n := 0$, choose an initial feasible point $\mathbf{x}^{(n)}$
2: **repeat** $\{n := n + 1\}$
3: Solve (11.9) and obtain optimal value $\mathbf{x}^*$
4: Update $\mathbf{x}^{(n)} := \mathbf{x}^*$
5: **until** Convergence
6: **Output:** $\mathbf{x}^{(n)}$

11.2.2 Local optimization method: successive convex approximation

For a general nonconvex program, optimal solutions could be found applying the branch-and-bound-type methods but with the price of extremely high computational complexity. Instead, using local optimization approaches to arrive at near-optimal solutions is more appealing in practice [20]. Among the local optimization approaches, the SCA method is increasingly applied for nonconvex problems in wireless communications and also other fields.

SCA is an iterative method that aims at locating a Karush–Kuhn–Tucker (KKT) solution to a nonconvex program. The central idea is to iteratively approximate the nonconvex parts by the proper convex ones [21,22]. The general procedure of the SCA is outlined in Algorithm 11.1, and the details are discussed next. Let us consider a general optimization program given by

$$\underset{\mathbf{x}}{\text{minimize}}\ f_0(\mathbf{x}) \quad \text{subject to} \quad \{f_i(\mathbf{x}) \le 0,\ i = 1, ..., m\}, \tag{11.7}$$

where $f_i(\mathbf{x})$ $(i = 0, \ldots, l)$ are all continuously differentiable functions over $\mathbb{R}^N$. The feasible set is denoted as $\mathscr{S}(\mathbf{x}) \triangleq \{\mathbf{x} \in \mathbb{R}^N |\ f_i(\mathbf{x}) \le 0,\ i = 1, \ldots, m\}$, which is assumed to be a compact set. In $\mathscr{S}(\mathbf{x})$, we also assume that $f_i(\mathbf{x})$ $(i = 0, \ldots, l)$ are convex functions and $f_i(\mathbf{x})$ $(i = l + 1, \ldots, m)$ are nonconvex ones. Clearly the problem is nonconvex due to the last $l - m$ constraints. The central idea of the SCA is to iteratively approximate $f_i(\mathbf{x})$ $(i = l + 1, \ldots, m)$ by its convex upper bounds [21,22]. More specifically, given a feasible point $\mathbf{x}' \in \mathscr{S}(\mathbf{x})$, $f_i(\mathbf{x})$ $(i = l + 1, \ldots, m)$ is replaced by a convex function $\tilde{f}_i(\mathbf{x}, g(\mathbf{x}'))$ satisfying the following properties:

$$\begin{aligned}
&\textbf{(a)}\ f_i(\mathbf{x}) \le \tilde{f}_i(\mathbf{x}; g_i(\mathbf{x}')) \\
&\textbf{(b)}\ f_i(\mathbf{x}') = \tilde{f}_i(\mathbf{x}'; g_i(\mathbf{x}')) \\
&\textbf{(c)}\ \nabla_{\mathbf{x}} f_i(\mathbf{x}') = \nabla_{\mathbf{x}} \tilde{f}_i(\mathbf{x}', g_i(\mathbf{x}')),
\end{aligned} \tag{11.8}$$

for all $\tilde{\mathscr{S}}(\mathbf{x}; \mathbf{x}') \triangleq \{\mathbf{x} \in \mathbb{R}^N | f_i(\mathbf{x}) \le 0, \quad i = 1, \ldots, l; \quad \tilde{f}_i(\mathbf{x}, g(\mathbf{x}')) \ne 0, \quad i = l + 1, \ldots, m\}$, where $g_i(\mathbf{x}')$: $\mathbb{R}^N \to \mathbb{R}^K$ is a parameter vector. We remark that $\tilde{\mathscr{S}}(\mathbf{x}; \mathbf{x}')$

needs to satisfy Slater's constraint qualification [21]. The replacement leads to the following convex subproblem:

$$\underset{\mathbf{x}}{\text{minimize}}\ f_0(\mathbf{x}) \quad \text{subject to} \quad \{\mathbf{x} \in \tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}')\}. \tag{11.9}$$

Since the optimal solution of (11.9), denoted as $\mathbf{x}^*$, belongs to $\tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}')$ due to the conditions (a) and (b), it is relevant to use $\mathbf{x}^*$ to form subproblem (11.9) for the next iteration. To be specific, let $\mathbf{x}^{(n)}$ denote the optimal solution at iteration n, then the feasible set of the subproblem at iteration $n+1$ is $\tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}^{(n)})$ (see Step 4 in Algorithm 11.1). The process is iteratively carried out until the convergence is established.

The update rule and the conditions (a) and (b) guarantee the convergence of Algorithm 11.1. In particular, recall that $\tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}^{(n)})$ is the feasible set of the subproblem at iteration $n+1$. Thus, as $\mathbf{x}^{(n+1)}$ is the optimal solution obtained by solving (11.9) with $\tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}^{(n)})$, we immediately have $f_0(\mathbf{x}^{(n+1)}) \leq f_0(\mathbf{x}^{(n)})$ since $\mathbf{x}^{(n)}$ is a point in $\tilde{\mathscr{S}}(\mathbf{x};\mathbf{x}^{(n)})$ due to (11.8)(a) and (b). This implies the monotonic decrease of the sequence of objective values. As a result, the convergence is ensured given that the feasible set $\mathscr{S}(\mathbf{x})$ is bounded. On the other hand, the convergence points satisfy the KKT optimality conditions due to the properties (b) and (c) as shown in [21,22].

It is worth noting that if $f_i(\mathbf{x})$ $(i = l+1,\ldots,m)$ are not differentiable but subdifferentiable at $\mathbf{x}'$, the algorithm can still produce a monotonically decreasing sequence of objective because (11.8)(a) and (b) remain satisfied in this case. However, the approximation function here may not hold (11.8)(c), and thus achieved solutions at the convergence may not satisfy the KKT optimality conditions. Another remark is that the objective $f_0(\mathbf{x})$ might be nonconvex in some applications. In such problems, we can apply the SCA principle for $f_0(\mathbf{x})$, i.e., $f_0(\mathbf{x})$ is replaced by $\tilde{f}_0(\mathbf{x};g_0(\mathbf{x}'))$ satisfying the conditions in (11.8) in the approximate problem (11.9). For more convergence results of the SCA, we refer the interested readers to references [21–23].

Obviously, it is important to find the convex approximations that hold the conditions in (11.8). We provide next some SCA-applicable functions and their approximated formulations that hold conditions in (11.8). Those are useful to derive SCA-based algorithms for the energy efficiency maximization problems.

Useful approximate formulations

- Fractional-linear function: $\phi(x,y) \triangleq \frac{x}{y}$ where $(x,y) \in \mathbb{R}^2_{++}$. Its convex approximation is given by

$$\phi^{\text{frac}}(x,y;\lambda) \triangleq 0.5\left(\lambda x^2 + \frac{1}{\lambda y^2}\right), \text{ where } \lambda = \frac{1}{x'y'}. \tag{11.10}$$

- Quadratic-over-linear function: $\phi(\mathbf{x},y;\mathbf{H}) \triangleq \frac{\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{y}$ where $\mathbf{x} \in \mathbb{C}^n$, $y \in \mathbb{R}_{++}$, and $\mathbf{H} \succeq 0$. The approximation of $\phi(\mathbf{x},y;\mathbf{H})$ can be achieved by means of the first-order Taylor series, i.e.,

$$\begin{aligned}\phi^{\text{qol}}(\mathbf{x},y;\mathbf{x}',y';\mathbf{H}) &\triangleq \phi(\mathbf{x}',y';\mathbf{H}) + \left\langle [\nabla_{\mathbf{x}}\phi(\mathbf{x}',y';\mathbf{H}), \nabla_y\phi(\mathbf{x}',y';\mathbf{H})]^{\mathrm{T}}, [\mathbf{x}-\mathbf{x}', y-y']^{\mathrm{T}} \right\rangle \\ &= \frac{(\mathbf{x}')^{\mathrm{H}}\mathbf{H}\mathbf{x}'}{(y')^2}y - \frac{2\Re((\mathbf{x}')^{\mathrm{H}}\mathbf{H}\mathbf{x})}{y'}.\end{aligned} \tag{11.11}$$

- Power function: $\phi(x;p) \triangleq x^p$ where $x \in \mathbb{R}_{++}$. An approximation of $\phi(x;p)$ is
$$\phi^{\text{po}}(x;x';p) \triangleq (p-1)(x')^p - p(x')^{p-1}x. \tag{11.12}$$

11.3 Cloud radio access networks: system model and energy efficiency optimization formulation

We focus on examining the problem of beamformer design maximizing the network energy efficiency of a downlink C-RAN taking into account the impact of capacity-finite fronthaul links. To address the fronthaul limitation, the idea is to reduce the amount of BB signals exchanged through the fronthaul links. This is done by selecting a properly small subset of users associated with an RRH, giving rise to the RRH-user association problem that is jointly designed with the transmit beamforming. In addition, to exploit the potential energy efficiency gain of the C-RAN, RRH selection scheme is used in which some RRHs are properly switched into a sleep mode for power-saving purposes. Noticeably, RRH-user association and RRH selection are commonly adopted in the resource allocation for the C-RAN to cope with the fronthaul constraints and/or to improve the energy efficiency.

The energy-efficient designs for the C-RAN employing RRH-user association and RRH selection have been addressed in some recent works [24,25]. Therein, however, energy efficiency is improved by minimizing the total power consumption [24,26]. Since the data rate is not jointly optimized in the objective, energy efficiency performance achieved by such methods may be far from the optimal one. Very recently, approaches addressing the energy efficiency objective in C-RAN have been developed in [10,12,27], which are summarized herein.

We present a global optimization method to the joint design problem based on the BRnB. For practical applications, we provide three suboptimal solutions based on the SCA that aim at achieving near-optimal solutions but with remarkably reduced complexity compared to the globally optimal method.

11.3.1 System model

We consider the downlink of a C-RAN in which R RRHs, each equipped with M antennas,* cooperatively serve K single-antenna users following joint transmission scheme [28]. In particular, data information for each user is transmitted from multiple RRHs. In this section, we focus on the data-sharing-based fronthauling strategy. Energy-efficient transmission designs for compression-based CRANs can be referred to [12,27]. For notational convenience, we denote by $\mathscr{R} \triangleq \{1,\ldots,R\}$ the set of RRHs, and $\mathscr{K} \triangleq \{1,\ldots,K\}$ the set of users. The transmit signals are processed at a common BBU pool, which are then forwarded to the RRHs via finite-capacity fronthaul links. A simplified model of the considered system is illustrated in Figure 11.2. The BBU pool is assumed to perfectly know the channel state information associated with all the users in the network. We denote by $\mathbf{h}_{i,k} \in \mathbb{C}^{1\times M}$ the channel vector between RRH

*The assumption on the number of equipped antennas for all RRHs is merely for notational simplicity.

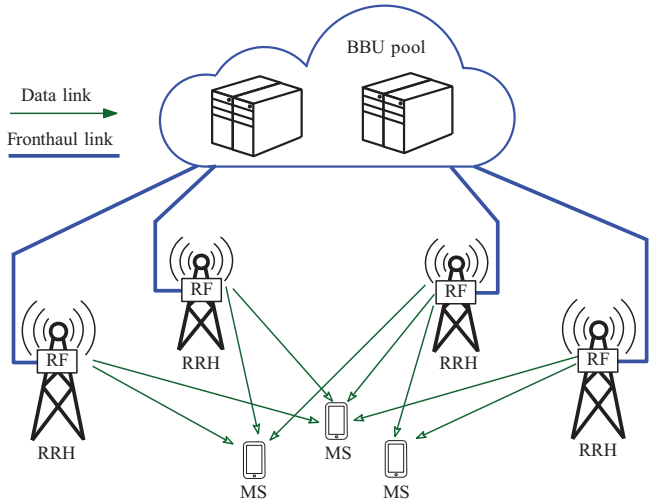


Figure 11.2 A simplified system model of the downlink of a multiuser C-RAN

i and user k. We also assume that linear beamforming scheme is adopted to form the RF transmit signals at the RRHs. Let d_k denote the unit-energy data symbol intended for user k, i.e., $\mathbb{E}[|d_k|^2] = 1$, and $\mathbf{w}_{i,k} \in \mathbb{C}^{M\times 1}$ be the transmit beamformer for d_k at RRH i. The received signal at user k can be written as

$$y_k = \underbrace{\Big(\sum_{i\in\mathscr{R}} \mathbf{h}_{i,k}\mathbf{w}_{i,k}\Big)d_k}_{\text{desired signal}} + \underbrace{\sum_{j\in\mathscr{K}\setminus k}\Big(\sum_{i\in\mathscr{R}} \mathbf{h}_{i,k}\mathbf{w}_{i,j}\Big)d_j}_{\text{interference}} + n_k$$

$$= \mathbf{h}_k\mathbf{w}_k d_k + \sum_{j\in\mathscr{K}\setminus k} \mathbf{h}_k\mathbf{w}_{i,j}d_j + n_k, \tag{11.13}$$

where $n_k \sim \mathscr{CN}(0,\sigma_k^2)$ is the additive white Gaussian noise at user k. In (11.13), we have denoted $\mathbf{h}_k \triangleq [\mathbf{h}_{1,k},\mathbf{h}_{2,k},\ldots,\mathbf{h}_{R,k}] \in \mathbb{C}^{1\times MR}$ and $\mathbf{w}_k \triangleq [\mathbf{w}_{1,k}^{\mathrm{T}},\mathbf{w}_{2,k}^{\mathrm{T}},\ldots,\mathbf{w}_{R,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{MR\times 1}$ that are the aggregate vectors of all channels and beamformers from all RRHs to user k, for notational simplicity. Assuming that each user decodes its own data symbol while treating interference symbol as noise. Then the signal-to-interference-plus-noise can be written as

$$\gamma_k(\mathbf{w}) \triangleq \frac{|\mathbf{h}_k\mathbf{w}_k|^2}{\sum_{j\in\mathscr{K}\setminus k}|\mathbf{h}_k\mathbf{w}_j|^2 + \sigma_k^2}. \tag{11.14}$$

Let r_k be the achievable data rate transmitted to user k. To guarantee reliable communication, the rate must be below the instantaneous mutual information between channel input and output or the constraint

$$r_k \le W\log(1+\gamma_k(\mathbf{w})),\ \forall k \in \mathscr{K},$$

where W is the transmission bandwidth, must hold.

11.3.2 Power constraints

The power used for transmission at RRH i should not exceed its available power budget, denoted as P_i. The power constraint for RRH i can be expressed as

$$\sum_{k\in\mathcal{K}} \|\mathbf{w}_{i,k}\|_2^2 \leq P_i,\ \forall i \in \mathcal{R}.$$

In addition, the power amplifier (PA) of each antenna chain has the maximum RF output power. Let $[\mathbf{w}_{i,k}]_m$ denote the beamforming coefficient in vector $\mathbf{w}_{i,k}$ corresponding to antenna m, and $\tilde{\mathbf{w}}_{i,m} \triangleq [[\mathbf{w}_{i,1}]_m; [\mathbf{w}_{i,2}]_m; \ldots; [\mathbf{w}_{i,K}]_m] \in \mathbb{C}^{K\times 1}$ stack all $[\mathbf{w}_{i,k}]_m$. The RF power radiated at antenna m should satisfy:

$$\sum_{k\in\mathcal{K}} |[\mathbf{w}_{i,k}]_m|^2 = \|\tilde{\mathbf{w}}_{i,m}\|_2^2 \leq P_{i,m}^{\text{ant}},$$

where $P_{i,m}^{\text{ant}}$ is the maximum RF output power at the antenna m of RRH i. Without loss of generality, we assume that $P_i = \bar{P}, \forall i$; and $P_{i,m}^{\text{ant}} = P^{\text{ant}}, \forall i, m$.

11.3.3 Fronthaul constraint

We assume that the maximum capacity of a fronthaul link is $\bar{C}$. For feasible transmission, the total data rate of the wireless physical layer of RRH i should not be larger than $\bar{C}$. It is noted that the fronthaul rate in each link is proportional to the number of users served by the corresponding RRH. As a result, under the fronthaul-constrained scenario, each RRH might not need to serve all users in the network, but a smaller group of selected users to which it can potentially provide the best network performance [9,25]. In such cases, the RRH-user association method is needed.

For the formulation purpose, let us define the following indicator function:

$$\mathbb{I}(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}. \tag{11.15}$$

Now we can present by the indicator function on the beamforming vector $\mathbf{w}_{i,k}$, whether RRH i sends information to user k as

$$\mathbb{I}(\|\mathbf{w}_{i,k}\|_2^2) = \begin{cases} 1 & \text{if } \|\mathbf{w}_{i,k}\|_2^2 > 0 \\ 0 & \text{if } \|\mathbf{w}_{i,k}\|_2^2 = 0 \end{cases}.$$

Accordingly, the aggregate data rate, which can be reliably transmitted by the wireless interface of RRH i, is calculated as $\sum_{k\in\mathcal{K}} \mathbb{I}(\|\mathbf{w}_{i,k}\|_2^2) r_k$. Thus, the fronthaul constraint can be expressed as

$$\sum_{k\in\mathcal{K}} \mathbb{I}(\|\mathbf{w}_{i,k}\|_2^2) r_k \leq \bar{C},\ \forall i \in \mathcal{R}.$$

In practice, the values of $\bar{C}$ range from tens of Mbits/s to tens of Gbits/s, depending on fronthaul transport solutions (e.g., passive optical networks, fiber, microwave, Ethernet cables, and wireless communications) and deployments/network requirements (e.g., indoor, outdoor, and urban or rural environments) [14,29,30].

11.3.4 Power consumption

The power consumption is the crucial element to quantify the energy efficiency measure. In this section, we model the network power consumption following those in [31–34] which includes three main components: power for operating electronic circuits, power for signal processing and fronthauling, and power dissipated on PAs.

11.3.4.1 Circuit power consumption

A significant amount of power is required for operating the RF chain of RRHs (e.g., analog-to-digital or digital-to-analog converters, filters, and mixers) as well as the elements of network infrastructure. Assuming that RRHs and the associated network units are configured to be switchable between sleeping and active modes, each of them consumes a dedicated power P^{act} and P^{sl}, respectively, with $P^{\text{act}} \gg P^{\text{sl}}$. For power-saving purposes, an RRH should be put into a sleep mode when it does not transmit to any user, and otherwise it is active when having data to transmit. To represent the operation modes of the RRH, we can use the indicator function introduced in (11.15). By noting that $\sum_{k\in\mathcal{K}} \|\mathbf{w}_{i,k}\|_2^2$ is the total radiated power at RRH i, the operation modes of the RRH can be interpreted as $\mathbb{I}(\sum_{k\in\mathcal{K}} \|\mathbf{w}_{i,k}\|_2^2)$. The circuit power consumption can be then modeled as [32,33]

$$P_{\text{cir}}(\mathbf{w}) \triangleq \sum_{i\in\mathcal{R}} \mathbb{I}\left(\sum_{k\in\mathcal{K}} \|\mathbf{w}_{i,k}\|_2^2\right) P^{\text{act}} + \sum_{i\in\mathcal{R}} \left(1 - \mathbb{I}\left(\sum_{k\in\mathcal{K}} \|\mathbf{w}_{i,k}\|_2^2\right)\right) P^{\text{sl}} + P_0, \quad (11.16)$$

where P_0 includes the fixed power consumption to keep the network operating, e.g., power supply for the network infrastructures [32].

11.3.4.2 Signal processing and fronthauling power

The data needs to be encoded/decoded and modulated/demodulated. In general, a higher data rate generally requires a larger codebook, and the larger number of bits incurs higher power for encoding and decoding on the BB circuit boards. Moreover, the fronthaul network transports the data signals, and, thus, also consumes power. In fact, the power consumption for fronthauling varies with the data rate. For example, in wireless fronthaul networks, higher fronthaul rate requires higher power consumption. From this standpoint, the amount of power consumed for signal processing and fronthauling is rate dependent. Assuming that this amount of power is linearly scaled with the total fronthaul rate [31], the signal processing and fronthauling power can be written as

$$P_{\text{SP-FH}}(\mathbf{w}, \mathbf{r}) \triangleq \delta \sum_{i\in\mathcal{R}} \sum_{k\in\mathcal{K}} \mathbb{I}(\|\mathbf{w}_{i,k}\|_2^2) r_k, \quad (11.17)$$

where δ is a constant coefficient with unit W/(Gbits/s) and $\mathbf{r} \triangleq [r_1, \ldots, r_K]^{\text{T}}$.

11.3.4.3 Dissipated power on PA

The amount of power dissipated on the PAs strongly depends on the PA's efficiency which is defined as the ratio between the total RF output power and the direct current input power. Conventionally, PA's efficiency is assumed to be a constant over the operating range. However, this assumption may hold true for only some transceiver types, in which advanced PA architectures can be implemented to make the PA's transfer function operating in its linear region (e.g., macro and micro BSs in LTE systems). For smaller size and/or low-cost transmitter such as RRHs, such techniques are expensive to employ [32]. Therefore, for this scenario, PA's efficiency is highly dependent on the output power region. To account for this, we consider the nonlinear power consumption model of the PA, in which the PA's efficiency is a function of its radiated power [34,35], i.e.:

$$\varepsilon_{i,m}(\mathbf{w}) \triangleq \frac{\sqrt{P^{\text{ant}}}}{\varepsilon_{\max}} \sqrt{\sum_{k\in\mathscr{K}} |[\mathbf{w}_{i,k}]_i|^2} = \varepsilon \|\tilde{\mathbf{w}}_{i,m}\|_2, \tag{11.18}$$

where $\varepsilon \triangleq \frac{\sqrt{P^{\text{ant}}}}{\varepsilon_{\max}}$ and P^{ant} and $\varepsilon_{\max} \in [0, 1]$ are the maximum power of the PA and the maximum PA's efficiency, respectively. Following the definition of the PA's efficiency, the power consumed at the PA m of RRH i for radiating the signals outward the antenna is

$$P_{\text{PA},(i,m)}(\mathbf{w}) \triangleq \frac{\|\tilde{\mathbf{w}}_{i,m}\|_2^2}{\varepsilon_{i,m}(\mathbf{w})} = \frac{1}{\varepsilon}\|\tilde{\mathbf{w}}_{i,m}\|_2, \tag{11.19}$$

which is clearly a function of the beamforming vector. From (11.19), the total power dissipated on all the PAs is calculated as

$$P_{\text{PA}}(\mathbf{w}) \triangleq \frac{1}{\varepsilon} \sum_{i\in\mathscr{R}} \sum_{m=1}^{M} \|\tilde{\mathbf{w}}_{i,m}\|_2. \tag{11.20}$$

11.3.4.4 Total power consumption

To summarize, the total consumed power in the considered system, denoted as $P(\mathbf{w}, \mathbf{r})$, can be expressed as

$$\begin{aligned} P(\mathbf{w}, \mathbf{r}) &\triangleq P_{\text{cir}}(\mathbf{w}) + P_{\text{SP-FH}}(\mathbf{w}, \mathbf{r}) + P_{\text{PA}}(\mathbf{w}) \\ &= \frac{1}{\varepsilon} \sum_{i\in\mathscr{R}} \sum_{m=1}^{M} \|\tilde{\mathbf{w}}_{i,m}\|_2 + (P^{\text{act}} - P^{\text{sl}}) \sum_{i\in\mathscr{R}} \mathbb{I}\left(\sum_{k\in\mathscr{K}} \|\mathbf{w}_{i,k}\|_2^2\right) \\ &\quad + \delta \sum_{i\in\mathscr{R}} \sum_{k\in\mathscr{K}} \mathbb{I}\left(\sum_{k\in\mathscr{K}} \|\mathbf{w}_{i,k}\|_2^2\right) r_k + \underbrace{RP^{\text{sl}} + P_0}_{\triangleq P_{\text{c}}}. \end{aligned} \tag{11.21}$$

11.3.5 Problem formulation

We focus on maximizing the overall network energy efficiency of the C-RAN while satisfying the user-specific QoS constraints, fronthaul constraints, and power constraints. In particular, the joint design problem of beamforming, data rate, and RRH-user association and RRH selection design is expressed as

$$\underset{\mathbf{w},\mathbf{r}}{\text{maximize}} \quad \frac{\sum_{k\in\mathscr{K}} r_k}{P(\mathbf{w},\mathbf{r})} \tag{11.22a}$$

$$\text{subject to} \quad r_k \le \log(1+\gamma_k(\mathbf{w})),\ \forall k \in \mathscr{K} \tag{11.22b}$$

$$r_k \ge r_0,\ \forall k \in \mathscr{K} \tag{11.22c}$$

$$\sum_{k\in\mathscr{K}} \mathbb{I}(\|\mathbf{w}_{i,k}\|_2) r_k \le \bar{C},\ \forall i \in \mathscr{R} \tag{11.22d}$$

$$\sum_{k\in\mathscr{K}} \|\mathbf{w}_{i,k}\|_2^2 \le \bar{P},\ \forall i \in \mathscr{R} \tag{11.22e}$$

$$\|\tilde{\mathbf{w}}_{i,m}\|_2^2 \le P^{\text{ant}},\ \forall i \in \mathscr{R}, m = 1,\ldots,M. \tag{11.22f}$$

where r_0 is the predefined data rate threshold of a user. Problem (11.22) is a general nonsmooth nonconvex optimization problem due to the nonconvex objective (11.22a) and the constraints (11.22b) and (11.22d). In fact, by means of the RRH-user association, the joint beamforming optimization in (11.22) can be viewed as finding group sparse solutions of $\mathbf{w}$. Hence we can apply some sort of sparse optimization algorithms to solve (11.22) [25,36].

Nevertheless, (11.22) can be alternatively cast as a mixed integer program to which a wider class of optimization tools are available. In particular, for beamforming designs in wireless communications, indicator function (11.15) can be simplified by the following constraints:

$$\|\mathbf{w}_{i,k}\|_2 \le x_{i,k}\sqrt{\bar{P}}, x_{i,k} \in \{0,1\},\ \forall i \in \mathscr{R}, k \in \mathscr{K}, \tag{11.23a}$$

$$s_i \ge x_{i,k}, \forall k \in \mathscr{K};\ s_i \le \sum_{k\in\mathscr{K}} x_{i,k}, s_i \in \{0,1\},\ \forall i \in \mathscr{R}, \tag{11.23b}$$

where $x_{i,k}$ can be seen as a preference variable representing the association between RRH i and user k, i.e., RRH i does not transmit to user k if $x_{i,k} = 0$, and it does if $x_{i,k} = 1$. Clearly we can see by (11.23a) that $\|\mathbf{w}_{i,k}\|_2 = 0$ if $x_{i,k} = 0$, and $\|\mathbf{w}_{i,k}\|_2 > 0$ otherwise. Similarly, s_i is the preference variable representing the operating state of RRH i, i.e., $s_i = 0$ if RRH i is in sleep mode, and $s_i = 1$ if RRH i is active. Constraint (11.23b) is to tighten the relationship between s_i and $x_{i,k}$, i.e., $s_i = 1$ if there exists (i,k) such that $x_{i,k} = 1$, and $s_i = 0$ otherwise. In other words, RRH i is active if it transmits information to at least one user and sleeps if having no data to transmit.

Based on the previous discussion, we can replace indicator $\mathbb{I}(\cdot)$ by binary variables $x_{i,k}$ and s_i and recast the considered problem as

$$\underset{\mathbf{w},\mathbf{x},\mathbf{s},\mathbf{r}}{\text{maximize}} \quad \frac{\sum_{k\in\mathscr{K}} r_k}{P(\mathbf{w},\mathbf{r},\mathbf{s},\mathbf{x})} \tag{11.24a}$$

$$\text{subject to} \quad r_k \le \log(1+\gamma_k(\mathbf{w})),\ \forall k \in \mathscr{K} \tag{11.24b}$$

$$r_k \ge r_0,\ \forall k \in \mathscr{K} \tag{11.24c}$$

$$\sum_{k\in\mathscr{K}} x_{i,k} r_k \le \bar{C},\ \forall i \in \mathscr{R} \tag{11.24d}$$

$$\|\mathbf{w}_{i,k}\|_2 \le x_{i,k}\sqrt{\bar{P}}, x_{i,k} \in \{0,1\},\ \forall k \in \mathscr{K}, \forall i \in \mathscr{R} \tag{11.24e}$$

$$\sum_{i\in\mathscr{R}} \|\mathbf{w}_{i,k}\|_2^2 \le \bar{P},\ \forall i \in \mathscr{R} \tag{11.24f}$$

$$\|\tilde{\mathbf{w}}_{i,m}\|_2^2 \le P^{\text{ant}},\ \forall i \in \mathscr{R} \tag{11.24g}$$

$$s_i \ge x_{i,k}, \forall k \in \mathscr{K};\ s_i \le \sum_{k\in\mathscr{K}} x_{i,k},\ s_i \in \{0,1\}, \tag{11.24h}$$

where $\mathbf{x} \triangleq [x_{1,1},\ldots,x_{i,k},\ldots,x_{R,K}]^{\mathrm{T}}$, $\mathbf{s} = [s_1,\ldots,s_R]^{\mathrm{T}}$, and

$$P(\mathbf{w},\mathbf{r},\mathbf{s},\mathbf{x}) \triangleq \frac{1}{\varepsilon}\sum_{i\in\mathscr{R}}\sum_{i=1}^{M}\|\tilde{\mathbf{w}}_{i,m}\|_2 + (P^{\text{act}} - P^{\text{sl}})\sum_{i\in\mathscr{R}} s_i + \delta\sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} x_{i,k} r_k + P_{\text{c}}.$$

The new formulation is still a nonconvex optimization problem. However, (11.24) allows the application of global optimization framework such as the ones in Section 11.2. In the following sections, we first introduce an optimal algorithm based on the BRnB procedure to solve (11.24). Then efficient algorithms achieving near-optimal performance are presented; two of which are derived based on(11.24) and the other is based on (11.22).

11.4 Energy-efficient methods for cloud radio access networks

11.4.1 Globally optimal solution via BRnB algorithm

The BRnB algorithm cannot be applied to (11.24) since the optimization problem is not in a standard monotonic form. In particular, we can quickly observe that the fractional objective in (11.24a) is not an increasing function of the involved variables. Thus, the idea is to apply the BRnB on an alternative problem, which admits a monotonic

formulation while sharing the same optimal solution set with (11.24). To this end, we consider the following formulation:

$$\underset{\eta,\mathbf{w},\mathbf{x},\mathbf{s},\mathbf{r},\mathbf{t}}{\text{maximize}} \quad \eta \tag{11.25a}$$

$$\text{subject to} \quad \eta P(\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t}) - \sum_{k\in\mathcal{K}} r_k \leq 0 \tag{11.25b}$$

$$\sum_{m=1}^{M} ||\tilde{\mathbf{w}}_{i,m}||_2 \leq t_i, \ \forall i \in \mathcal{R} \tag{11.25c}$$

$$(11.24b)\text{–}(11.24h), \tag{11.25d}$$

where η and $\mathbf{t} \triangleq [t_1,\ldots,t_R]^{\mathrm{T}}$ are newly introduced variables and $P_{\text{total}}(\mathbf{w},\mathbf{r},\mathbf{s},\mathbf{x})$ is redefined as $P(\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t}) \triangleq \frac{1}{\varepsilon}\sum_{i\in\mathcal{R}} t_i + (P^{\text{act}} - P^{\text{sl}})\sum_{i\in\mathcal{R}} s_i + \delta\sum_{i\in\mathcal{R}}\sum_{k\in\mathcal{K}} x_{i,k} r_k + P_{\text{c}}$. We can easily justify that (11.25b) and (11.25c) hold equality at the optimality that guarantees the equivalence between (11.24) and (11.25) in terms of optimal solution set. The application of the BRnB to (11.25) is inspired by the following key observations: (i) objective (11.25a) and constraints (11.24c), (11.24d), (11.24h), and (11.25b) are monotone with respect to η, $\mathbf{x}$, $\mathbf{s}$, $\mathbf{r}$, and $\mathbf{t}$; (ii) given feasible values of $(\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t})$, we can determine the corresponding beamforming vector $\mathbf{w}$ by the following lemma.

Lemma 11.1. *Let* $(\hat{\mathbf{x}},\hat{\mathbf{s}},\hat{\mathbf{r}},\hat{\mathbf{t}},\hat{\mathbf{w}})$ *be a feasible point of (11.25). Given the values of* $(\hat{\mathbf{x}},\hat{\mathbf{s}},\hat{\mathbf{r}},\hat{\mathbf{t}})$*, then* $\hat{\mathbf{w}}$ *can be computed as*

$$\hat{\mathbf{w}} = \text{find}\{\mathbf{w}|(11.24b),(11.24e)\text{–}(11.24g),(11.25c)\}, \tag{11.26}$$

in which we replace $(\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t})$ *by* $(\hat{\mathbf{s}},\hat{\mathbf{x}},\hat{\mathbf{r}},\hat{\mathbf{t}})$.

The proof of Lemma 11.1 can be found in [10]. Based on these observations, we can develop a BRnB procedure to find optimal solution $(\mathbf{x},\mathbf{s},\mathbf{r},\mathbf{t},\eta)$ of (11.25). The optimal beamforming $\mathbf{w}$ can be derived accordingly by means of Lemma 11.1. This is the central idea of the proposed algorithm as described next.

To apply the BRnB method introduced in Section 11.2.1, we first determine the monotonic problem with respect to $(\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t},\eta)$, i.e.,

$$\underset{\eta,\mathbf{x},\mathbf{s},\mathbf{r},\mathbf{t}}{\text{maximize}} \quad \eta \tag{11.27a}$$

$$\text{subject to} \quad (11.24c),(11.24d),(11.24h),(11.25b) \tag{11.27b}$$

$$[\mathbf{s},\mathbf{x},\mathbf{r},\mathbf{t},\eta] \in [\mathbf{a},\mathbf{b}], \tag{11.27c}$$

in which $\mathbf{a} \triangleq [\underline{\mathbf{s}},\underline{\mathbf{x}},\underline{\mathbf{r}},\underline{\mathbf{t}},\underline{\eta}]$ and $\mathbf{b} \triangleq [\overline{\mathbf{s}},\overline{\mathbf{x}},\overline{\mathbf{r}},\overline{\mathbf{t}},\overline{\eta}]$ define the lower and upper bounds of the initial box. Vertices in $\mathbf{a}$ and $\mathbf{b}$ are calculated as follows. Since $\mathbf{s}$ and $\mathbf{x}$ are binary

variables, hence $\underline{s}_b = 0, \overline{s}_b = 1, \underline{x}_{b,k} = 0, \overline{x}_{b,k} = 1$. By (11.22c), we have $r_k \geq \underline{r}_k = r_0$, and by (11.22b)

$$r_k \leq \overline{r}_k = \min\left\{\bar{C}, \log\left(1 + |\mathbf{h}_k \mathbf{w}_k|^2/\sigma_k^2\right)\right\} \leq \min\left\{\bar{C}, \log\left(1 + R\bar{P}\|\mathbf{h}_k\|_2^2/\sigma_k^2\right)\right\},$$

as $|\mathbf{h}_k \mathbf{w}_k|^2 \leq \|\mathbf{h}_k\|_2^2 \|\mathbf{w}_k\|_2^2$ by the Cauchy–Schwarz inequality, and $\|\mathbf{w}_k\|_2^2 \leq R\bar{P}$. We also have $t_i \geq \underline{t}_i = 0$ and $t_i \leq \overline{t}_i = M\sqrt{P^{\text{ant}}}$; and by (11.25b)

$$\underline{\eta} \triangleq \frac{\sum_{k\in\mathscr{K}} \underline{r}_k}{P(\overline{\mathbf{s}}, \overline{\mathbf{x}}, \overline{\mathbf{r}}, \underline{\mathbf{t}})} \quad \text{and} \quad \overline{\eta} \triangleq \frac{\sum_{k\in\mathscr{K}} \overline{r}_k}{P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \overline{\mathbf{t}})}. \tag{11.28}$$

Since feasible set of (11.27) is upper bounded by the power and fronthaul constraints, it satisfies the normal and finite properties required by the BRnB algorithm (see Definition 11.3), thus, (11.27) is now in standard form of MO problem and can be solved by the BRnB method.

Remark that our aim is to find optimal solutions of (11.25) through solving (11.27). Therefore, we need to ensure that optimal solutions obtained by the BRnB method are feasible to (11.25), thereby the desired optimal solutions can be found by Lemma 11.1. To this end, every newly created box needs to be checked if potentially containing feasible solutions to (11.25), otherwise that box should be eliminated. This can be done by solving the following feasibility problem for a box V:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \sum_{i\in\mathscr{R}} \sum_{m=1}^{M} \|\tilde{\mathbf{w}}_{i,m}\|_2 \tag{11.29a}$$

$$\text{subject to} \quad \mathbf{h}_k \mathbf{w}_k \geq \sqrt{(e^{\underline{r}_k} - 1)\left(\sum_{j\neq k}^{K} |\mathbf{h}_k \mathbf{w}_j|^2 + \sigma_k^2\right)} \tag{11.29b}$$

$$\underline{s}_i \underline{t}_i \leq \sum_{m=1}^{M} \|\tilde{\mathbf{w}}_{i,m}\|_2 \leq \overline{s}_i \overline{t}_i, \ i \in \mathscr{R} \tag{11.29c}$$

$$\|\tilde{\mathbf{w}}_{i,m}\|_2^2 \leq \overline{s}_i P^{\text{ant}}, \ \|\mathbf{w}_{i,k}\|_2 \leq \overline{x}_{i,k}\sqrt{\bar{P}}, i \in \mathscr{R} \tag{11.29d}$$

$$\sum_{k\in\mathscr{K}} \|\mathbf{w}_{i,k}\|_2^2 \leq \overline{s}_i \bar{P} \ \forall i \in \mathscr{R}. \tag{11.29e}$$

Specifically, (11.29) can be viewed as minimizing the power consumption subject to minimum users' rate requirement $\underline{\mathbf{r}}$ given the power budget $\overline{\mathbf{t}}$ and RRH-user association $\overline{\mathbf{x}}$ and $\overline{\mathbf{s}}$. Obviously, if there does not exist any beamformer set making $\underline{\mathbf{r}}$ achievable, then V is infeasible to (11.25) and needs to be immediately discarded.

To summarize, the globally optimal method solving (11.25) is described in Algorithm 11.2. Noticeably, we have made modifications to the standard BRnB procedure in the algorithm so as to accelerate the convergence rate of Algorithm 11.2. These

Algorithm 11.2: The BRnB algorithm solving (11.25)

1: **Initialization:** Compute $\mathbf{a}$, $\mathbf{b}$ and apply box reduction to box $[\mathbf{a};\mathbf{b}]$. Let $n := 1$, $\mathscr{B}_1 = \mathrm{r}([\mathbf{a};\mathbf{b}])$ and $f_{\mathrm{UpB}}(B) = \underline{\eta}$
2: **repeat** $\{n := n + 1\}$
3: **Branching:** select a box $V_{\mathrm{c}} = [\mathbf{p};\mathbf{q}] \subset \mathscr{B}_{n-1}$ and branch V_{c} into two smaller ones V_{c}^1 and V_{c}^2, then remove V_{c} from $\mathscr{B}_{n-1}$.
4: **Reduction:** apply box reduction to each box V_{c}^l ($l = \{1, 2\}$) and obtain reduced box $\mathrm{r}(V_{\mathrm{c}}^l)$.
5: **Bounding:** for each box $\mathrm{r}(V_{\mathrm{c}}^l)$
6: **if** solving (11.29) is feasible **then**
7: Achieve $\mathbf{w}^*$, calculate $\mathbf{t}^*$ and extract $\mathbf{x}^*$.
8: Update $\underline{\mathbf{t}} := \mathbf{t}^*$ and calculate $f_{\mathrm{UpB}}(\mathrm{r}(V_{\mathrm{c}}^l))$ by (11.31).
9: Check $\mathbf{x}^*$ with (11.33), if true, compute feasible objective $\hat{\eta}$ as (11.32), and update new CBO, i.e., $\mathrm{CBO} := \max\{\hat{\eta}, \mathrm{CBO}\}$.
10: Update $\mathscr{B}_n := \mathscr{B}_{n-1} \cup \{\mathrm{r}(V_{\mathrm{c}}^l) | f_{\mathrm{UpB}}(\mathrm{r}(V_{\mathrm{c}}^l)) \geq \mathrm{CBO}\}$.
11: **end if**
12: **until** Convergence
13: **Output:** With $(\mathbf{x}^*, \mathbf{s}^*, \mathbf{r}^*, \mathbf{t}^*)$, recover $\mathbf{w}^*$ by (11.26) to achieve the globally optimal solution of (11.24), i.e., $(\mathbf{w}^*, \mathbf{x}^*, \mathbf{s}^*, \mathbf{r}^*)$.

changes exploit some useful properties of the energy efficiency maximization in the C-RAN. The modifications (compared to the generic framework) made in Algorithm 11.2 are presented next.

Improved branching dimension

Normally, one entry of $[\mathbf{s}, \mathbf{x}, \mathbf{r}, \mathbf{t}, \eta]$ is branched at each iteration, and, thus, the total number of iterations may increase exponentially with the problem size. Intuitively, we can lower the computational complexity by minimizing the branching dimensions while still guaranteeing convergence. For (11.25), it turns out that we can skip branching on η and $\mathbf{t}$. In particular, recall that lower and upper bounds of η can be determined via those of $[\mathbf{s}, \mathbf{x}, \mathbf{r}, \mathbf{t}]$ as (11.28). Thus, branching on η is not essential. On the other hand, consider problem (11.29) and denote by $\mathbf{w}^*$ the optimal solution if solving (11.29) is successful, and $\mathbf{t}^* \triangleq \{t_i^*\}_i$ with $t_i^* = \sum_{m=1}^{M} \|\tilde{\mathbf{w}}_{i,m}^*\|_2$. Obviously, $\mathbf{t}^*$ is the minimum power required to achieve $\underline{\mathbf{r}}$, and it holds $\underline{\mathbf{t}} \leq \mathbf{t}^*$. Also, $\mathbf{t}^*$ is unique solution because of the structure of the objective in (11.29) [37, Chapter 3]. At this point, we can replace $\underline{\mathbf{t}}$ by $\mathbf{t}^*$ to yield a tighter lower bound on $\mathbf{t}$. Thus, it is sufficient to only branch $(\mathbf{x}, \mathbf{s}, \mathbf{r})$ as the lower bound on $\mathbf{t}$ is always improved with $\underline{\mathbf{r}}$. The presented properties significantly accelerate the convergence of the BRnB.

Improved branching order

In principle, we can randomly select a variable to perform branching in each iteration of the BRnB algorithm, because this operation does not delete any feasible solution. For the considered problem, we can potentially reduce the computational complexity

if we start the algorithm by first branching on $\mathbf{s}$ due to its dependency on other factors. Intuitively, the number of active RRHs provides the degrees of freedom that can make the desired data rate $\underline{\mathbf{r}}$ achievable. Moreover, we can immediately obtain $x_{i,k} = 0, \forall k \in \mathcal{K}$ whenever $s_i = 0$, implying that the effective dimension in V is reduced by K times. Therefore, by first keeping branching on $\mathbf{s}$ until $\underline{\mathbf{s}} = \overline{\mathbf{s}}$, we can quickly determine that the combinations of $\{s_i\}_i$ make (11.25) infeasible. This is done by solving (11.29) with given $\overline{\mathbf{s}}$ and target rate r_0 for all users. Moreover, since the length of $\mathbf{s}$ is much smaller than that of $\mathbf{x}$ in most of wireless communications applications, it is expected that branching on $\mathbf{s}$ may take a relatively small number of iterations and, thus, may create a relatively small number of new boxes.

Improved upper bound

Upper bounding as in (11.28) is not tight enough and thus is inefficient for the considered problem. To be specific, let us recall the lower bound of $P(\mathbf{s}, \mathbf{x}, \mathbf{r}, \mathbf{t})$, i.e.:

$$P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \underline{\mathbf{t}}) = \frac{1}{\varepsilon} \sum_{i \in \mathcal{R}} \underline{t}_i + (P^{\text{act}} - P^{\text{sl}}) \sum_{i \in \mathcal{R}} \underline{s}_i + \delta \sum_{i \in \mathcal{R}} \sum_{k \in \mathcal{K}} \underline{x}_{i,k} \underline{r}_k + P_{\text{c}}, \tag{11.30}$$

and observe that $P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \underline{\mathbf{t}}) = \sum_{i \in \mathcal{R}} \underline{t}_i/\varepsilon + P_{\text{c}}$ if $\underline{\mathbf{x}} = \mathbf{0}$ and $\underline{\mathbf{s}} = 0$. However, it holds that $\sum_{i \in \mathcal{R}} s_i \geq 1$, since at least one RRH is active for transmission, and $\sum_{i \in \mathcal{R}} \sum_{k \in \mathcal{K}} x_{i,k} r_k \geq \sum_{k \in \mathcal{K}} \underline{r}_k$ because the fronthaul networks must at least convey the minimum required data rate for all users. In other words, the second and third terms in (11.30) are always nonzero. Moreover, from the result of solving (11.29) feasible, we have $\underline{\mathbf{t}} \leq \mathbf{t}^*$. Thus, we can upper tighten the bound of $P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \underline{\mathbf{t}})$ as

$$\begin{aligned} P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \underline{\mathbf{t}}) \leq \underline{P}(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \mathbf{t}^*) \triangleq & \sum_{i \in \mathcal{R}} \frac{1}{\varepsilon} t_i^* + (P^{\text{act}} - P^{\text{sl}}) \max \left\{ 1, \sum_{i \in \mathcal{R}} \underline{s}_i \right\} \\ & + \delta \max \left\{ \sum_{k \in \mathcal{K}} \underline{r}_k, \sum_{i \in \mathcal{R}, k \in \mathcal{K}} \underline{x}_{i,k} \underline{r}_k \right\} + P_{\text{c}}. \end{aligned}$$

One can see that replacing $P(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \underline{\mathbf{t}})$ by $\underline{P}(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \mathbf{t}^*)$ does not remove any feasible solution. As such a tighter upper bound on η over V can be recalculated as

$$\overline{\eta} = \frac{\sum_{k \in \mathcal{K}} \overline{r}_k}{\underline{P}(\underline{\mathbf{s}}, \underline{\mathbf{x}}, \underline{\mathbf{r}}, \mathbf{t}^*)}. \tag{11.31}$$

Updating the best current objective

As mentioned earlier in Section 11.2.1, boxes, the upper bounds of which are smaller than the CBO, can be removed out of set $\mathcal{B}_n$ to reduce the complexity. Thus, efficiently updating the CBO is deemed important for improving the algorithm's efficiency. To do so, we need to find a feasible point within each box in order to calculate the

feasible objective value. In particular, let $[\hat{\mathbf{s}}, \hat{\mathbf{x}}, \hat{\mathbf{r}}, \hat{\mathbf{t}}]$ be some feasible point in box V. The feasible objective is

$$\hat{\eta} = \frac{\sum_{k\in\mathcal{K}} \hat{r}_k}{P(\hat{\mathbf{s}}, \hat{\mathbf{x}}, \hat{\mathbf{r}}, \hat{\mathbf{t}})}. \tag{11.32}$$

If $\hat{\eta} >$ CBO, $\hat{\eta}$ is updated to be the new CBO. We now present a heuristic method to obtain a feasible point within box V. We first recall the feasibility problem (11.29) and note that $\underline{\mathbf{r}}$ is feasible if solving (11.29) is so. Thus, our idea is to set $\hat{\mathbf{r}} = \underline{\mathbf{r}}$ and determine the values of the other variables $(\hat{\mathbf{x}}, \hat{\mathbf{s}}, \hat{\mathbf{t}})$ corresponding to $\underline{\mathbf{r}}$. Particularly for $\hat{\mathbf{x}}$, we can simply compute $\hat{\mathbf{x}}$ by setting $\hat{x}_{i,k} = 0$ if $\|\mathbf{w}^*_{i,k}\|_2 = 0$ and vice versa $\hat{x}_{i,k} = 1$ if $\|\mathbf{w}^*_{i,k}\|_2 > 0$ where $\mathbf{w}^*$ is an optimal solution obtained by solving (11.29). The obtained solution is verified feasible if it stays in the feasible set of (11.27), i.e.,

$$\hat{\mathbf{x}} \in \left\{ \mathbf{x} \mid \sum_{i\in\mathcal{R}} x_{i,k} \geq 1, k \in \mathcal{K}, \sum_{k\in\mathcal{K}} x_{i,k}\underline{r}_k \leq \bar{C}, i \in \mathcal{R} \right\}. \tag{11.33}$$

The value of $\hat{\mathbf{s}}$ is found based on $\hat{\mathbf{x}}$ according to (11.24h), while the value of $\hat{\mathbf{t}}$ is $\hat{t}_i = \sum_{m=1}^{M} \|\tilde{\mathbf{w}}^*_{i,m}\|_2, \forall i \in \mathcal{R}$.

We can easily check that these modifications made in Algorithm 11.2 do not drop off any feasible solution. As a result, Algorithm 11.2 is still guaranteed to converge and its output is the globally optimal solutions of (11.24).

11.4.2 Suboptimal solutions via successive convex approximation

In general, a global optimization algorithm often takes enormous complexity to output a solution that is in fact not practically preferable. Instead, it often serves as the performance benchmark to evaluate the quality of other methods used in practice. Toward more efficient algorithms, we next present three iterative suboptimal approaches that aim at achieving near-optimal solutions but using much reduced computation. These approaches apply different techniques to deal the binary variables, before employing the SCA framework in Section 11.2.2 to find a stationary solution of the design problem. Noticeably, the first two methods are developed in accordance with formulation (11.24), while the third method is based on (11.22).

11.4.2.1 SCA-based mixed integer programming

In the first method, we keep the binary variables unchanged while using the SCA framework to convexify the nonconvex parts of problem (11.24) then solved using modern mixed integer convex solvers. We shall refer this algorithm as SCA-MI in the sequel.

To apply the SCA, recall that the nonconvexity must have a convex bound that satisfies the conditions in (11.8). However, this is not the case of (11.24). Thus, we first need to transform the problem into an equivalent form suitable for application of the SCA. To be simple, we shall consider the following transformation of (11.24) of

which the nonconvex parts have similar formulations with those introduced in Section 11.2.2, i.e.:

$$\underset{\substack{t,\mathbf{w},\mathbf{s},\mathbf{x},\\ \mathbf{r},\mathbf{g},\mathbf{q},\mathbf{v}}}{\text{maximize}} \quad \frac{\sum_{k\in\mathscr{K}} r_k}{t+\delta\sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} v_i^2} \tag{11.34a}$$

$$\text{subject to} \quad t \geq \frac{1}{\varepsilon}\sum_{i\in\mathscr{R}}\sum_{m=1}^{M}\|\tilde{\mathbf{w}}_{i,m}\|_2 + \sum_{i\in\mathscr{R}}(P^{\text{act}}-P^{\text{sl}})s_i + P_{\text{c}} \tag{11.34b}$$

$$\log(1+g_k) \geq r_k \ \forall k \in \mathscr{K} \tag{11.34c}$$

$$\sum_{j\neq k}^{K}|\mathbf{h}_k\mathbf{w}_j|^2 + \sigma_k^2 \leq \frac{|\mathbf{h}_k\mathbf{w}_k|^2}{g_k}, \ \forall k \in \mathscr{K} \tag{11.34d}$$

$$x_{i,k} \leq \frac{v_{i,k}^2}{r_k}, \ \forall k \in \mathscr{K}, \forall i \in \mathscr{R} \tag{11.34e}$$

$$\sum_{k\in\mathscr{K}} v_{i,k}^2 \leq \bar{C}, \ \forall i \in \mathscr{R} \tag{11.34f}$$

$$(11.24\text{c}), (11.24\text{e})\text{–}(11.24\text{h}), \tag{11.34g}$$

where t, $\mathbf{g} \triangleq [g_1,\ldots,g_K]^{\text{T}}$, $\mathbf{q} \triangleq [q_1,\ldots,q_K]^{\text{T}}$, $\mathbf{v} \triangleq [v_1,\ldots,v_R]^{\text{T}}$ are newly introduced slack variables. Specifically, the objective (11.24a) is equivalently rewritten by (11.34a) and (11.34b) with t; (11.24b) is rewritten by (11.34c) and (11.34d) using $\mathbf{g}$; (11.24d) is replaced by (11.34e) and (11.34f) using $\mathbf{v}$. It is easily checked that (11.24) and (11.34) are equivalent in terms of the optimal set. We keep rewriting (11.34) as

$$\underset{\boldsymbol{\theta}\in\mathscr{S}}{\text{minimize}} \quad f(\boldsymbol{\theta}) \triangleq \frac{t}{\sum_{k\in\mathscr{K}} r_k} + \frac{\delta\sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} v_i^2}{\sum_{k\in\mathscr{K}} r_k} \tag{11.35a}$$

$$\text{subject to} \quad (11.34\text{d}), (11.34\text{e}), \tag{11.35b}$$

where $\boldsymbol{\theta} \triangleq \{t, \mathbf{w}, \mathbf{s}, \mathbf{x}, \mathbf{r}, \mathbf{g}, \mathbf{q}, \mathbf{v}\}$ denotes the involved optimization variables and

$$\mathscr{S} \triangleq \{\boldsymbol{\theta} | (11.24\text{c}), (11.24\text{e})\text{–}(11.24\text{h}), (11.34\text{b}), (11.34\text{c}), (11.34\text{f})\},$$

is the set of convex constraints of (11.34). The nonconvexity of (11.35a) and (11.35b) is obvious due to the first term of objective and the right side of (11.34d) and (11.34e). These are the SCA-applicable functions (cf. Section 11.2.2). The application of the SCA to (11.35a) and (11.35b) is now straightforward. In particular, the nonconvex constraints can be replaced by the following convex ones:

$$\sum_{j\neq k}^{K}|\mathbf{h}_k\mathbf{w}_j|^2 + \sigma_k^2 \leq \phi^{\text{qol}}\left(\mathbf{w}_k, g_k; \mathbf{w}_k^{(n)}, g_k^{(n)}; \mathbf{h}_k^{\text{H}}\mathbf{h}_k\right) \tag{11.36}$$

$$x_{i,k} \leq \phi^{\text{qol}}\left(v_{i,k}, r_k; v_{i,k}^{(n)}, r_k^{(n)}\right), \ \forall i \in \mathscr{R}, k \in \mathscr{K}, \tag{11.37}$$

in which the right side of (11.34d) and (11.34e) have been approximated using (11.11). For the objective, the nonconvexity can be approximated by the function in (11.10). As a result, we have the following approximate objective:

$$f(\boldsymbol{\theta};\boldsymbol{\theta}^{(n)}) \triangleq \phi^{\text{frac}}\left(t; \sum_{k\in\mathscr{K}} r_k; \lambda^{(n)}\right) + \frac{\delta \sum_{i\in\mathscr{R}} \sum_{k\in\mathscr{K}} v_{i,k}^2}{\sum_{k\in\mathscr{K}} r_k}, \tag{11.38}$$

where $\lambda^{(n)} = \frac{1}{t^{(n)} \sum_{k\in\mathscr{K}} r_k^{(n)}}$.

In summary, we solve the following approximate convex program of (11.35a) and (11.35b) at every iteration of the SCA procedure outlined in Algorithm 11.3

$$\underset{\boldsymbol{\theta}\in\mathscr{S}}{\text{minimize}} \quad f(\boldsymbol{\theta};\boldsymbol{\theta}^{(n)}) \quad \text{subject to} \quad \{(11.36), (11.37)\}. \tag{11.39}$$

The convergence of Algorithm 11.3 can be proved following the arguments in Section 11.2.2 and those in the related references. Problem (11.39) is a mixed integer convex program that can be solved effectively by off-the-shelf convex solvers such as MOSEK [38].

It is interesting to see that most of constraints in (11.39) are second-order-cone (SOC) representable except the logarithmic constraint in (11.34c) (cf. [39] for SOC-presentable formulations). Thus, the complexity of solving (11.39) would remarkably be reduced if (11.34c) can be approximated by SOC ones. For this purpose, we can use approaches presented in [40]. By applying such method, the resulting problem becomes SOC programming (SOCP) that allows more available off-the-shelf convex solvers, such as CPLEX, GUROBI, applicable to solve it with lower complexity.

The SCA-MI uses a significantly smaller number of iterations to find the solutions compared to the BRnB method. In addition, it can achieve exact binary solutions. However, since the SCA-MI method deals directly with binary variables in each iteration, the per-iteration complexity is still relatively high, and so is the total run time, as we shall see via numerical demonstrations. In the following, we present the second and third approaches that have further lower per-iteration complexity.

Algorithm 11.3: Proposed method for solving (11.34)

1: **Initialization:** Set $n := 0$, choose initial values for $\boldsymbol{\theta}^{(0)}$
2: **repeat** $\{n := n + 1\}$
3: Solve (11.39) and achieve $\boldsymbol{\theta}^*$
4: Update $\boldsymbol{\theta}^{(n)} := \boldsymbol{\theta}^*$
5: **until** Convergence

11.4.2.2 SCA-based regularization method

The second method also aims at using the SCA framework to solve (11.24), achieving exact binary solutions. Unlike the SCA-MI method, the binary variables here are equivalently represented by a constraint of continuous variables. A regularization technique is then applied to make the resulting problem to be SCA applicable. More specifically, we first use the well-known polynomial constraint of binary variables for $\mathbf{x}$ which is given by

$$x_{i,k} \in \{0,1\}, \forall i,k \Leftrightarrow \sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} x_{i,k}^2 - x_{i,k} \geq 0,\ x_{i,k} \in [0,1]. \tag{11.40}$$

We can adjust the previous representation by the fact that $x_{i,k}^2 - x_{i,k} < 0$ for $x_{b,k} \in (0,1)$. For $\mathbf{s}$, we simply relax it as $s_i \in [0,1]$ because s_i is automatically binary when $x_{b,k}$ is so, which is due to (11.24h). In the sequel of this subsection, $x_{i,k}$ and s_i are seen as continuous over $[0,1]$, $\forall i,k$. Now, to apply the SCA method, we can utilize the transformation presented in the previous subsection which leads to the following equivalent problem of (11.24):

$$\underset{\boldsymbol{\theta}\in\mathscr{S}}{\text{minimize}}\ f(\boldsymbol{\theta}) \quad \text{subject to } \{x_{i,k}, s_i \in [0,1],\ (11.40), (11.34\text{d}), (11.34\text{e})\}, \tag{11.41}$$

while two nonconvex constrains (11.34d) and (11.34e) can be handled by the same way as done in Section 11.4.2.1. Unfortunately, (11.40) is not suitable for the use of the SCA. This is because its convex approximation does not satisfy Slater's constraint qualification, which is the necessary condition for the SCA. To understand this, let us apply the SCA principle to (11.40) resulting in the following approximate constraint:

$$\phi^{\text{po}}(\mathbf{x};\mathbf{x}';2) \triangleq \sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} \left(2x_{i,k}x'_{i,k} - (x'_{i,k})^2\right) - \sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} x_{i,k} \geq 0, \tag{11.42}$$

where $x'_{i,k} \in [0,1]$ is a feasible point in (11.41). It is not difficult to check that the set $\{x_{i,k} \in (0,1) | \phi^{\text{po}}(\mathbf{x};\mathbf{x}';2) > 0\}$ is empty violating Slater's condition. To cope with this issue, we can apply the regularization technique [41,42]. In particular, auxiliary variable $q \geq 0$ is added to (11.41) to arrive at the following regularized problem:

$$\underset{\boldsymbol{\theta}\in\mathscr{S}}{\text{minimize}}\ f(\boldsymbol{\theta}) + \alpha q \tag{11.43a}$$

$$\text{subject to } \sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} x_{i,k}^2 - x_{i,k} \geq q,\ x_{i,k}, s_i \in [0,1] \tag{11.43b}$$

$$(11.34\text{d}), (11.34\text{e}). \tag{11.43c}$$

In (11.43), the last term in the objective stands for the cost that (11.43b) is violated when $x_{i,k} \neq \{0,1\}$, which is the objective to be minimized; $\alpha > 0$ is the penalty parameter. We can justify that the use of q allows (11.43b) to be satisfied for any

$x_{i,k} \in [0,1]$, and (11.43) is identical with (11.41) when $q = 0$. At this point, the SCA algorithm can be applied to (11.43) similarly to Section 11.4.2.1, leading to the following convex approximation problem solved at each iteration:

$$\underset{\boldsymbol{\theta} \in \mathscr{S}, q}{\text{minimize}} \quad f(\boldsymbol{\theta}; \boldsymbol{\theta}^{(n)}) + \alpha q \tag{11.44a}$$

$$\text{subject to} \quad \sum_{i \in \mathscr{R}} \sum_{k \in \mathscr{K}} \left(2 x_{i,k} x_{i,k}^{(n)} - (x_{i,k}^{(n)})^2 \right) - \sum_{i \in \mathscr{R}} \sum_{k \in \mathscr{K}} x_{i,k} \geq q, \; x_{i,k}, s_i \in [0,1] \tag{11.44b}$$

$$(11.36), (11.37). \tag{11.44c}$$

We outline the procedure solving (11.43) in Algorithm 11.4. The convergence of the method can be justified following the discussion in Section 11.2.2. An important point in Algorithm 11.4 is that the value of penalty parameter α is increased at each iteration, i.e., Step 5. We note that a high value of α will encourage q to be zero which means that $x_{i,k}$ $\forall i, k$ are binary values. Thus, the idea is to start Algorithm 11.4 with a small value of α to prioritize minimizing the original objective and then increase α in subsequent iterations to force q to be zero.

Remark: For Algorithm 11.4, our expectation is that the obtained solutions at the convergence shall eventually be exact binary values. To ensure this, we can replace (11.24e) by

$$\|\mathbf{w}_{i,k}\|_2 \leq x_{i,k}^{p} \sqrt{\bar{P}}, \tag{11.45}$$

where p is a constant. This maneuver is inspired by two observations. First, for all value $p > 1$, (11.45) is equivalent to (11.24e) if $x_{b,k} \in \{0,1\}$. Second, $x_{i,k}^{p} \geq x_{i,k}^{p+1}$ for $x_{i,k} \in [0,1]$ with $p > 0$ which means that:

$$\|\mathbf{w}_{i,k}\|_2 \leq x_{i,k}^{p+1} \sqrt{\bar{P}} \leq x_{i,k}^{p} \sqrt{\bar{P}} \leq \cdots \leq x_{i,k} \sqrt{\bar{P}}. \tag{11.46}$$

Thus, this replacement is nothing but to tighten the feasible set of (11.41) such that a tighter continuous relaxation can be obtained with higher values of p. Remark that (11.45) for $p > 1$ is nonconvex constraints and its right side is the power function.

Algorithm 11.4: Proposed method for solving (11.43)

1: **Initialization:** Set $n := 0$, choose initial values for $\boldsymbol{\theta}^{(0)}$ and set $\alpha^{(0)}$ small
2: **repeat** $\{n := n + 1\}$
3: Solve (11.44) and achieve $\boldsymbol{\theta}^{*}$
4: Update $\boldsymbol{\theta}^{(n)} := \boldsymbol{\theta}^{*}$
5: Update $\alpha^{(n)} := \min\{\alpha_{\max}; \alpha^{(n-1)} + \varepsilon\}$ for small ε
6: **until** Convergence

In light of the SCA principle, we can replace (11.45) by the following approximate constraint when solving (11.44) (Step 3 in Algorithm 11.4), i.e.,

$$\|\mathbf{w}_{i,k}\|_2 \leq \phi^{\text{po}}\left(x_{i,k}; x_{i,k}^{(n)}; p\right)\sqrt{\bar{P}}. \tag{11.47}$$

We note that (11.45) has not been used in the development of the global optimization algorithm and the SCA-MI method, since these methods do not work on the continuous relaxation of the binary variables.

11.4.2.3 SCA-based ℓ_0-approximation method

In the third method, we use the formulation (11.22) to derive solutions for the joint design problem. By viewing the joint design as finding sparse solutions of the beamformers, we can leverage the sparsity-based approach combining with SCA method to solve the considered problem. In particular, the nonsmooth function $\mathbb{I}(\cdot)$ is first approximated by a continuous one using ℓ_0/ℓ_2 norm method [25,43]. As such the SCA for the continuous optimization problem can be applied.

To proceed, we consider the simplified expression of (11.15)

$$\mathbb{I}\left(\|\mathbf{w}_{i,k}\|_2^2\right) \Leftrightarrow \mathbb{I}(u_{i,k}); \mathbb{I}\left(\sum_{k\in\mathscr{K}}\|\mathbf{w}_{i,k}\|_2^2\right) \Leftrightarrow \mathbb{I}\left(\sum_{k\in\mathscr{K}} u_{i,k}\right), \text{ where } \|\mathbf{w}_{i,k}\|_2 \leq u_{i,k}, \tag{11.48}$$

and $u_{i,k}$ is the slack variable associated with the power of $\mathbf{w}_{i,k}$. We justify (11.48) by the fact that $u_{i,k} = 0$ if $\|\mathbf{w}_{i,k}\| = 0$ and otherwise. We note that $u_{i,k}$ is additionally introduced merely to make the presentation of the algorithm easier to follow, as one can directly apply the method presented next on $\mathbb{I}(\|\mathbf{w}_{i,k}\|_2^2)$ to find the solutions. Following ℓ_0/ℓ_2 norm method, $\mathbb{I}(u_{i,k})$ can be continuously approximated by some (concave) functions. For example, Table 11.1 lists continuous approximations of the indicator function that have often been used in sparse optimization problems in wireless communications. We can see that $\mathbb{I}(u_{i,k}) \approxeq \psi_\beta(u_{i,k})$ when approximation

Table 11.1 ℓ_0-Approximation function $\psi_\beta(y)$ and the corresponding subgradient $\partial\psi_\beta(y)$ and first-order approximations $\bar{\psi}_\beta(y;y^{(n)})$ [23]

	Function $\boldsymbol{\psi_\beta(y)}$	**Subgradient $\boldsymbol{\partial\psi_\beta(y)}$**	**Approximation $\boldsymbol{\bar{\psi}_\beta(y;y^{(n)})}$**
Exp	$1-\exp(-\beta y)$	$\beta e^{-\beta y}$	$1-e^{-\beta y^{(n)}}(1-\beta(y-y^{(n)}))$
Capped-ℓ_1	$\min\{1,\beta y\}$	$\begin{cases}0 & \text{if } y \geq \frac{1}{\beta}\\ \beta & \text{if otherwise}\end{cases}$	$\begin{cases}1 & \text{if } y^{(n)} \geq \frac{1}{\beta}\\ \beta y & \text{otherwise}\end{cases}$

parameter β is sufficiently *large*. Replacing $\mathbb{I}(u_{i,k})$ by $\psi_\beta(u_{i,k})$, we arrive at the following continuously approximated problem (11.22):

$$\underset{\mathbf{w},\mathbf{r},\mathbf{u}}{\text{maximize}} \quad \frac{\sum_{k\in\mathscr{K}} r_k}{P(\mathbf{w},\mathbf{r},\mathbf{u})} \tag{11.49a}$$

$$\text{subject to} \quad \|\mathbf{w}_{i,k}\|_2 \le u_{i,k},\ \sum_{k\in\mathscr{K}} u_{i,k}^2 \le \bar{P},\ \forall b \in \mathscr{B} \tag{11.49b}$$

$$\sum_{k\in\mathscr{K}} \psi_\beta(u_{i,k}) r_k \le \bar{C} \tag{11.49c}$$

$$(11.22\text{b}), (11.22\text{c}), (11.22\text{f}), \tag{11.49d}$$

where $P(\mathbf{w},\mathbf{r},\mathbf{u}) = \sum_{i\in\mathscr{R}}(\frac{1}{\varepsilon}\sum_{i=1}^{M}\|\tilde{\mathbf{w}}_{i,m}\|_2 + (P^{\text{act}} - P^{\text{sl}})\psi_\beta(\sum_{k\in\mathscr{K}} u_{i,k}) + \delta\sum_{k\in\mathscr{K}} \psi_\beta(u_{i,k})r_k) + P_{\text{c}}$. We note that (11.49) is still nonconvex but can be solved following the same manner with those in the previous subsections. In particular, we first recall the transformation:

$$\underset{t,\mathbf{w},\mathbf{r},\mathbf{g},\mathbf{u},\mathbf{v},\tilde{\mathbf{x}},\tilde{\mathbf{s}}}{\text{minimize}} \quad \frac{t}{\sum_{k\in\mathscr{K}} r_k} + \frac{p_{\text{SP}}\sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}} v_i^2}{\sum_{k\in\mathscr{K}} r_k} \tag{11.50a}$$

$$\text{subject to} \quad \tilde{x}_{i,k} \ge \psi_\beta(u_{i,k}),\ \forall i \in \mathscr{R}, k \in \mathscr{K} \tag{11.50b}$$

$$\tilde{s}_i \ge \psi_\beta(\textstyle\sum_{k\in\mathscr{K}} u_{i,k}),\ \forall i \in \mathscr{R} \tag{11.50c}$$

$$t \ge \frac{1}{\varepsilon}\sum_{i\in\mathscr{R}}\sum_{k\in\mathscr{K}}\|\tilde{\mathbf{w}}_{i,m}\|_2 + \sum_{i\in\mathscr{R}}(P^{\text{act}} - P^{\text{sl}})\tilde{s}_i + P_{\text{c}} \tag{11.50d}$$

$$\tilde{x}_{i,k} \le \frac{v_{i,k}^2}{r_k},\ \forall i \in \mathscr{R}, k \in \mathscr{K} \tag{11.50e}$$

$$\sum_{k\in\mathscr{K}} v_{i,k}^2 \le \bar{C},\ \forall i \in \mathscr{R} \tag{11.50f}$$

$$(11.22\text{c}), (11.22\text{f}), (11.34\text{c}), (11.34\text{d}), (11.49\text{b}). \tag{11.50g}$$

In (11.50), $\tilde{\mathbf{x}} \triangleq [\tilde{x}_{1,1},\ldots,\tilde{x}_{i,k},\ldots,\tilde{x}_{R,K}]^{\text{T}}, \mathbf{s} = [\tilde{s}_1,\ldots,\tilde{s}_R]^{\text{T}}$ are additionally introduced to facilitate the transformation such that (11.50) has similar form with those nonconvex programs in Sections 11.4.2.1 and 11.4.2.2. Second, we apply the SCA to solve the continuous problem (11.50). Observing that the nonconvexity of (11.50) is due to the constraints (11.50b), (11.50c), (11.34d), (11.50e), and the objective. For the first two constraints, they can be convexified as

$$\tilde{x}_{i,k} \ge \bar{\psi}_\beta\left(u_{i,k}; u_{i,k}^{(n)}\right) \tag{11.51}$$

$$\tilde{s}_i \ge \bar{\psi}_\beta\left(\textstyle\sum_{k\in\mathscr{K}} u_{i,k}; \sum_{k\in\mathscr{K}} u_{i,k}^{(n)}\right), \tag{11.52}$$

respectively, where $\bar{\psi}_\beta(\cdot)$ is the first-order approximation provided in Table 11.1. The remaining parts of the nonconvexity are approximated as done in (11.36), (11.37),

and (11.38), respectively. Finally, we arrive at the approximate convex program of problem (11.50), i.e.:

$$\text{minimize}_{\tilde{\boldsymbol{\theta}}\in\tilde{\mathscr{S}}} \quad f(\tilde{\boldsymbol{\theta}};\tilde{\boldsymbol{\theta}}^{(n)}) \quad \text{subject to} \quad \{(11.36),(11.37),(11.51),(11.52)\}, \tag{11.53}$$

where $\tilde{\boldsymbol{\theta}} \triangleq \{t, \mathbf{w}, \mathbf{r}, \mathbf{v}, \mathbf{g}, \mathbf{u}, \tilde{\mathbf{x}}, \tilde{\mathbf{s}}\}$ includes involved optimization variables and

$$\tilde{\mathscr{S}} \triangleq \{\tilde{\boldsymbol{\theta}} | (11.22c),(11.22f),(11.50d),(11.34c),(11.50f),(11.49b)\},$$

defines the convex parts of (11.50). We describe the third suboptimal method in Algorithm 11.5. Similarly to Algorithm 11.4, parameter β is also updated after each iteration. The idea is the same as β is viewed to provide the tightness of the binary approximation $\psi_\beta(\cdot)$.

Convergence of Algorithm 11.5 is guaranteed following the discussion in Section 11.2.2. We remark that if the Capped-ℓ_1 function is adopted in the algorithm, the achieved stationary point is not ensured to hold the first-order optimality of (11.50) since the Capped-ℓ_1 function is not smooth.

11.4.3 Complexity analysis of the presented optimization algorithms

We now estimate the computational complexity of the optimization algorithms presented in this section. In particular, Table 11.2 provides the worst case per-iteration complexity of Algorithms 11.2–11.5. Here, we remark that the complexity of

Algorithm 11.5: Proposed method for solving (11.50)

1: **Initialization:** Set $n := 0$, choose initial values for $\tilde{\boldsymbol{\theta}}^{(0)}$ and set $\beta^{(0)}$ small
2: **repeat** $\{n := n + 1\}$
3: Solve (11.53) and achieve $\tilde{\boldsymbol{\theta}}^{*}$
4: Update $\tilde{\boldsymbol{\theta}}^{(n)} := \tilde{\boldsymbol{\theta}}^{*}$
5: Update $\beta^{(n)} := \min\{\beta_{\max}; \beta^{(n-1)} + \varepsilon\}$ for small ε
6: **until** Convergence and output $\tilde{\boldsymbol{\theta}}^{*}$

Table 11.2 Estimated worst case per-iteration complexity of the optimization algorithms presented in Section 11.2

	Solved optimization program	**Per-iteration complexity**
Algorithm 11.2	SOCP	$\mathscr{O}(\sqrt{R(K+M)}R^3K^3M^3)$
Algorithm 11.3	MI-SOCP	$\mathscr{O}(2^{RK+R}\sqrt{R(K+M)}R^3K^3M^3)$
Algorithm 11.4	SOCP	$\mathscr{O}(\sqrt{R(K+M)}R^3K^3M^3)$
Algorithm 11.5	SOCP	$\mathscr{O}(\sqrt{R(K+M)}R^3K^3M^3)$

Algorithm 11.2 is dominated by the step of solving the feasibility problem (11.29). For suboptimal methods, the approximated programs solved at each iteration are general convex programs but can be represented as SOCPs as discussed in [10,40]. Recall that an SOCP can be solved by path-following interior-point method in most of the available off-the-shelf convex solvers. Therefore, the complexity of the presented optimization procedures can be estimated following [44, Lecture 6.6]. It is worth noting that Algorithm 11.3 deals with the MI-SOCP (11.39) in each iteration. Indeed, MI-SOCP is a combinatorial optimization problem. In some powerful convex solvers such as MOSEK, the MI-SOCP can be solved by implementing conic branch-and-cut and conic outer approximation frameworks [38]. Thus, its complexity is scaled exponentially with the length of the involved Boolean variables and the size of the solved SOCP.

We can see that Algorithms 11.2, 11.4, and 11.5 have similar computational cost for each iteration. This is because they all deal with beamforming variables. This operation dominates the size of solved problems, i.e., the length of $\mathbf{w}$ is significantly larger than the size of the other optimization variables. However, the use of Algorithm 11.2 is prohibited in practice due to the extremely high number of iterations required as will be demonstrated in the numerical experiments. For Algorithm 11.3, its complexity is higher than that of the other two suboptimal methods, because it needs to handle the Boolean variables. Thus, Algorithm 11.3 is more suitable for problems of small and moderate sizes. On the other hand, Algorithms 11.4 and 11.5 can be applied for solving large-scale problems.

11.5 Numerical examples

We consider a network with $R = 3$ RRHs, each is equipped with $M = 2$ antennas, and $K = 4$ single-antenna users. The RRHs are placed at the coordinates $(-100, 0)$, $(100, 0)$, and $(0, 100\sqrt{3})$ m. Users are randomly distributed in the area covered by all RRHs. We use the path loss model for the small-cell BS, i.e., $30 \log_{10}(D_{i,k}) + 38 + \mathcal{N}(0, 8)$ in dB where $D_{i,k}$ is the distance in meters between RRH i and user k. The operating bandwidth is 10 MHz and the noise power density is -143 dBW. We set $\bar{P} = MP^{\text{ant}}$, $\varepsilon_{\max} = 0.55$ [34] and $\delta = 0.5$ (W/Mbits/s). The minimum required data rate for each user is $r_0 = 1$ bit/s/Hz. For the penalty parameters, we take $\alpha_{\max} = 10^3$, $\beta_{\max} = 10^6$ and initialize $\alpha^{(0)} = 10^{-3}$ and $\beta^{(0)} = 0.1$. All the convex programs in this chapter are solved by MOSEK solver in MATLAB® environment [38].

11.5.1 Convergence results

The first set of experiments demonstrates the convergence behavior of the presented optimization algorithms. We first show the convergence of the globally optimal method, i.e., Algorithm 11.2, for one channel realization in Figure 11.3. In particular, the figure plots the upper bound returned when Algorithm 11.2 proceeds for a random channel realization. It is seen that the upper bound monotonically converges to the optimal value. In the figure, we also illustrate the convergence of the BRnB procedure

without applying the presented modifications in Algorithm 11.2. As expected, the result clearly shows that the modifications provide remarkable improvement to the convergence speed of the BRnB.

The convergences of the SCA-based suboptimal methods are illustrated in Figure 11.4, where we plot the sequence of energy efficiency objectives obtained by Algorithms 11.3, 11.4, and 11.5 over iterations for two random channel realizations. For comparison purposes, the optimal objective value from the BRnB method for each case of considered channel is also provided. As can be seen, the three suboptimal approaches are able to converge to a very close optimal performance while requiring much smaller number of iterations compared to the globally optimal method. We also note that the SCA procedure in Algorithms 11.4 and 11.5 does not show monotonic convergence behavior as first few iterations. This is because the impact of the updating the parameters in Algorithms 11.4 and 11.5, which is specifically discussed in the

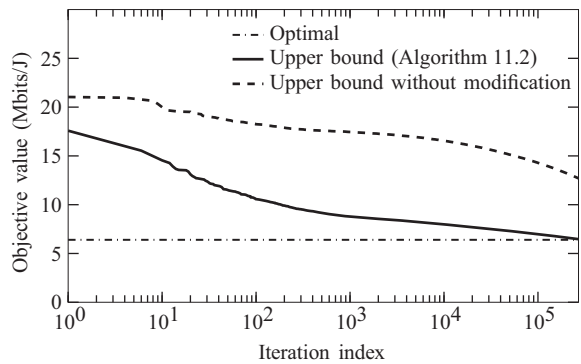


Figure 11.3 Convergence behavior of Algorithm 11.2 for one channel realization with $\bar{P} = 27$ dBm, $\bar{C} = 200$ Mbits/s

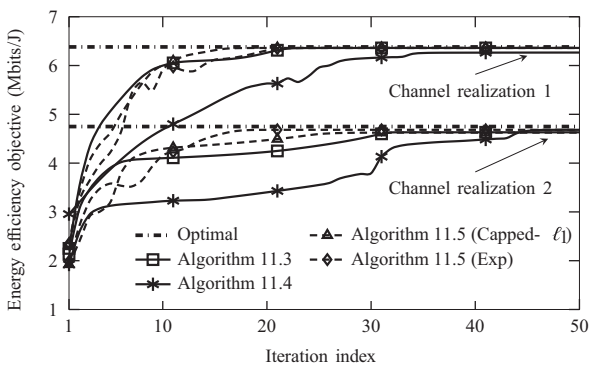


Figure 11.4 Convergence behavior of the SCA-based suboptimal algorithms for two random channel realizations with $\bar{P} = 27$ dBm, $\bar{C} = 200$ Mbits/s

context of the next figure. In fact, when the parameters are fixed after some number of iterations (i.e., $\alpha^{(n)} = \alpha_{\max}$ for Algorithm 11.4 and $\beta^{(n)} = \beta_{\max}$ for Algorithm 11.5), these approaches perform the monotonic convergence.

Figure 11.5 illustrates the convergence of Algorithms 11.4 and 11.5 when the parameters are fixed, i.e., $\alpha^{(n)} = \alpha_{\max}$ and $\beta^{(n)} = \beta_{\max}$ $\forall n$, respectively. Note that the result is plotted using Channel realization 2 in Figure 11.4 for the ease of comparison. We can see that the considered methods converge quickly and monotonically (compared to the result in Figure 11.4), but the achieved performances are far from the optimal values. On the other hand, increasing the value of the parameters over the iterations uses more iterations. However, doing so provides a better performance.

To compare the suboptimal methods in terms of time complexity, we provide in Table 11.3 the average per-iteration and total run time of Algorithms 11.3, 11.4, and 11.5. As expected, the SCA-MI requires significantly higher per-iteration run time compared to the other methods and so does the total run time. This is because the SCA-MI in fact still deals with the combinatorial optimization problem and, thus, has high computational complexity, whereas the other two approaches only need to solve the continuous one. We can also see that Algorithms 11.4 and 11.5 have approximately

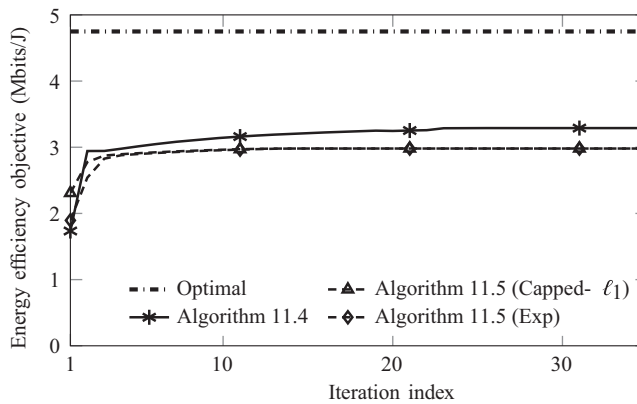


Figure 11.5 Convergence of Algorithms 11.4 and 11.5, when the parameters are fixed, for one random channel realization with $\bar{P} = 27$ dBm, $\bar{C} = 200$ Mbits/s

Table 11.3 Comparison on average per-iteration and total run time (s) of the SCA-based suboptimal methods with $\bar{P} = 27$ dBm, $\bar{C} = 200$ Mbits/s

Run time (s)	**Algorithm 11.3**	**Algorithm 11.4**	**Algorithm 11.5**
Per-iteration	0.9	0.03	0.02
Total	25	2	1.5

similar run time complexity. This result is understandable since they solve the problem with the similar size of involved variables.

11.5.2 Energy efficiency performance

The second set of experiments illustrates the energy efficiency performance of the C-RAN. For comparison purposes, we provide the achieved energy efficiency of the existing design that is the sum rate maximization for C-RAN in [9], dubbed as maxSR. In addition, the optimal energy efficiency of the design without applying RRH-user association and RRH selection (i.e., full cooperation scenario) is also plotted to demonstrate the significance of these selection schemes.[†]

Figure 11.6 shows the average energy efficiency versus the different setting of transmit power $\bar{P}$. Our first observation is that, in the average sense, the performances achieved by the suboptimal algorithms are relatively close to that of the optimal one in all cases of $\bar{P}$. We can also see that the energy efficiency performances achieved by the introduced methods first increase and then, after some point, decrease. The result can be explained as follows. In an energy efficiency maximization problem, when we increase the transmit power in the small value region of $\bar{P}$, the increase in the data rate often outweighs the increase of the power consumption. On the other hand, when $\bar{P}$ is large enough, the energy efficiency optimization mechanism tends not to use all available transmit power. For the linear model of PA's efficiency, the optimized transmit power and data rate would remain unchanged as $\bar{P}$ increases beyond a certain

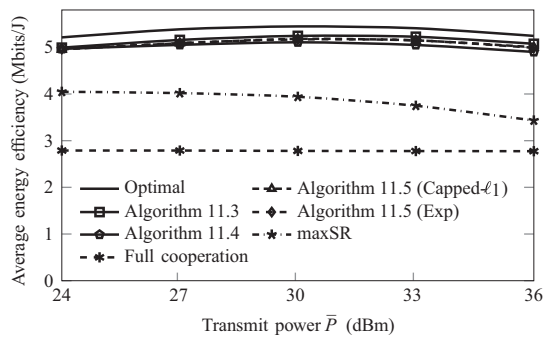


Figure 11.6 Average energy efficiency performances of the considered schemes versus the maximum transmit power with $\bar{C} = 200$ Mbits/s

[†]In the full cooperation scenario, we solve problem (11.24) but with variables $\mathbf{x}$ and $\mathbf{s}$ being fixed as $\mathbf{x} = \mathbf{1}$ and $\mathbf{s} = \mathbf{1}$. In this regard, the fronthaul constraints in (11.24d) are replaced by the following one:

$$\sum_{k \in \mathcal{K}} r_k \leq \bar{C}.$$

The performance of this scheme is found by applying the BRnB algorithm to the resulting problem.

value [45,46]. However, for the considered nonlinear model, a larger value of $\bar{P}$ reduces the PA's efficiency due to the nonlinear behavior of the PA as can been seen by (11.18). This leads to the significant increase in the power consumption on PA, and, thus, downgrades the energy efficiency. For the maxSR scheme in [9], the energy efficiency performance monotonically decreases versus $\bar{P}$. This is understandable, since the sum rate performance of the maxSR is constrained by the fronthaul capacity $\bar{C}$. Thus, increasing $\bar{P}$ does not result in the sum rate improvement when $\bar{C}$ is fixed but leads to the increase of the power consumption on the PAs, as discussed earlier. On the other hand, the performance of the full cooperation scheme is inferior to the others. To explain this, we recall that the optimized sum data rate for the full cooperation transmission scheme is less than or equal to $\bar{C}$. As a result, the achieved sum rate is small when $\bar{C}$ is so, and thus the optimized transmit power is small and far from the maximum RF output power of the PAs. This implies the low efficiency of the PAs resulting in high power dissipated on the RF chains. We shall see via the next experiment that energy efficiency of the full cooperation scheme increases significantly when $\bar{C}$ grows sufficiently large.

In Figure 11.7, we plot the average energy efficiency performances versus the maximum fronthaul capacity $\bar{C}$. Similarly to Figure 11.6, we can observe a small gap between the performance of the optimal and suboptimal schemes. Another important observation is that the energy efficiency increases as $\bar{C}$ increases for all considered schemes and becomes saturated when $\bar{C}$ is sufficiently large. The result can be understood as follows. Recall that the capacity-finite fronthaul links restrict the number of users that can be served by an RRH. In the small value region of $\bar{C}$, increasing $\bar{C}$ allows each RRH serving a larger set of users, which improves the cooperation among the RRHs, i.e., more users can be served by multiple RRHs. This increases the system cooperation gain and so does the system performance. When the fronthaul capacity is large enough such that the additional cooperation gain provides no improvement in the achieved performance, increasing $\bar{C}$ does not change the performance. For the full cooperation scheme, the energy efficiency increases proportionally with $\bar{C}$ as expected. However, in all cases of $\bar{C}$, its achieved performance is far from that of the

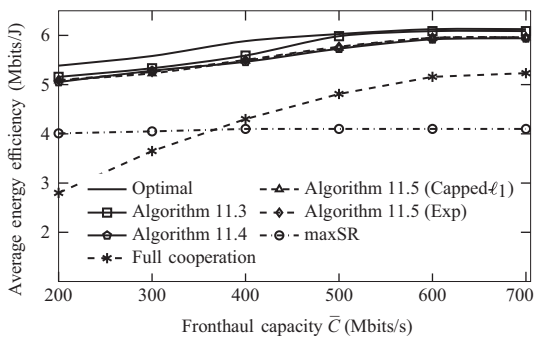


Figure 11.7 Average energy efficiency performances of the considered schemes versus the maximum fronthaul capacity with $\bar{P} = 27$ dBm

introduced methods. This result shows the benefit of the RRH-user association and RRH selection in the fronthaul-constrained C-RAN designs.

11.6 Conclusion

In the C-RANs, the energy-efficient transmission strategies necessarily involve the RRH-user association and RRH selection schemes to deal with the constraints of the limited fronthaul capacity and to boost the energy efficiency. As such, the energy efficiency optimization needs to deal with mixed integer nonconvex programs of which optimal solutions are generally difficult to find. We have introduced a globally optimal method via MO framework to access the optimal energy efficiency performances of the considered C-RAN design. We have also discussed the possible modifications on the generic global optimization framework to improve the optimal algorithm's efficiency, which are based on exploiting the specific properties of the C-RAN. The presented global optimization methods can serve as benchmarks for the energy efficiency optimization approaches in the C-RANs. For practically appealing methods, we have discussed three suboptimal approaches via the SCA procedure. The presented methods have been numerically shown to achieve very close to optimal performance with much reduced complexity.

The introduced optimization algorithms in this chapter provide the design guidelines for not only the energy efficiency maximization, but also other resource allocation problems in the C-RANs. Beyond the C-RAN architectures, many of the presented methods are readily applicable to other wireless communications design problems as well.

Some major challenges in the C-RAN need to be addressed in future research [6,7,47]. One of the main practical issues is the requirement on strict phase-synchronization among cooperative RRHs for coherent transmission. This poses a barrier for practical C-RAN designs, especially in ultradense C-RANs where a very large number of RRHs are co-deployed. A promising solution for this issue is to adopt noncoherent transmission to C-RAN [48]. Another challenge for practical implementation of C-RAN is to meet the requirement of latency. In particular, the limited capacity of the fronthaul links may cause a high latency as 5G RAN is expected to transport a huge amount of data. Therefore, the BB signal rate requirement on the fronthaul should be reduced. This could be done, e.g., by functional splitting such that parts of BB signal-processing functionalities can be done at the RRHs and only the really needed ones at the BBU [6,47].

References

[1] NGMN Alliance. 5G white paper. NGMN Alliance; 2015.
[2] Ericsson. Ericsson mobility report. Ericsson, EAB-18:014144 Uen; 2019.
[3] Mobile C. C-RAN: the road towards green RAN. White Paper. 2011;2.

[4] Rost P, Bernardos CJ, Domenico AD, *et al.* Cloud technologies for flexible 5G radio access networks. IEEE Commun Mag. 2014;52(5):68–76.
[5] Hossain E and Hasan M. 5G cellular: key enabling technologies and research challenges. IEEE Instrum Meas Mag. 2015;18(3):11–21.
[6] 5G PPP Architecture Working Group. View on 5G architecture. White Paper; 2016.
[7] Nokia. 5G immersive service opportunities with Edge Cloud and Cloud RAN. White Paper; 2019.
[8] Ericsson. Virtualized 5G RAN: why, when and how? Ericsson. 2020 [Online]. https://www.ericsson.com/en/blog/2020/2/virtualized-5g-ran-why-when-and-how.
[9] Dai B and Yu W. Sparse beamforming and user-centric clustering for downlink cloud radio access network. IEEE Access. 2014;2(6):1326–1339.
[10] Nguyen KG, Vu QD, Juntti M, and Tran LN. Energy efficiency maximization for C-RANs: discrete monotonic optimization, penalty, and ℓ_0-approximation methods. IEEE Trans Signal Process. 2018;66(17):4435–4449.
[11] Park SH, Simeone O, Sahin O, and Shitz SS. Fronthaul compression for cloud radio access networks: signal processing advances inspired by network information theory. IEEE Signal Process Mag. 2014;31(6):69–79.
[12] Nguyen K, Vu Q, Juntti M, and Tran L. Energy efficient precoding C-RAN downlink with compression at fronthaul. In: 2017 IEEE International Conference on Communications (ICC), Paris, 2017, pp. 1–6,
[13] Peng M, Wang C, Lau V, and Poor HV. Fronthaul-constrained cloud radio access networks: insights and challenges. IEEE Wireless Commun. 2015;22(2): 152–160.
[14] Peng M, Sun Y, Li X, and Wang C. Recent advances in cloud radio access networks: system architectures, key techniques, and open issues. IEEE Commun Surv Tutor. 2016 Third quarter;18(3):2282–2308.
[15] Schaible S and Shi J. Recent developments in fractional programming: single ratio and max-min case. Nonlinear Anal Convex Anal. 2004;493506:493–505.
[16] Zappone A and Jorswieck E. Energy efficiency in wireless networks via fractional programming theory. Found Trends Commun Inf Theory. 2015; 11(3–4):185–396.
[17] Zhang YJ, Qian L, and Huang J. Monotonic optimization in communication and networking systems. Found Trends Networking. 2013;7(1):1–75.
[18] Tuy H, Minoux M, and Hoai-Phuong NT. Discrete monotonic optimization with application to a discrete location problem. SIAM J Optim. 2006;17(1):78–97.
[19] Tuy H, Al-Khayyal F, and Thach PT. Monotonic optimization: branch and cut methods. In: Essays and Surveys in Global Optimization. Springer; 2005. pp. 39–78.
[20] Boyd S and Vandenberghe L. Convex optimization. Cambridge University Press; 2004.
[21] Marks BR and Wright GP. A general inner approximation algorithm for nonconvex mathematical programs. Oper Res. 1978;26(4):681–683.

[22] Beck A, Ben-Tal A, and Tetruashvili L. A sequential parametric convex approximation method with applications to nonconvex truss topology design problem. J Global Optim. 2010;47(1):29–51.

[23] Le Thi HA, Dinh TP, Le HM, and Vo XT. DC approximation approaches for sparse optimization. Eur J Oper Res. 2015;244(1):26–46.

[24] Dai B and Yu W. Energy efficiency of downlink transmission strategies for cloud radio access networks. IEEE J Sel Areas Commun. 2016;34(4): 1037–1050.

[25] Shi Y, Zhang J, and Letaief KB. Group sparse beamforming for green cloud-RAN. IEEE Trans Wireless Commun. 2014;13(5):2809–2823.

[26] Hajisami A, Tran TX, and Pompili D. Elastic-Net: Boosting Energy Efficiency and Resource Utilization in 5G C-RANs. In: 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). Orlando, FL, USA; 2017. pp. 466–470.

[27] Vu TT, Ngo DT, Dao MN, Durrani S, Nguyen DH, and Middleton RH. Energy efficiency maximization for downlink cloud radio access networks with data sharing and data compression. IEEE Trans Wireless Commun. 2018;17(8):4955–4970.

[28] Marsch P and Fettweis GP. Coordinated multi-point in mobile communications: from theory to practice. Cambridge University Press; 2011.

[29] Checko A, Christiansen HL, Yan Y, *et al.* Cloud RAN for mobile networks – a technology overview. IEEE Commun Surv Tutor. 2015 First quarter;17(1):405–426.

[30] Pfeiffer T. Next generation mobile fronthaul and midhaul architectures [Invited]. IEEE/OSA J Opt Commun Networking. 2015;7(11):B38–B45.

[31] Isheden C and Fettweis GP. Energy-Efficient Multi-Carrier Link Adaptation with Sum Rate-Dependent Circuit Power. In: 2010 IEEE GLOBECOM 2010; 2010. pp. 1–6.

[32] Auer G, Giannini V, Desset C, *et al.* How much energy is needed to run a wireless network? IEEE Wireless Commun. 2011;18(5):40–49.

[33] Dhaini AR, Ho PH, Shen G, and Shihada B. Energy efficiency in TDMA-based next-generation passive optical access networks. IEEE/ACM Trans Netw. 2014;22(3):850–863.

[34] Persson D, Eriksson T, and Larsson EG. Amplifier-aware multiple-input multiple-output power allocation. IEEE Commun Lett. 2013;17(6): 1112–1115.

[35] Mikami S, Takeuchi T, Kawaguchi H, Ohta C, and Yoshimoto M. An Efficiency Degradation Model of Power Amplifier and the Impact Against Transmission Power Control for Wireless Sensor Networks. In: 2007 IEEE Radio and Wireless Symposium; 2007. pp. 447–450.

[36] Candès EJ, Wakin MB, and Boyd SP. Enhancing sparsity by reweighted ℓ_1 minimization. J Fourier Anal Appl. 2008;14(5):877–905.

[37] Dattorro J. Convex optimization & Euclidean distance geometry. Lulu.com; 2010.

[38] ApS M. The MOSEK optimization toolbox for MATLAB manual. Version 9.0; 2019. Available from: http://docs.mosek.com/9.0/toolbox/index.html.
[39] Alizadeh F and Goldfarb D. Second-order cone programming. Math Program. 2001;95:3–51.
[40] Nguyen KG, Tervo O, Vu QD, Tran LN, and Juntti M. Energy-efficient transmission strategies for CoMP downlink—overview, extension, and numerical comparison. EURASIP J Wireless Commun Netw. 2018;2018(1):207.
[41] Mehanna O, Huang K, Gopalakrishnan B, Konar A, and Sidiropoulos ND. Feasible point pursuit and successive approximation of non-convex QCQPs. IEEE Signal Process Lett. 2015;22(7):804–808.
[42] Vu QD, Tran LN, Farrell R, and Hong EK. An efficiency maximization design for SWIPT. IEEE Signal Process Lett. 2015;22(12):2189–2193.
[43] Le Thi HA and Nguyen MC. DCA based algorithms for feature selection in multi-class support vector machine. Ann Oper Res. 2017;249(1–2):273–300.
[44] Ben-Tal A and Nemirovski A. Lectures on modern convex optimization. Philadelphia: MPS-SIAM Series on Optimization, SIAM; 2001.
[45] Nguyen D, Tran LN, Pirinen P, and Latva-aho M. Precoding for full duplex multiuser MIMO systems: spectral and energy efficiency maximization. IEEE Trans Signal Process. 2013;61(16):4038–4050.
[46] Nguyen KG, Vu QD, Juntti M, and Tran LN. Distributed solutions for energy efficiency fairness in multicell MISO downlink. IEEE Trans Wireless Commun. 2017;16(9):6232–6247.
[47] Westerberg E. 4G/5G RAN architecture: how a split can make the difference. Ericsson Technol Rev. 2016;93(6):1–15.
[48] Vu Q, Tran L, and Juntti M. Distributed Noncoherent Transmit Beamforming for Dense Small Cell Networks. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019. pp. 4599–4603.

Chapter 12

Energy-efficient full-duplex networks

José Mairton B. da Silva Jr.[1], *Christodoulos Skouroumounis*[2], *Ioannis Krikidis*[2], *Gábor Fodor*[3,4] *and Carlo Fischione*[1]

As the standardization specifications of the fifth generation (5G) of cellular networks mature [1], the deployment phase is starting up [2]. Hence, peaks of data rates in the order of tens of Gbit/s as well as more energy-efficient deployments in the order of tens of bit/J are expected [3,4]. Nevertheless, the quick development of new applications and services encourages the research community to look beyond 5G and explore new technological components [5]. Indeed, to meet the increasing demand for mobile broadband as well as internet of things type of services, the research and standardization communities are currently investigating novel physical and medium access layer technologies. The new technologies include further virtualization of networks, the use of the lower terahertz bands, even higher cell densification, and full-duplex (FD) communications [5].

FD has been proposed as one of the enabling technologies to increase the spectral efficiency of conventional wireless transmission modes [6]. The FD technology overcomes our prior understanding that it is not possible for radios to transmit and receive simultaneously on the same time-frequency resource. Due to this, we can also refer to FD communications as in-band FD. In-band FD transceivers have the potential of improving the spectral efficiency of networks operating with half-duplex (HD) transceivers by a factor close to two [7]. In addition to the spectral efficiency gains, FD can provide gains in the medium access control layer. For Wi-Fi networks, problems such as the hidden/exposed nodes and collision detection can be mitigated and the energy consumption can be reduced [6]. For cellular networks, together with the gains in spectral efficiency, the end-to-end delay may be reduced, which can lead to an increase in the energy efficiency and a reduction of the latency [6].

[1]Division of Network and Systems Engineering, Department of Computer Science, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
[2]Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus
[3]Ericsson Research, Stockholm, Sweden
[4]Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden

Until recently, in-band FD was not considered as a feasible technology for wireless networks due to the inherent interference created from the transmitter to its own receiver, the so-called self-interference (SI). However, recent advancements in antenna and analog/digital SI interference cancellation techniques demonstrate FD transmissions as a viable alternative to traditional HD transmissions [8–11]. Given the recent architectural progression of 5G toward higher densification, higher number of antennas and utilizing the millimeter wave (mmWave) band, the integration of FD communications into such scenarios is appealing. Although the SI remains a challenge, in-band FD communications are more suited for short-range communication due to lower transmission power at the base station (BS) and shorter distance between users and the BS. The use of multiple antennas and the transmission in the mmWave band are allies that help to mitigate the SI in the spatial domain and provide even more gains for spectral and energy efficiency. Therefore, FD communications appear as an important technology to improve the spectral and energy efficiency of current communication systems and help to meet the goals of 5G and beyond. To this end, it is important to understand the challenges and solutions of using FD communications to achieve the spectral and energy efficiency gains. In this chapter, the solutions for the challenges are divided into resource allocation, protocol design, hardware design, and energy harvesting.

This chapter starts with an overview of FD communications, including its challenges and solutions. Next, a comprehensive literature review of energy efficiency in FD communications is presented along with the key solutions to improve energy efficiency. Finally, we evaluate the key aspects of energy efficiency in FD communications for two scenarios: single-cell with multiple users in a pico-cell scenario and a system-level evaluation with macro- and small-cells with multiple users.

12.1 Introduction

In-band FD in wireless networks is not recent, the concept of transmission and reception in the same frequency channel has been used since 1940 in radar systems [12]. SI was already a key challenge at that time, and the first circuits to mitigate the SI were proposed and provided low levels of cancellation. Such mild SI cancellation levels limited the transmit power and provided a reduced range of detectable targets. In the last decade, wireless broadband systems, such as Wi-Fi and cellular, started to experiment with in-band FD communications. Depending on the transmitter power, the SI needs to be cancelled by more than 100 dB to reduce its value to the noise floor [8], which is a significant level to be cancelled.

To cancel SI, we need to rely on techniques that can be tackled from the propagation, analog, and digital domains [12]. Usually, the first cancellation happens in the propagation domain, which includes the physical separation of the transmitter and receiver antennas, the use of duplexers or circulators to share the transmitter and receiver antenna, and in multiple antenna scenarios the use of the spatial separation [13,14]. Next, SI cancellation in the analog and/or digital domain suppresses the residual SI, which may include the use of additional circuits.

The best SI cancellation so far appears to have been achieved in [8], in which 110 dB of SI cancellation in Wi-Fi networks with bandwidth of 20 MHz were achieved. According to the authors, such performance can also be achieved in current fourth-generation (4G) systems regardless of the frequency band. More recently, SI cancellation has achieved significant levels also for mobile devices [10], FD multiple input–multiple output (MIMO) relays [15], device-to-device nodes using FD [16], in mmWave bands [11,17], wideband communications [18], and real-time cancellation [19]. As an example of practical success, researchers from Stanford University have founded a start-up company, Kumu Networks, to develop practical FD radios that are currently being deployed in small cell BSs [20].

With respect to the transmission configurations on different radio access technologies, such as Wi-Fi or cellular networks, we can envision in-band FD communications in three application areas: bidirectional FD, three-node FD, and relaying FD [21]. Figure 12.1 shows an HD system, using a three-node configuration with a BS, and two nodes, and the aforementioned FD configurations. In bidirectional FD, two FD-capable nodes (either a user, access point, or BS) transmit and receive on the same time–frequency resource that creates SI for both the nodes. For Wi-Fi or IoT applications, this configuration is attractive due to the distributed aspects of such networks. In contrast, three-node FD involves three nodes, but it requires FD capability only at the BS (or access point). The FD-capable node transmits to its receiver node while receiving from another transmitter node on the same frequency channel, in which SI is present only at the FD-capable node. For cellular networks, this configuration is attractive because instead of requiring all users to be FD, only the BS can be assumed FD-capable, while users remain HD (Figure 12.1). In relaying FD, one node transmits to an FD-capable relay and then retransmits the signal to the second node, where all transmissions occur in the same time–frequency resource, and Node_1 cannot transmit directly to Node_2 due to propagation conditions. Since the relay is the only FD-capable node, SI is present only at the relay. This configuration is appealing to general radio access technologies because it can represent a master–worker architecture in IoT, as well as a relay or backhaul architecture in cellular networks.

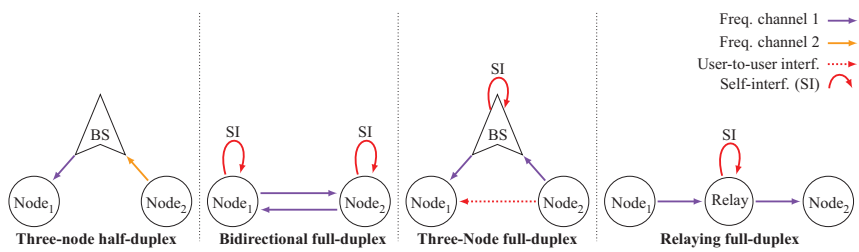


Figure 12.1 *In-band FD schemes are divided into three configurations, bidirectional FD, three-node FD, relaying FD, and the additional HD mode in a three-node configuration. Some configurations are more suitable to cellular networks with the BS as access points, whereas others are also suited for general radio access technologies*

These new configurations extend the design options and allow for many benefits such as higher spectral usage of the already available frequency resource, lower power consumption in the physical and medium access layer, higher physical-layer security through friendly jamming, higher detection of nodes, and higher coverage area. Overall, the common challenges that all configurations face are the SI, which must be mitigated using different cancellation methods, and the user-to-user interference. For instance, the bidirectional FD configuration suffers from SI at both nodes, while the three-node FD has to deal with SI only at the BS and has to mitigate the user-to-user interference between uplink (UL) and downlink (DL) users.

Therefore, the FD technology is quickly approaching the phase of commercial deployments in low-power wireless networks [20], which highlights the need of addressing the energy efficiency aspects in FD communications. When the analog domain cancellation of the SI signal is used, the inclusion of additional circuits may incur higher energy consumption in the transceiver architecture. It is important to take into account energy efficiency aspects in FD to guarantee that, together with spectral efficiency gains, the energy efficiency of the system increases despite the potentially higher circuit power consumption. For an energy-efficient design of FD communications, the solutions for the challenges consist of techniques and technologies divided into four categories: resource allocation, protocol design, hardware design, and energy harvesting.

Resource allocation techniques consider energy efficiency aspects in the optimization of the physical layer, for scenarios considering a single- or a multicell configuration. Protocol designs, either for cellular or Wi-Fi networks, use the benefits of the FD capability in the medium access control layer to reduce the power spent in retransmissions. Hardware designs include new analog/digital devices with lower power consumption and the smaller sizes of current SI cancellation circuits. Energy harvesting is a technology that uses different harvesting techniques to supply wireless devices, including common energy sources such as solar, wind, and radio frequency signals. For FD communications, the SI created by the FD node becomes an additional energy source that can be harvested, which suggests that the energy harvesting technology can be used to further improve the energy efficiency of FD communications. The next section presents a comprehensive literature review and the current advancements for the four solution categories.

Information-theoretic tools helped in clarifying the potential gains of the FD technology compared to the HD operation by exploring the achievable rate regions. Despite the maturity reached in the design of a single-link and the remarkable performance gains achieved in small-scale networks, such as single-cell scenarios, a deep understanding of the role played by FD technology in more complex networks deployments is, however, still elusive. Specifically, the concept of FD radio in large-scale networks, such as multicell scenarios, triggers a nontrivial trade-off between spatial reuse and aggregate interference. With the simultaneous bidirectional data exchange within node pairs, more links per unit area can be active and can potentially lead to enhanced performance. However, the additional amount of interference caused by a more aggressive access to the medium decreases the probability of successfully retrieving information. To understand the performance of FD communications on

both scenarios, we address separately single- and multicell scenarios in our literature review and analysis on Sections 12.3 and 12.4. For each section, we propose novel solutions to address energy efficiency using tools ranging from optimization theory to stochastic geometry.

12.2 Literature review

As it has been emphasized in many recent research reports, for FD communications to become part of mainstream communication technologies, energy efficiency is an important aspect. Figure 12.2 presents key references in the four categories of techniques and technologies that enable energy-efficient FD communications.

12.2.1 Resource allocation

Resource allocation techniques include power control, user-frequency assignment, user scheduling, and, in the case of multiple antenna systems, transmitter and receiver

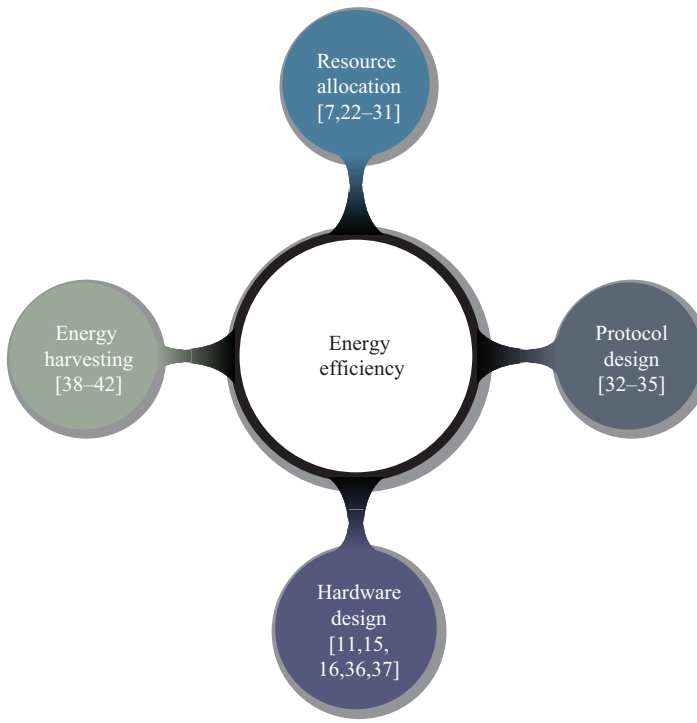


Figure 12.2 For an energy-efficient design of FD communications, the solutions consist of techniques and technologies divided into four categories: resource allocation, protocol design, hardware design, and energy harvesting

beamforming. The techniques aiming at energy efficiency may appear in terms of explicit energy efficiency maximization, total power minimization under quality of service (QoS) constraints, such as minimum rate, or variants of both approaches aforementioned with or without QoS constraints [22]. Notice that FD communications experience different types of interference, such as SI and user-to-user interference, and the aforementioned resource allocation techniques are used to mitigate the interference while promoting an energy-efficient FD network. For example, transmitting with high power at the BS will increase the SI and impact negatively the rate of the UL user. Then, the BS and UL should have a proper power control to not allow the situation mentioned previously. Since single- and multicell scenarios experience different types of interference, such as the BS-to-BS interference, resource allocation techniques can be further categorized into single- and multicell techniques.

In [23], the authors define two different optimization goals in an FD single-cell MIMO environment: the first one is to maximize spectral efficiency, and the second is to maximize energy efficiency. The energy efficiency objective function is defined as the sum rate of both UL and DL users divided by the total power consumption of both UL and DL. For total power consumption, the authors take into account both the power amplifier efficiency and the circuit power. The numerical results indicate that FD systems with an energy-efficient design are slightly outperformed by HD systems in terms of energy efficiency. The reason is that FD systems increase the spectral efficiency but consume more energy than HD systems, which is due to continuous transmission and reception on FD systems that impact negatively the energy efficiency. In addition, the authors show that both designs, an energy and a spectral efficient, achieve approximately the same energy efficiency when the transmitted power at the BS is small, whereas the gap between the two designs widens as the transmit power increases.

The authors in [24] present a survey of energy-efficient techniques for mmWave FD relaying systems. An important insight in this work is that energy efficiency in FD relaying systems is largely affected by the power consumption of the power amplifier, circuit power, and the SI circuit. Due to FD communications, both transmit and receiver chains are active at the same time, which leads to a naturally higher power consumption than in HD systems. In addition, the power consumption may increase if both analog and digital SI cancellation are employed. Analyzing the energy efficiency of passive suppression, analog and digital cancellation, the authors show that FD with passive suppression and analog cancellation outperforms FD employing both suppression and digital cancellation.

In [25], the authors analyze the trade-off between energy and spectral efficiency in a pico-cell. Assuming either a constant or linearly increasing residual SI term, the authors propose power control and resource allocation solutions to maximize the energy efficiency defined as in [23] with QoS constraints. The numerical results indicate that the energy efficiency of FD systems outperforms HD systems, and that the gap increases with the number of user pairs. In addition, the authors also show that the point that reaches maximum energy efficiency in FD systems provides at least the same spectral efficiency of HD systems. Hence, given a spectral efficiency target, FD systems consume much less power than what HD systems do.

The authors in [26] propose a green networking solution for a bidirectional FD network with MIMO nodes and relying on precoding for interference mitigation between nodes. The objective is to the minimize the total power consumption in the network subject to QoS constraints related to minimum rate for each node. Using game theory, the authors prove the existence and uniqueness of the Nash equilibrium and design a medium access control protocol to implement the proposed solution in a distributed fashion. The numerical results indicate that the proposed solution significantly outperforms a carrier sense multiple access-based design using FD communications and HD schemes in terms of both throughput and energy/power efficiency.

In [27], the objective is to design an energy-efficient hybrid beamforming solution for FD mmWave relaying systems. The objective function is to maximize the energy efficiency while taking into account QoS constraints, the analog/digital precoders and combiners at the relay, the power amplifier efficiency, and the circuit power consumption for splitters, phase shifters, and SI cancellation. The numerical results indicate that the number of antennas is also an important factor, and that a very large number of antennas may, in fact, decrease energy efficiency. The reason is that there exist an optimal number of antennas that maximize the energy efficiency. Hence, when deploying a higher number of antennas than the optimal, increasing the spectral efficiency incurs a significant increase in the power consumption. Moreover, a small number of radio-frequency chains provide higher energy efficiency than a higher number of radio-frequency chains. When comparing the spectral and energy efficiency for different number of antennas, both increase until reaching the optimal point mentioned previously. Therefore, an important message of this work is that the optimal energy efficiency can be reached by employing a certain number of antennas.

Different from the works mentioned earlier, the authors in [28] use FD jamming to improve the system's secrecy energy efficiency. The scenario takes into account the additional power used for friendly jamming and SI cancellation at the FD nodes in a MIMO scenario with and without perfect channel state information. The goal is to maximize the secrecy energy efficiency, and the authors propose an iterative solution to solve the problem. The numerical results indicate that the secrecy energy efficiency gains are small for many scenarios. However, the gains increase when the SI cancellation is high enough, or the distance between the FD node and the eavesdropper is small, or when the transmitter and receiver use bidirectional FD.

Reference [7] lists energy efficiency challenges for FD communications in a multicell indoor environment. When the scheduler does not aim to achieve high energy efficiency, energy efficiency in FD systems can be lower than in HD systems due to the strong impact of the SI and the user-to-user interference. Recognizing these inherent problems of FD systems, the authors propose two solutions. The first one considers beamforming and sectorization to cancel or mitigate the additional interference due to FD systems. The second one considers a scheduler with a utility function that takes into account the power allocation of users along with the total system throughput. In the first solution, FD systems outperform HD systems in terms of spectral and energy efficiency for high and low SI cancellation values. For the second solution, FD systems still outperform HD systems in terms of spectral and energy efficiency, but there are losses in the spectral efficiency when compared to

the previous scheduling algorithm. Nevertheless, there are significant gains in terms of energy efficiency.

In [29], the authors consider two fundamental FD architectures, two-node and three-node, in the context of cellular networks where the terminals employ directional antennas. Using a stochastic geometry model, the authors investigate how directional antennas can control and mitigate the co-channel interference and also provide a model that characterizes the way directional antennas manage the SI in order to passively suppress it. Finally, a composite architecture network is considered, where both the architectures (two-node and three-node architecture) are employed in the network. Numerical results illustrate the effect of self- and multiuser interference on the FD performance, highlighting the significance of directional antennas in reducing the overall interference and therefore increase the network throughput and energy efficiency.

In [30], the authors investigate the local delay for FD heterogeneous networks, in which all BSs and user devices are FD capable. The objective is to model and analyze the expected time required for a successful transmission from an arbitrary node to its respective receiver in terms of system key parameters. Particularly, the authors consider a distance-based mode selection scheme for the optimal mode selection for the network nodes, and a silent-mode capability for all BSs to reduce the energy consumption. The silent-mode capability allows BSs to be inactive, resulting in the reduced interference observed by the receiver of interest and correspondingly, to the increased chance of successful transmissions. In this chapter, by utilizing tools from the stochastic geometry, the authors investigate the energy efficiency in single-layer cellular networks and the energy efficiency in the FD-capable heterogeneous networks by considering local delay in the analysis. Numerical results shown that, rising the silent probability in the high signal-to-interference ratio (SIR) regime can improve both the local delay and the energy efficiency. Finally, the authors shown that the considered hybrid heterogeneous deployments can increase the optimal SIR range for the energy efficiency.

The authors in [31] study the performance of FD radio in small cellular systems, where all BSs are FD capable, while user equipment's operate in the HD mode. Using stochastic geometry, the authors propose a mixed system model that allows the derivation of the outage and area spectral efficiency of such a system. The objective is to find the proportion of FD BSs such that some given constraints in terms of area spectral efficiency or, alternatively, of energy efficiency, can be met. Among the main findings, the authors have shown that the fraction of FD cells can be used as a design parameter to target different area spectral efficiency versus coverage trade-offs for the network operator. In particular, by increasing the amount of FD cells in the mixed system, the overall throughput increases at the cost of a drop in terms coverage and vice-versa.

Reference [32] develops an analytical framework that comprises the codesign of hybrid HD/FD heterogeneous cellular networks with mmWave communications, where the users are classified as either cell-center user (CCU) or cell-edge user (CEU) based on their location within a cell. In addition, the authors consider the employment of a distance-proportional fractional power control for the users, aiming the mitigation of the multicell interference caused by the FD operation. Based on stochastic geometry

tools, the coverage and the sum-rate performance are derived, with and without the employment of an ideal successive interference cancellation technique for both the DL and UL transmissions. Numerical results indicate that the achieved performance highly depends on the user location within the cell, and that heterogeneous mmWave cellular networks can significantly improve the sum-rate performance of FD systems and consequently the obtained energy efficiency, as compared with conventional sub-6 GHz cellular networks.

Many energy efficiency aspects, as part of the resource allocation procedure, for FD communications on single- and multicell scenarios have been addressed in the literature. However, the abovementioned literature review clearly suggests that there is a need to address energy efficiency aspects for FD communications when combined with other technologies that are a natural part of 5G. For FD massive MIMO systems, it is important to analyze the effect of a low number of quantization bits due to their impact on the energy consumption and SI cancellation. For FD mmWave communications, the phase shifters may also use low number of quantization bits to further reduce complexity and energy consumption. In both cases, the beamforming solutions need to be designed taking into account the energy efficiency of the system.

12.2.2 Protocol design

The protocol design techniques include enhancements in the medium access control layer protocols aiming at a more energy-efficient FD network. Such techniques span different applications and FD configurations, including Wi-Fi, sensor networks, and cellular networks.

In [33], the authors propose an energy model for FD that is compatible with current medium access control layer protocols used in wireless sensor networks. Due to the nature of simultaneous transmission and reception of FD, the authors propose a solution for the collision detection and the hidden node problems that takes into account the energy consumption and delay. To address the energy consumption, the authors analyze the different transmission states of an FD transceiver, such as shutdown, idle, transmit, receive, and FD. Given these states, the authors derive average power consumption and different delays along the FD transmission states. The numerical results indicate that FD transmissions allow DL and UL traffic to coexist without losses to the UL users, which does not happen in the HD mode because of the collisions and prioritization of DL traffic. A key result of this work is that when more users are connected to the network, FD mode has much lower delays and higher energy efficiency than HD mode, which happens because of a better collision detection in FD transmissions.

Reference [34] proposes a protocol focused on reducing the transmission power of data and acknowledgment packets in order to improve the energy efficiency of the network. Assuming bidirectional and relaying networks, the proposed protocol is able to mitigate the hidden node problem while maximizing the energy efficiency of the network. The numerical results indicate that the proposed protocol outperforms HD and other FD protocols in terms of energy efficiency, and that the energy consumption slightly increases with the bandwidth. In addition, the results indicate that the SI

must be properly addressed to harness the gains of an energy-efficient FD network. An interesting message of this work is the impact of the bandwidth in the energy efficiency of the proposed protocol, which to the best of our knowledge has not been carefully addressed in the literature.

In [35], the authors analyze on-and-off schemes in FD networks through the medium access control layer. The proposed technique reduces the power consumption by using specific beacon cycles and setting the medium on and off according to such cycles. The authors assume three communication scenarios, including bidirectional and three-node FD, and the HD, and consider scheduling schemes to reduce the package exchange overhead and to allow users to send multiple packets in a single beacon cycle. The numerical and experimental results indicate that the proposed schemes outperform HD and other FD medium access solutions in terms of energy efficiency, throughput, and low power consumption for different arrival rates, number of users, SI circuit power consumption, and traffic asymmetry between UL and DL.

The model developed in [36] takes into account spectrum-sharing schemes for FD cognitive networks, in which the secondary users are willing to use the licensed spectrum of primary users. The authors design a frame model that allows the secondary users to be in HD mode, referred to as learn-before-transmit, and in FD mode, referred to as learn-and-transmit, depending on the traffic the primary users experience. Using the proposed frame model, the authors derive analytically the energy efficiency and propose an adaptive spectrum sensing scheme to reduce the energy consumption. The numerical results show that the proposed adaptive scheme provides more energy efficiency for low and high traffic arrival scenarios, and that it outperforms the conventional HD mode and a modified HD mode using the proposed adaptive scheme.

Overall, we notice a gap in the literature for protocols supporting FD in 5G and 4G networks in terms of addressing energy efficiency. Using the insights from the FD literature on protocol design for Wi-Fi radio access technologies, energy-efficient protocols supporting FD communications can be studied in various standardization bodies. For example, 3rd Generation Partnership Project (3GPP) standardized the 5G radio interface as 5G New Radio, in which there are different capabilities that handle a large number of antennas and a sleep state of the radio network. Hence, it opens up for possibilities for enabling an energy-efficient use of FD communications.

12.2.3 Hardware design

Due to recent advancements in SI cancellation techniques, the circuits and techniques to mitigate the SI signal may consume less power for the cancellation and reduce the circuit sizes allowing FD compact devices. With a compact device, the circuit power consumption is expected to drop due to lower power dissipation, more efficient circuits, and also allow the use of FD communications in handheld devices, such as mobile phones, to improve the spectral efficiency. Therefore, reducing the scale of the SI cancellation circuits and their energy consumption can impact directly in the energy efficiency.

The authors in [37] propose a novel SI cancellation technique using a lossless coupling network, which interconnects the transceiver's chains to the antenna elements. The technique provides an SI cancellation method at the analog domain that can be applied into FD MIMO transceiver arrays. Instead of focusing on a small-scale circuit to cancel the SI interference, the energy benefits from this technique are the reduction of the power dissipated for SI cancellation purposes. The authors conducted experiments to validate the SI cancellation for a MIMO configuration with two antennas, and the SI cancellation levels exceed 70 dB for all the analyzed scenarios.

Reference [38] focuses on compact analog SI cancellation solutions that can be integrated in BSs, notebooks, sensor nodes, and mobile phones. The authors propose two designs to achieve the desired small form factor. The first design aims at small-cell BSs and notebooks as application areas, whereas the second design targets more mobile phones and sensor nodes. The first design combines propagation and analog cancellations by using a dual-port polarized antenna and a self-tunable circuit, which offers up to 75 dB of SI cancellation. The preliminary implementation indicates an antenna size of 90 mm $\times$ 90 mm, but it can be scaled down to 30 mm $\times$ 30 mm. The second design uses a tunable electrical balance isolator in combination with a single-port antenna, which offers up to 50 dB. The implementation of this design based on low-cost plain complementary metal oxide semiconductor (CMOS) technology (excluding the antenna) would measure less than 1 mm^2.

A new architecture for a compact multi-antenna FD relay for different bandwidth and transmit powers is proposed in [15]. Given the desired small size, it is necessary to use other passive isolation techniques that are not only based on physical antenna separation. Hence, the authors use wave traps to provide up to 70 dB of passive SI cancellation while maintaining the compact size. In addition, the authors analyze two scenarios to combine the propagation domain cancellation with either analog or/and digital cancellation. The SI cancellation level increases to approximately 110 dB when all SI cancellations on all domains are assumed, while more than 100 dB of SI cancellation is achieved using only propagation and digital domain cancellation. When in a scenario with bandwidth of 80 MHz with either 30 or 24 dBm of transmit power level, the SI cancellation levels are significant and also reach the 100 dB mark.

The work reported in [16] aims at providing a real-time compact multi-antenna FD along with device-to-device communications that take into account not only the SI cancellation level but also the compactness and power efficiency. To provide FD operation, the authors consider dual polarization-based radio-frequency front-end with high cross polarization for the passive SI cancellation. The benefit of using this dual polarization is the high isolation without needing to consume additional power from analog components such as the circulator. In order to further improve the SI cancellation, the authors propose a digital cancellation that is robust against frequency selective fading and compatible with current 4G systems. The proposed SI cancellation scheme achieves more than 90 dB of cancellation for a wide range of transmit power levels.

Finally, reference [11] summarizes recent efforts from the literature to make compact FD radios and highlight some of their results in creating compact FD radios. To

create small form factor FD radios, one of the biggest challenges is the implementation of low-loss shared-antenna interface with high transmitter–receiver isolation. Using shared-antenna interfaces, the compact form factor can be achieved and could provide an easier application to multi-antenna systems while easing the system design through channel reciprocity. One of their proposed architectures is a wideband, reconfigurable, and orthogonally polarized antenna interface, which was employed in 4.6 GHz with a bandwidth of 300 MHz and achieved 50 dB of SI cancellation. Moreover, the authors extended the prototype to 60 GHz, which has a size of 1.3 mm $\times$ 3.4 mm and was implemented on a CMOS chip. Along with digital cancellation, the proposed design achieves approximately 80 dB over the 60 GHz and with a bandwidth of 1 GHz.

Due to these recent advances reported earlier, the hardware design has improved since the first works in FD, and the implementation of smaller devices with FD transceivers has gained momentum. Moreover, the use of polarized antennas and implementation on CMOS is an important direction that needs further research, not only in sub-6 GHz but also in the mmWave spectrum. Nevertheless, the literature on SI cancellation techniques that reduce the power dissipation, such as the ones with lossless networks, is still small and requires further research.

12.2.4 Energy harvesting

In general, the energy harvesting technology is attractive because it provides a cost-efficient and almost perpetual energy supply to wireless networks. Specifically to FD communications, energy harvesting is a natural ally for providing energy efficiency. Both technologies are suited for short communication range, and in FD communications the SI signal can also be a source of energy to be harvested. We can divide the techniques used for energy harvesting along two lines: simultaneous wireless information and power transfer (SWIPT) and wireless-powered communications (WPC). In SWIPT, the same transmitted signal is used simultaneously for wireless information and energy, whereas in WPC the terminals communicate using the energy harvested from wireless power transmissions [39].

In [39], the authors analyze WPC in FD communication systems in which the access point broadcasts energy and the users transfer information to the access point. The authors propose a novel protocol that enables energy transfer from the FD access point in the DL and information transfer in the UL using time division. The proposed protocol aims at weighted sum-rate maximization of the users while optimizing the allocated time for the energy transfer in the DL and the information transfer in the UL, and the average and peak transmit power at the access point. The numerical results indicate that the proposed FD protocol outperforms HD in terms of average sum-rate when the SI cancellation level and number of users are high. In scenarios with high SI cancellation level, the performance gain in terms of average sum-rate between the FD and the HD mode increases with the average power.

Reference [40] analyzes a model that is similar to that used in [39] but takes into account a causality constraint for the users. The causality constraint implies that the users cannot use the harvested energy after its transmission time, i.e., users can only use all the energy harvested thus far before their transmissions. This constraint creates a coupling between the energy harvested and the transmission time, which

makes the problem more complicated than [39]. The authors have two different goals: sum rate maximization with total charge and transmission time minimization. For each goal, the authors provide an optimal and a suboptimal solution that are numerically evaluated in the results. The numerical results indicate that the optimal solutions outperform the baseline solutions with equal time allocation, and that the system is able to maximize the throughput or minimize the total time for different number of users and transmission powers.

Reference [41] considers an FD relaying system with multiple antennas in which WPC is employed at the relay. The multi-antenna relay receives energy from the source and transmits information to the destination, and it aims at instantaneous and delay-constrained throughput maximization. The authors propose optimal and suboptimal solutions for the linear transmitter and receiver beamformers, and the time-sharing variable that allows the time switching between energy and information transfer. The numerical results indicate that the time used for energy harvesting decreases as either the number of relay's receive antennas or the source's transmit antennas increase, which indicates that in such cases the relay can harvest the same amount of energy in a shorter time. Moreover, the results show that the beamforming increases the SI cancellation capability at the FD relay and that the suboptimal solutions based on linear beamforming provide close performance to the optimum in many situations.

The interesting question whether SI is a friend or a foe in FD communications is examined in [42]. The main assumption is that the SI signal could be used as an energy harvesting source at the BS by switching off the SI cancellation circuit for a period of time to allow the energy to be harvested, and then maintaining it on when the concurrently UL and DL transmissions occur. The goal is to maximize the energy efficiency, and as such the authors propose a solution with guaranteed convergence. The numerical results indicate that the proposed solution outperforms in terms of energy efficiency the solution without the energy harvesting through the SI signal.

A secure WPC scheme in an FD scenario with a power station, an information transmitter, an eavesdropper, and multiple intended users is proposed in [43]. The power station broadcasts a multicast service for multiple users and a confidential unicast service to one user while protecting this information from the eavesdropper. The objective of the authors is to maximize the secrecy-multicast rate while taking into account the energy transmission, the multicast and the unicast service, and the artificial noises to degrade the eavesdropper performance. To meet the objective, the authors derive an approximate solution for scenarios with perfect and imperfect channel state information. The numerical results indicate that the proposed solution provides a better secrecy-multicast trade-off performance than HD systems for both perfect and imperfect channel state information. Moreover, the authors show that a larger number of antennas at the information transfer provide a better secrecy-multicast trade-off. Hence, FD communications together with the energy harvesting technology, through WPC techniques, also provide gains in terms of security when compared to HD systems.

Techniques using the energy harvesting technology have been analyzed thoroughly, specially WPC. Based on our literature review, we believe that there is a strong need for additional techniques that use the technology as a source of power in different scenarios, including those employing a large number of antennas. With

such techniques, it may be possible to benefit from the SI signal power to improve the energy efficiency of the system.

12.3 Single-cell analysis

Coordination mechanisms are required in order to mitigate the negative effects of interference, both SI and user-to-user interference, on the spectral and energy efficiency of the system [7]. The two most important mechanisms are user-pair-frequency assignment and power control [7,21,44]; together, these determine which users transmit simultaneously on the frequency channels and with which powers the users and the BS will transmit. Therefore, it is crucial to understand the trade-offs between UL and DL performance of FD small-cell systems in the design of spectral and energy-efficient medium access control protocols and also coordination mechanisms that help to realize the FD potential. Hence, in this subsection, we focus on the problem of joint power control and user-frequency channel assignment, assuming three-node FD transmissions and a frequency selective wireless environment. Specifically, we investigate the fundamental problem of sum spectral efficiency maximization while minimizing the sum power consumption in a single-cell system. Using a joint approach of spectral efficiency maximization and sum power minimization, the aim is to achieve both a spectral and energy-efficient solution to three-node FD in a single-cell system.

12.3.1 System model

We consider a single-cell and single-antenna cellular system in which the BS is FD capable, whereas the users served by the BS are only HD capable. We assume that within a large number of served users, a smaller subset is selected by an appropriate scheduler for data transmission in the UL and reception in the DL. The numbers of users in the UL and DL are denoted by I and J, respectively, which are constrained by the total number of frequency channels in the system F, i.e., $I \leq F$ and $J \leq F$. We assume that the UL and DL users have already been preselected by a scheduler for transmission, which explains the inequalities $I \leq F$ and $J \leq F$.

We assume a frequency selective environment, such that the composite channel gains depend on the frequency and consist of small- and large-scale fading. Let G_{ibf} denote the channel gain between transmitter user i and the BS on frequency channel f, G_{bjf} denote the channel gain between the BS and the receiving user j on frequency channel f, and G_{ijf} denote the interfering channel gain between the UL transmitter user i and the DL receiver user j on frequency channel f. To take into account the residual SI power that leaks to the receiver, we define β as the SI cancellation coefficient, such that the SI power at the receiver of the BS is βP_j^d when the transmit power is P_j^d.

The vector of transmit powers in the UL for all the users is denoted by $\mathbf{p^u} = [P_1^u \cdots P_I^u]^T$, whereas the vector of DL transmit powers by the BS is denoted by $\mathbf{p^d} = [P_1^d \cdots P_J^d]^T$. The user-to-user interference depends heavily on the geometry of the co-scheduled UL and DL users, i.e., the *pairing* of UL and DL users, which also depends on the frequency channels assigned to UL and DL users. Therefore, user pairing and frequency channel allocation are key functions of the system. Accordingly,

we define the assignment matrix, $\mathbf{X} \in \{0,1\}^{I\times J\times F}$, such that $x_{ijf} = 1$ if user i is paired with user j on frequency f, and 0 otherwise.

The signal-to-interference-plus-noise ratio (SINR) at the BS of transmitting user i and the SINR at the receiving user j of the BS, both assigned to frequency channel f, are given by

$$\gamma_{if}^{u} = \frac{P_i^u G_{ibf}}{\sigma^2 + \sum_{j=1}^{J} x_{ijf} P_j^d \beta}, \quad \gamma_{jf}^{d} = \frac{P_j^d G_{bjf}}{\sigma^2 + \sum_{i=1}^{I} x_{ijf} P_i^u G_{ijf}}, \tag{12.1}$$

respectively, where x_{ijf} in the denominator of γ_{if}^{u} accounts for the SI at the BS, whereas x_{ijf} in the denominator of γ_{jf}^{d} accounts for the user-to-user interference caused by user i to user j on frequency f, and σ^2 is the noise power.

The achievable spectral efficiency by each user is given by the Shannon equation (in bits/s/Hz) for the UL and DL as

$$C_i^u = \sum\nolimits_{f=1}^{F} \log_2(1+\gamma_{if}^{u}), \quad C_j^d = \sum\nolimits_{f=1}^{F} \log_2(1+\gamma_{jf}^{d}), \tag{12.2}$$

where the sums are taken along the frequency dimension because at most one SINR, out of F, is nonzero for every UL/DL user. Thus, the energy efficiency of the system can be defined as

$$EE = \frac{\sum_{i=1}^{I} C_i^u + \sum_{j=1}^{J} C_j^d}{\sum_{i=1}^{I} P_i^u + \sum_{j=1}^{J} P_j^d + P_c + P_s}, \tag{12.3}$$

where P_c accounts for the circuit power consumption at the transmitter and receiver chains, and P_s accounts for the total power consumption at the SI cancellation circuit. The term P_s is general enough to consider the power consumption regardless of the SI cancellation assumed on the propagation, analog, or digital domains.

Our goal is to devise the pairing of UL and DL users, assigning each pair to frequency channels and allocating the minimum sum transmit power such that the overall spectral efficiency is maximized. Specifically, we formulate the problem as

$$\begin{aligned}
\underset{\mathbf{X},\mathbf{p}^u,\mathbf{p}^d}{\text{maximize}} \quad & \sum_{i=1}^{I} C_i^u + \sum_{j=1}^{J} C_j^d && \text{(12.4a)}\\
\text{subject to} \quad & P_i^u \le P_{\max}^u, \ \forall i, && \text{(12.4b)}\\
& P_j^d \le P_{\max}^d, \ \forall j, && \text{(12.4c)}\\
& \sum_{j=1}^{J}\sum_{f=1}^{F} x_{ijf} \le 1, \ \forall i, && \text{(12.4d)}\\
& \sum_{i=1}^{I}\sum_{f=1}^{F} x_{ijf} \le 1, \ \forall j, && \text{(12.4e)}\\
& \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ijf} \le 1, \ \forall f, && \text{(12.4f)}\\
& x_{ijf} \in \{0,1\}, \ \forall i,j,f. && \text{(12.4g)}
\end{aligned}$$

The main optimization variables are $\mathbf{p}^u$, $\mathbf{p}^d$, and $\mathbf{X}$. Constraints (12.4b) and (12.4c) limit the transmit powers per-user and per-channel DL power constraint, whereas constraints (12.4d)–(12.4f) assure that at most one user in the DL can share the frequency resource with a user in the UL and vice versa.

We propose a close-to-optimal solution of problem (12.4), named G-FLIP, by splitting it into two parts corresponding to power control and user-frequency channel assignment problems. Due to lack of space, we discuss herein only the key steps of our solution approach that combines the Fast-Lipschitz (FL) optimization to solve the power control problem with a greedy assignment solution. For more details about the power control and greedy assignment solutions, we refer the interested reader to [44].

For the power control, we assume that the assignment has already been done and rewrite problem (12.4) in an equivalent form in terms of the transmit powers and the adaptive spectral efficiency (or SINR) targets of a user pair. We notice that the transmit powers are intertwined with the spectral efficiency targets through a sequence of feasibility checking subproblems. The subproblems provide an inequality solely between the spectral efficiency targets and the channel coefficients of a user pair (see [44, Lemma 1]). The remaining problem does not have the transmit powers as variables, and it aims to maximize the sum spectral efficiency target of a user pair while guaranteeing nonzero spectral efficiency for both users. With the aforementioned inequality, the remaining problem is not concave, and we resort to the FL optimization framework to obtain a close-to-optimal and fast solution. The main advantages of the FL optimization framework [45] are that the objective function or constraints are not required to be convex, and the optimal solution is found in a distributed manner using fixed-point iterations that are known for fast convergence. We rewrite the spectral efficiency target maximization problem in an equivalent FL optimization form and prove that the optimal solution can be obtained through the FL optimization framework (see [44, Lemma 2]). Using the spectral efficiency target obtained from the FL optimization, we use the aforementioned sequence of feasibility checking subproblems to obtain in a distributed manner the transmit powers with minimum sum. Notice that the problem formulation (12.4) has no minimum QoS constraint, i.e., no minimum SINR or spectral efficiency that the users should obtain.

Due to our solution approach, it is not necessary to have fixed QoS constraints because we optimize in an adaptive manner the spectral efficiency targets. The proposed power control solution uses FL optimization to obtain the optimal spectral efficiency targets with maximum sum and nonzero spectral efficiency for both users and obtain the powers that achieve these spectral efficiency targets with minimum sum. In the case that predefined QoS constraints on the SINR or spectral efficiency are strictly enforced, the proposed solution approach can consider the SINR targets as minimum spectral efficiencies on the feasibility set of the optimization problem solved by the FL optimization (see [44, Problem 17]). For a summary of the FL power control solution, we refer the reader to Algorithms 1 and 2 in [44]. For the user-frequency channel assignment problem, notice that the problem can be written in an equivalent form as the Axial 3-Dimensional Assignment Problem [44, Section VI], which is a well-known non-deterministic polynomial-time (NP)-hard problem. To solve this problem, we propose a greedy solution with provable guarantees to obtain

the user-frequency assignment in polynomial time. For a summary of the greedy solution, we refer the reader to Algorithm 3 in [44].

12.3.2 Numerical results

In order to evaluate the energy efficiency aspects in an FD cellular network, we consider a single-cell system operating in two small-cell environments [46,47]: urban micro and pico-cell environment, respectively. We assume 2.5 GHz carrier frequency and a system bandwidth of 5 MHz, in which the maximum number of frequency channels is $F = 25$. The total transmit power budget is 38 dBm [48, Table 6.0A], and the transmitting power of the BS is 24 dBm per frequency channel [49, Table 4.5], due to the bandwidth of 5 MHz and the 25 frequency channels considered. We assume that the SI cancellation coefficient β varies from -150 to -70 dB with steps of 10 dB while showing the perfect SI cancellation ($-\infty$ dB). Although the highest SI cancellation obtained experimentally was 110 dB (i.e., $\beta = -110$ dB), we consider higher cancellation values to understand the asymptotic performance until the SI is perfectly cancelled. The parameters of this system are set according to Table 12.1.

In the following, we compare the performance of these algorithms in urban-micro and pico-cell scenarios. Moreover, we refer to the algorithms as

1. proposed greedy assignment solution with the proposed FL power control, named G-FLIP;
2. assignment according to the greedy solution but with equal power allocation, named G-EPA;
3. random assignment with equal power allocation, named R-EPA;
4. random assignment but with the proposed FL power control, named R-FLIP; and
5. HD with maximum transmitted power, named HD.

Table 12.1 Simulation parameters

Parameter	**Value**
Cell radius (urban-micro pico-cell)	(100 m 40 m)
Number of users ($I = J$)	(25)
Monte Carlo iterations	400
Carrier frequency	2.5 GHz
System bandwidth	5 MHz
Number of freq. channels (F)	(25)
Urban-micro LOS/NLOS path-loss model	Set according to [46, Table 4-1]
Pico-cell LOS/NLOS path-loss model	Set according to [47, Table 6.2-1]
Thermal noise power (σ^2)	-116.4 dBm/channel
SI cancelling coefficient (β)	$(-\infty\ -150 \cdots -70)$ dB
Total circuit power (P_c)	24 dBm
SI circuit power (P_s)	24 dBm
User max power ($P^u_{\max}$)	24 dBm
BS max power ($P^d_{\max}$)	24 dBm/channel

Notice that the HD algorithm does not need a channel assignment because the UL and DL do not share the resource. Since there is no user-to-user interference, the spectral efficiency of the BS and UL transmissions is maximized by transmitting with maximum power. Hence, the total power consumption is the sum of the DL-transmitted power from the BS to all users and the sum of the UL-transmitted power by the users.

Figure 12.3 shows the average spectral efficiency, total power consumption, and energy efficiency versus the SI level in the urban-micro scenario. Figure 12.3(a) indicates that the proposed joint solution with greedy assignment and optimal power control, G-FLIP, outperforms HD by approximately 15% and 4% for high and achievable SI cancellation levels (-110 and -100 dB). Due to the higher levels of interference in the urban-micro scenario, the gains achieved by the FD algorithms such as G-FLIP and G-EPA, are small. Although FD communications bring gains in scenarios with wider cell radius, the gains are reduced.

We notice that the solutions with greedy assignment, G-FLIP and G-EPA, have a similar performance when the SI cancellation levels are high and achievable but diverge as the SI cancellation levels decrease. Hence, the impact of a smart assignment is more important when the SI is mild, which is explained by that the system is limited

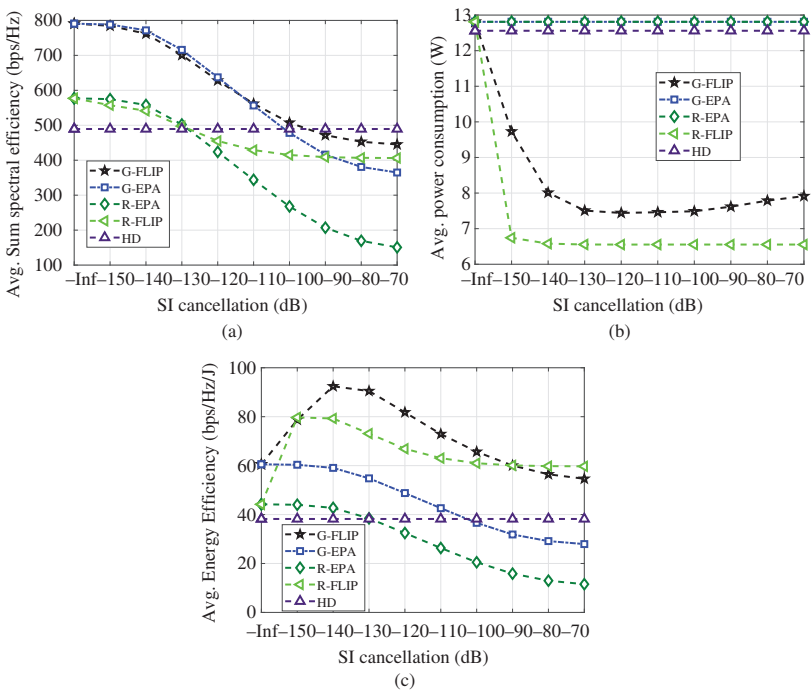


Figure 12.3 The average sum spectral efficiency in the urban-micro scenario versus (a), average power consumption (b), and average energy efficiency (c) SI level

by the user-to-user interference instead of the SI. When the SI cancellation worsens, i.e., the SI cancellation levels increase, the solution with smart power control, R-FLIP, outperforms the solution with smart assignment, G-EPA. When the SI is almost or perfectly cancelled, i.e., β from $-\infty$ to -120 dB, both G-FLIP and G-EPA maintain the best performance among FD algorithms. In addition, G-EPA slightly outperforms G-FLIP, which is due to the proposed FL power control limiting the SINR targets that can be achieved while minimizing the power consumption. Overall, the impact of a smart power control is more important when the SI is strong, which is explained by that the system is limited by SI instead of the user-to-user interference. Nevertheless, the solution with a joint smart power control and assignment, G-FLIP, outperforms other solutions that use either smart assignment or power control.

Figure 12.3(b) shows that the total power consumption of the solutions using the smart FL power control is much lower than the total power consumption when using equal power allocation. For instance, the proposed solution G-FLIP saves approximately 41% in total power when compared with HD. Interestingly, G-FLIP uses more power than R-FLIP. With better user pairs due to a smarter assignment, G-FLIP uses more power to maximize the spectral efficiency than the R-FLIP solution. Notice that G-EPA and R-EPA consume slightly more power than HD. The reason is the SI power consumption term, P_s, which is an additional power consumption not present in the HD. This indicates that FD may consume more power than HD if a proper power control is not applied. When the SI is almost or perfectly cancelled, the power consumption of G-FLIP and R-FLIP gradually increases with a better SI cancellation. This behavior is expected because with a lower interference SI level, the system will dedicate as much power as it is allowed to improve the total spectral efficiency.

Figure 12.3(c) shows the average energy efficiency as defined in (12.3), and we first notice that the proposed solution G-FLIP is always more energy efficient than HD. Using G-FLIP, for -110 dB SI cancellation the relative gain in energy efficiency is approximately 91%, whereas for -70 dB the relative gain is approximately 43%. Interestingly, when G-FLIP is outperformed by HD at -90 dB SI cancellation level, the energy efficiency of G-FLIP is outperformed by R-FLIP. This shows that although the spectral efficiency of G-FLIP is still higher than R-FLIP, the low power consumption of R-FLIP compensates the lack of spectral efficiency. When the SI is almost or perfectly cancelled, the energy efficiency initially increases when the SI cancellation improves from -120 to -140 dB. However, the energy efficiency starts to decrease when the SI cancellation reaches -150 dB. From Figure 12.3(a) and (b), the increase in the spectral efficiency is very small, while the increase in the power consumption is very high, which leads to a sharp decrease in the energy efficiency. Therefore, this behavior shows that further improvements in the SI cancellation may improve the energy efficiency at first but may also decrease the energy efficiency if the transmit power level is too high.

For the urban-micro scenario, FD networks are more energy efficient than HD networks, although not always providing more spectral efficiency than HD. The key role for such behaviors is the smart FL power control solution proposed in [44], which maximizes the spectral efficiency while minimizing the sum power consumption to achieve such optimal spectral efficiencies.

Figure 12.4 shows the average spectral efficiency, total power consumption, and energy efficiency versus the SI level in the pico-cell scenario. Figure 12.4(a) shows that G-FLIP outperforms HD in almost all the SI cancellation levels analyzed, where the relative gain is approximately 29% for −110 dB SI cancellation level and it is outperformed by HD with relative gain of approximately 2% for −70 dB SI cancellation level. Similar to the urban-micro scenario in Figure 12.3(a), the solutions with smart assignment outperform the solutions with smart power control when the SI cancellation level is high, whereas the opposite happens when the SI cancellation level is low. When the SI cancellation is almost or perfectly cancelled, the spectral efficiency becomes similar to Figure 12.3(a); it increases quickly with an improvement in the SI cancellation.

Figure 12.4(b) shows that G-FLIP uses approximately 29% less power than HD and other solutions without smart power control. Similarly, to Figure 12.3(b), G-FLIP consumes approximately 28% more power than R-FLIP, which is explained by the better user-to-user interference with the greedy assignment and reduced path-loss of the pico-cell scenario. As expected from Figure 12.3(b), G-EPA and R-EPA consume slightly more power than HD, and the power consumption quickly increases when the SI cancellation is almost or perfectly cancelled.

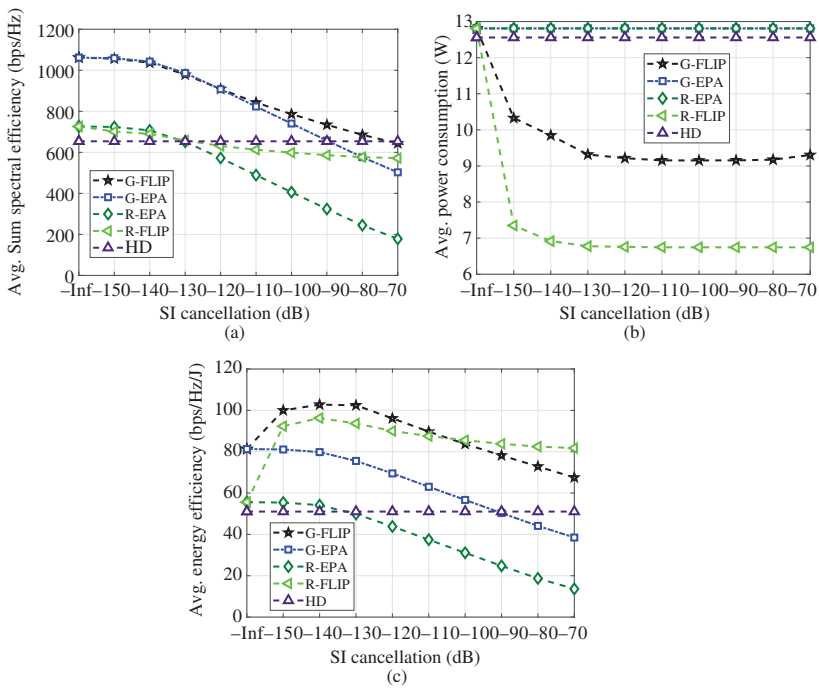


Figure 12.4 The average sum spectral efficiency in the pico-cell scenario versus (a), average power consumption (b), and average energy efficiency (c) SI level

Figure 12.4(c) shows that the energy efficiency of G-FLIP is outperformed by R-FLIP quickly with a decrease in the SI cancellation level, i.e., from −110 to −70 dB. Nevertheless, the solutions with smart power control, G-FLIP and R-FLIP, always outperform HD. For instance, the relative gains of G-FLIP when compared to HD and −110 dB SI cancellation level is approximately 77%, whereas the relative gains of R-FLIP compared to HD and −70 dB SI cancellation level is approximately 63%. Moreover, we notice that the low power consumption from R-FLIP impacts more the energy efficiency than the high spectral efficiency of G-FLIP. Similarly, to Figure 12.3(c), when the SI cancellation is almost or perfectly cancelled, the energy efficiency initially increases and later decreases due to the quick increase in the power consumption.

Therefore, provided that a smart power control is available, FD communications are more energy efficient than HD communications, and specifically for the pico-cell scenario, are also more spectral efficient.

12.4 Multicell analysis

Several techniques have been proposed to alleviate the SI and multiuser interference caused by the FD operation in the context of large-scale wireless networks. mmWave communications are considered as a suitable environment for integrating FD systems due to their unique features, such as the large available bandwidth and the antenna directivity, which can boost the quality of the direct link [50]. Furthermore, recent studies have shown that the higher path-losses of the mmWave signals and their sensitivity to blockages can improve the network performance by mitigating the overall interference [32]. Thus, the codesign of FD radio and mmWave networks is of critical importance in order to combat the severe multiuser interference caused by the FD technology by exploiting the prominent properties of mmWave communications. Moreover, the majority of works mainly focus on the average network performance of a user at a random location within a cell. However, the link quality of a user is subjected to its location. Motivated by the earlier, in this subsection, we propose an analytical framework based on stochastic geometry, which comprises the codesign of FD radio and mmWave cellular networks, where the users can operate either in FD or in HD mode. In addition, the users employ distance-proportional fractional power control and are classified either as CCUs or CEUs based on their location within a cell. Analytical expressions for the coverage performance and the energy efficiency are derived for the considered user classifications.

12.4.1 System model

Consider an mmWave cellular network, where the BSs are spatially distributed according to a homogeneous Poisson Point Process (PPP) Φ with spatial density λ BS/m^2 and transmit power $P^d_{\max}$. Furthermore, the locations of the users follow an arbitrary independent point process Φ_u with spatial density $\lambda_u \ll \lambda$ and transmit power $P^u_{\max}$. Assume that all BSs operate in FD mode, while the users can operate either in FD or

in HD mode (e.g., legacy terminals) with probabilities δ^{F} and δ^{H}, respectively, where $\delta^{\mathrm{F}} + \delta^{\mathrm{H}} = 1$. It is assumed that an FD user is served by its closest BS for both the DL and UL transmissions, while for an HD user it is only considered the DL direction, for which it is served by its closest BS. Moreover, all the FD nodes are assumed to have both DL and UL communication links. Without loss of generality and by following Slivnyak's theorem [51], the analysis concerns the typical user located at the origin but the results hold for all users of the network.

Both the BSs and the user are equipped with multiple transmit/receive antennas, and the actual beam pattern is approximated by the sectorized model [32]. Particularly, the antenna array gain is parameterized by three values: (1) the main-lobe beamwidth $\phi \in [0, 2\pi]$, (2) the main-lobe gain M (dB), and (3) the side-lobe gain m (dB), where $M > m$. It is assumed that perfect beam alignment can be achieved between each user and its serving BS, resulting in an antenna array gain of M^2, denoted by G_0. On the other hand, the beams of interfering links are assumed to be randomly oriented with respect to each other. Therefore, the gain of an interference link seen by a user is a discrete random variable and is given by $G = \{M^2, Mm, m^2\}$ with the corresponding probabilities $p_G = \left\{(\phi/\pi)^2, 2(\phi/\pi)(\pi - \phi)/\pi, ((\pi - \phi)/\pi)^2\right\}$.

All wireless signals are assumed to experience both large-scale path-loss effects and small-scale fading. Specifically, the small-scale fading between two nodes is modeled by Rayleigh fading with unit average power, where different links are assumed to be independent and identically distributed. Hence, the power of the channel fading is an exponential random variable with unit mean, i.e., $h \sim \exp(1)$. For the large-scale path-loss, an unbounded singular path-loss model is considered, i.e., $L(X, Y) = \|X - Y\|^{-a}$, which assumes that the received power decays with the distance between the transmitter located at X and the receiver located at Y, where $a > 2$ is the path-loss exponent. Furthermore, all wireless links exhibit additive white Gaussian noise with zero mean and variance σ^2. Regarding the SI, it is assumed that the FD-capable users and BSs employ imperfect cancellation mechanisms [52]. As such, the residual SI channel coefficient is considered to follow a Nagakami-μ distribution with parameters (μ, β). Therefore, the power gain of the residual SI channel follows a Gamma distribution with mean μ and variance β/μ, i.e., $h_{\mathrm{SI}} \sim \Gamma(\mu, \beta/\mu)$.

A link can be either line of sight (LOS) or non-LOS, depending on whether the BS is visible to the user or not. Let $p(r)$ be a step function that denotes the probability that a link of length r is LOS, i.e., $p(r) = p_{\mathrm{LOS}} \mathbb{1}(r \leq R_B)$, where R_B is the maximum length of an LOS link, $\mathbb{1}(X)$ is the indicator function where $\mathbb{1}(X) = 1$ if X is true, otherwise $\mathbb{1}(X) = 0$, and the constant $p_{\mathrm{LOS}} \in [0, 1]$ represents the average fraction of the LOS area in the ball of radius R_B [32]. It is assumed that the interference effect from the NLOS signals is ignored, since the dominant interference is caused by the LOS signals [32]. Thus, for the PPP Φ, a PPP Φ^{L} with inhomogeneous density $\lambda^{\mathrm{L}}(r) = \lambda p(r)$ is formed, based on the thinning property of the PPPs [51].

Let γ^{D} and γ^{U} represent the achieved DL and UL SINR for the typical receiver (i.e., a user or a BS), which can be expressed as follows:

$$\gamma^{\mathrm{D}} = \frac{P^d_{\max} G_0 L_0(x) h_x}{\sum_{y \in \Phi^{\mathrm{L}} \setminus x} P^d_{\max} G L_0(y) h_y + \sum_{v \in \Psi} P_u G L_0(v) h_v + \mathbb{1}_{\mathrm{FD}} h_{\mathrm{SI}} P^u_{\max} + \sigma^2}, \tag{12.5}$$

$$\gamma^{\mathrm{U}} = \frac{P^{u}_{\max} G_0 L_0(x) \widetilde{h}_x}{\sum_{y \in \Phi^{\mathrm{L}}} P^{d}_{\max} G L_0(y) h_y + \sum_{v \in \Psi \setminus x} P_u G L_0(v) h_v + h_{\mathrm{SI}} P^{d}_{\max} + \sigma^2}, \tag{12.6}$$

where $L_0(x)$ is the path-loss between the transmitter located at X and the receiver located at the origin, h_x and $\widetilde{h}_x$ are the channel fadings between the typical receiver and its serving BS and user, respectively, and $\mathbb{1}_{\mathrm{FD}}$ is the indicator function for the event "the typical user is FD-capable." It is important to mention here that the interference observed by the typical receiver (i.e., a user or a BS) is caused by both the nearby BSs $y \in \Phi^{\mathrm{L}}$ and the active users $v \in \Psi$, where Ψ is the point process that represents the active LOS users. The positions of the active users can be seen as a Voronoi-perturbed lattice process that is not mathematically tractable. Hence, we adopt the approximation proposed in [53] and characterize the point process Ψ as an inhomogeneous PPP with density $\bar{\lambda}^{\mathrm{L}}(r) = \lambda^{\mathrm{L}}(r)\left(1 - \exp\left(-\pi \lambda \delta^{\mathrm{F}} r^2\right)\right)$. Note that the user's performance is strongly dependent on its location with respect to the nearby BSs. In what follows, the details of the considered location-based classification of a user is presented based on the previous system model.

12.4.2 Location-based classification criteria

The spatial partition of a cell into two disjoint subregions, center and edge region, aims to classify the users according to their location. The cell-center region is defined as the region in which the user has significantly smaller distance from its serving BS compared to the distance from the dominant interferer. On the other hand, a user is located in the cell-edge region, if the distances from the serving BS and the dominant interferer are relatively equal.

Let $\mathscr{B}$ represent the region in which a user is classified within a cell, i.e., $\mathscr{B} = \{C, E\}$, where C and E are the cell-center and the cell-edge regions, respectively. A user is presumed to be in the cell-edge region, i.e., $\mathscr{B} = E$, if its distances from the first and the second closest BSs are relatively equal; otherwise, the user is presumed to be in the cell-center region, i.e., $\mathscr{B} = C$. Thus, a user is classified as CEU, if $R_s/R_d > \zeta$, otherwise as CCU, where R_s and R_d denote the distances of the first and the second closest BSs and $\zeta \in [0, 1]$ is a predefined fraction. Specifically, ζ is a metric that determines the size of the cell-center and the cell-edge regions and consequently the classification of users. Specifically, as $\zeta \to 0$ ($\zeta \to 1$), the cell-edge (cell-center) region becomes larger; thus, the number of users that are classified in the cell-edge (cell-center) region is also increased.

Let $p_{\mathscr{B}}$ denote the probability that a user is classified in the region $\mathscr{B}$. Hence, with the use of the joint distance distribution of R_s and R_d, a user is classified in the cell-center region with probability:

$$\begin{aligned} p_C = \Pr\left[\frac{R_s}{R_d} \leq \zeta\right] &= \int_0^{R_B} \left(\int_0^{vR_B} (2\pi\lambda p_{\mathrm{LOS}})^2 r v \exp\left(-\pi \lambda p_{\mathrm{LOS}} v^2\right) dr \right) dv \\ &= \zeta^2 \left(1 - \exp\left(-\pi \lambda p_{\mathrm{LOS}} R_B^2\right)\left(1 + \pi \lambda p_{\mathrm{LOS}} R_B^2\right)\right), \end{aligned} \tag{12.7}$$

while a user is classified in the cell-edge region with probability:

$$p_E = \Pr\left[\frac{R_s}{R_d} > \zeta\right] = 1 - p_C. \tag{12.8}$$

Lemma 12.1. *The cumulative distribution function of the distance between a user, which is classified in the region $\mathscr{B}$, and its serving BS is given by*

$$F_R(r|\mathscr{B}) = \begin{cases} \dfrac{e^{-\frac{\pi\lambda p_{\mathrm{LOS}} r^2}{\zeta^2}} - e^{-\frac{\pi\lambda p_{\mathrm{LOS}} R_B^2}{\zeta^2}} - \pi\lambda p_{\mathrm{LOS}}(R_B^2 - r^2)e^{-\pi\lambda p_{\mathrm{LOS}} R_B^2}}{p_C} & \text{if } \mathscr{B} = C, \\ \dfrac{\zeta^2\left(e^{-\frac{2\pi\lambda R_B^2 p_{\mathrm{LOS}}}{\zeta^2}} - e^{-\frac{\pi\lambda r^2 p_{\mathrm{LOS}}}{\zeta^2}}\right) - e^{-\pi\lambda p_{\mathrm{LOS}} R_B^2} + e^{-\pi\lambda p_{\mathrm{LOS}} r^2}}{p_E} & \text{if } \mathscr{B} = E. \end{cases} \tag{12.9}$$

and therefore its probability density function (PDF), is given by

$$f_R(r|\mathscr{B}) = \frac{d}{dr}\left(1 - F_R(r|\mathscr{B})\right). \tag{12.10}$$

12.4.3 Hybrid-duplex heterogeneous networks

Success probability: An important performance metric is the ability of a receiver to successfully decode the received signal. This metric is characterized by the probability that the observed SINR at the receiver (i.e., a user or a BS), which is classified in the region $\mathscr{B}$, is greater than a predefined threshold θ, i.e., $\Pr[\gamma^{\mathrm{D}} > \theta|\mathscr{B}]$ and $\Pr[\gamma^{\mathrm{U}} > \theta|\mathscr{B}]$ for the DL and UL transmission, respectively. Assuming that the typical user is spatially classified in the region $\mathscr{B}$, the DL success probability is given by

$$\begin{aligned} \Pr[\gamma^{\mathrm{D}} > \theta|\mathscr{B}] &= \delta^{\mathrm{H}} \int_0^{R_B} \mathscr{L}^{\mathrm{D}}(s_{\mathrm{D}})\mathscr{L}^{\mathrm{UE}}(s_{\mathrm{D}}, 0)\exp\left(-s_{\mathrm{D}}\sigma^2\right) f_R(u|\mathscr{B})du \\ &\quad + \delta^{\mathrm{F}} \int_0^{R_B} \frac{\mathscr{L}^{\mathrm{D}}(s_{\mathrm{D}})\mathscr{L}^{\mathrm{UE}}(s_{\mathrm{D}}, 0)}{\left(1 + \frac{s_{\mathrm{D}}\beta P_{\max}^u}{\mu}\right)^{\mu}} \exp\left(-s_{\mathrm{D}}\sigma^2\right) f_R(u|\mathscr{B})du, \end{aligned} \tag{12.11}$$

while the UL success probability is given by

$$\Pr[\gamma^{\mathrm{U}} > \theta|\mathscr{B}] = \int_0^{R_B} \frac{\mathscr{L}^{\mathrm{U}}(s_{\mathrm{U}})\mathscr{L}^{\mathrm{UE}}(s_{\mathrm{U}}, u)}{\left(1 + \frac{s_{\mathrm{U}}\beta P_{\max}^d}{\mu}\right)^{\mu}} \exp\left(-s_{\mathrm{U}}\sigma^2\right) f_R(u|\mathscr{B})du, \tag{12.12}$$

where

$$\mathscr{L}^{\mathrm{D}}(s, u) = \prod_G \exp\left(-\frac{2\pi\lambda p_G}{a-2}\frac{sGP_{\max}^d}{u^{a-2}}\,{}_2F_1\left[1, \frac{a-2}{a}, 2-\frac{2}{a}, -\frac{sGP_{\max}^d}{u^a}\right]\right), \tag{12.13}$$

$$\mathscr{L}^{\mathrm{U}}(s,u)=\prod_{G}\exp\left(-\frac{2\pi^2\lambda p_G}{a}\left(sGP_{\max}^{d}\right)^{\frac{2}{a}}\csc\left(\frac{2\pi}{a}\right)\right), \tag{12.14}$$

$$\mathscr{L}^{\mathrm{UE}}(s,x)=\prod_{G}\exp\left(-2\pi p_G sGP_u\int_{x}^{R_B}\frac{r^{-a+1}\bar{\lambda}^{\mathrm{L}}(r)}{1+sGP_u r^{-a}}dr\right), \tag{12.15}$$

$s=\{s_{\mathrm{D}},s_{\mathrm{U}}\}=\left\{\frac{\theta u^a}{P_{\max}^{d}G_0},\frac{\theta u^a}{P_{\max}^{u}G_0}\right\}$, $G=\{M^2,Mm,m^2\}$, $\csc(\cdot)$ and ${}_2F_1[\cdot,\cdot,\cdot,\cdot]$ denote the cosecant and the hypergeometric functions, respectively.

Energy efficiency: Energy efficiency is another important performance metric for wireless communication systems. Based on the user's operation, i.e., FD or HD mode, the energy efficiency is calculated as the ratio of the throughput to the total power consumption and is given by

$$EE_{\mathscr{B}}=\delta^{\mathrm{F}}\frac{T_{\mathscr{B}}^{\mathrm{D}}+T_{\mathscr{B}}^{\mathrm{U}}}{P_{\mathrm{tot}}^{\mathrm{F}}}+\delta^{\mathrm{H}}\frac{T_{\mathscr{B}}^{\mathrm{D}}}{P_{\mathrm{tot}}^{\mathrm{H}}}, \tag{12.16}$$

where $T_{\mathscr{B}}^{\mathrm{D}}$ and $T_{\mathscr{B}}^{\mathrm{U}}$ represent the achieved throughput for the DL and the UL direction, respectively; $P_{\mathrm{tot}}^{\mathrm{F}}$ and $P_{\mathrm{tot}}^{\mathrm{H}}$ denote the total power consumption for scenario where the typical user operates in FD and in HD mode, respectively, containing the power consumed for the user's communication, for the circuit and the SI cancellation process. Hence, $P_{\mathrm{tot}}^{\mathrm{F}}$ and $P_{\mathrm{tot}}^{\mathrm{H}}$ can be expressed as

$$P_{\mathrm{tot}}^{\mathrm{F}}=\frac{P_{\max}^{u}+P_{\max}^{d}}{\omega}+P_c+P_s \quad \text{and} \quad P_{\mathrm{tot}}^{\mathrm{H}}=\frac{P_{\max}^{u}}{\omega}+P_c, \tag{12.17}$$

where ω is the power amplifier efficiency, P_c and P_s represent the power consumed by the circuit and the SI cancellation process, respectively.

Regarding the user's throughput in the context of the considered system model, analytical expressions are obtained by using the derived success probabilities. Specifically, by the change of variable $\tau\to 2^x-1$, the DL throughput of a user that is classified in the region $\mathscr{B}$ is given by

$$T_{\mathscr{B}}^{\mathrm{D}}=\int_{0}^{\infty}\Pr[\log_2(1+\gamma^{\mathrm{D}})>x|\mathscr{B}]dx=\frac{1}{\ln(2)}\int_{0}^{\infty}\frac{\Pr[\gamma^{\mathrm{D}}>\tau|\mathscr{B}]}{\tau+1}d\tau, \tag{12.18}$$

while the UL throughput of a user that is classified in the region $\mathscr{B}$ is given by

$$T_{\mathscr{B}}^{\mathrm{U}}=\int_{0}^{\infty}\Pr[\log_2(1+\gamma^{\mathrm{U}})>x|\mathscr{B}]dx=\frac{1}{\ln(2)}\int_{0}^{\infty}\frac{\Pr[\gamma^{\mathrm{U}}>\tau|\mathscr{B}]}{\tau+1}d\tau, \tag{12.19}$$

where $\Pr[\gamma^{\mathrm{D}}>\tau|\mathscr{B}]$ and $\Pr[\gamma^{\mathrm{U}}>\tau|\mathscr{B}]$ represent the DL and UL success probabilities of a user which is classified in the region $\mathscr{B}$ and are given by (12.11) and (12.12), respectively.

12.4.4 Numerical results

In this section, numerical examples are given for the energy efficiency of the FD-enabled cellular networks presented earlier. The simulation system uses the parameters: $a = 4$, $\lambda = 5 \times 10^{-5}$, $P^{d}_{\max} = 30$ dBm, $P^{u}_{\max} = 20$ dBm, $\phi = \pi/3$, $M = 10\,\text{dB}, m = -10\,\text{dB}, \mu = 4, p_{\text{LOS}} = 0.7, R_B = 200\,\text{m}, \sigma^2 = -40\,\text{dB}, \omega = 0.25$, $P_c = 24$ dBm, and $P_s = 24$ dBm. For the considered network deployment, a user can be spatially classified either as CCU or as CEU.

Figure 12.5(a) plots the energy efficiency as a function of the fraction of FD users for the spatial classifications considered with $\zeta = \{0.6, 0.9\}$ and $\beta = -60$ dB. The energy efficiency curves are given by Monte Carlo simulations and perfectly match with the expressions that are obtained by numerically evaluating (12.16). The numerical evaluation of the theoretical expressions corresponds to the solid lines. The first important observation is that CCUs obtain significantly better energy efficiency compared to the energy efficiency obtained by the CEUs. This was expected since the serving BS of a CCU is closer compared to its interfering BSs, and therefore, higher SINR can be achieved. On the other hand, a CEU observes significantly

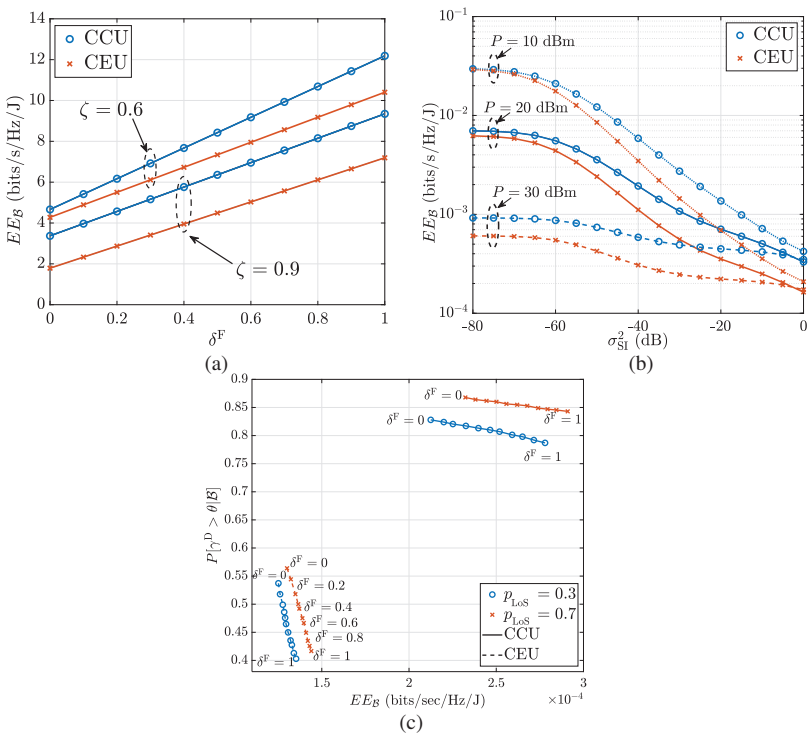


Figure 12.5 The energy efficiency (EE) versus (a) the fraction of FD users, (b) SI level, and (c) the trade-off between the coverage performance and energy efficiency

higher interference compared to the received signal strength, achieving a lower SINR compared to the observed SINR of a CCU. Another important observation is that by decreasing the cell-center region and consequently by increasing the cell-edge region, i.e., $\zeta \to 0$, the energy efficiency of both CCUs and CEUs increases. This observation was expected, since by decreasing the parameter ζ, the distance of both the CCUs and the CEUs from their serving BS is decreased, and the distance from their dominant interfering BS is increased; therefore, an enhanced SINR is observed. Finally, the difference in terms of energy efficiency between a CCU and a CEU decreases as $\zeta \to 0$, since by increasing the cell-edge region, the distance between a CEU and its serving BS is comparable to the distance between a CCU and its serving BS.

Figure 12.5(b) illustrates the effect of the residual SI on the energy efficiency of the considered system model for $P^{d}_{\max} = \{10, 20, 30\}$ dBm and $\zeta = 0.7$. It can be easily observed that by increasing the ability of the network's nodes to cancel the SI, i.e., $\beta \to -\infty$, the energy efficiency of the considered network is increased for both user classifications. This observation was expected, since by decreasing the residual SI at the BSs and the users, the aggregate received interference at the nodes is decreased, and therefore an increased SINR is observed. Furthermore, Figure 12.5(b) demonstrates the effect of transmit power on the energy efficiency. As can be seen, by increasing the transmit power of the network's nodes, the energy efficiency for both user's classification decreases.

Figure 12.5(c) demonstrates the trade-off between the energy efficiency and the DL coverage performances with respect to the fraction of FD users δ^{F}, for different blockage constants p_{LOS}. Each point in the curves represents the trade-off between the two performance metrics for a given fraction of FD users. As expected, by increasing the fraction of FD users, the coverage performance decreases due to the increased multiuser interference. Conversely, the increased number of FD users positively affects the energy efficiency, since the FD mode provides increased network throughput. Another important observation is that the increased number of FD users and consequently the increased multiuser interference mainly compromise the performance of the CEUs. The main justification is that the CEUs are more vulnerable in interference changes due to their small distances from the BSs that cause interference. Finally, Figure 12.5(c) reveals the negative impact of blockages on both the energy efficiency and the DL coverage performance of the network. This was expected since high blockage densities reduce the probability of a receiver to communicate with an LOS transmitter, and therefore the observed SINR is decreased.

12.5 Conclusion

With the recent advancements in the SI cancellation for many scenarios, the promising enhancements of FD in the spectral efficiency, medium access control layer, and short distances are forthcoming. To make a smooth and efficient deployment, the energy efficiency aspects must be taken into account and this is the goal of this chapter.

After the introduction of FD communications and some application areas and the discussion about the challenges of energy-efficient FD networks, we proposed four main design techniques to deal with such challenges. For each technique, we presented

a comprehensive literature review highlighting different directions that were taken to improve the energy efficiency. In addition, for each of the four techniques, we provided interesting research directions that are still open and need further investigation from the FD community.

Then, we presented results for enhancing the energy efficiency in a single-cell scenario. Including urban-micro and pico-cell, we proposed a power control and user-frequency assignment to maximize the sum spectral efficiency while minimizing the total power consumption. For both scenarios, the results indicated that the impact of a smart assignment in the spectral and energy efficiency is higher when the SI cancellation level is high, whereas the power control impacts more when the SI cancellation level is low. Moreover, using our proposed solution, FD communications are more energy efficient than HD communications for various SI cancellation levels and, specifically for the pico-cell scenario, are also more spectral efficient.

In the multicell scenario, we proposed an analytical framework based on stochastic geometry and studied the performance of FD-mmWave cellular networks. The developed framework takes into account the ability of users to operate either in FD or in HD mode and their classification either as CCUs or CEUs. We derived analytical expressions for both the coverage performance and energy efficiency for each user classification, and the impact of blockage density, residual LI, and fraction of FD users has been discussed. The results reveal that HD is beneficial for CEUs since it offers better coverage performance. On the other hand, FD provides a significantly larger energy efficiency to CCUs, with the cost of a slightly reduced coverage performance. Finally, we have shown that the combination of FD radio with mmWaves provides significant gains to cellular networks, as it can potentially increase the energy efficiency as well as mitigate the severe multiuser interference.

For both small- and multicell scenarios, the results illustrated the gains in terms of energy efficiency provided due to FD communications. Nevertheless, the key challenges of FD communications—to provide energy and spectral efficient networks, the SI and user-to-user interference, and the design techniques detailed in this chapter—are still an area of ongoing research that requires further investigation. With the vast application domain, including relaying, bidirectional, and cellular networks, as well as the suitability to techniques essential to 5G such as massive MIMO and mmWave, it is expected that FD communications will play an increasingly important role in beyond 5G and sixth-generation communications.

References

[1] Parkvall S, Dahlman E, Furuskar A, *et al.* NR: The New 5G Radio Access Technology. IEEE Communications Magazine. 2017;1(4):24–30.

[2] Ericsson. Ericsson Launches Enhanced 5G Deployment Options; 2019. Available from: https://bit.ly/2KWoI3m.

[3] ITU-R. Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s). ITU; 2017.

[4] ITU-R. IMT Vision – Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond. ITU; 2015.
[5] Giordani M, Polese M, Mezzavilla M, Rangan S, and Zorzi M Towards 6G Networks: Use Cases and Technologies. IEEE Communications Magazine. 2020;58(3):55–61.
[6] Zhang Z, Long K, Vasilakos A, *et al.* Full-Duplex Wireless Communications: Challenges, Solutions, and Future Research Directions. IEEE Communication Surveys and Tutorials. 2016;104(7):1369–1409.
[7] Goyal S, Liu P, Panwar SS, *et al.* Full Duplex Cellular Systems: Will Doubling Interference Prevent Doubling Capacity? IEEE Communications Magazine. 2015;53(5):121–127.
[8] Bharadia D, McMilin E, and Katti S. Full Duplex Radios. SIGCOMM Computer Communication Review. 2013;43(4):375–386.
[9] Duarte M, Sabharwal A, Aggarwal V, *et al.* Design and Characterization of a Full-Duplex Multiantenna System for WiFi Networks. IEEE Transactions on Vehicular Technology. 2014;63(3):1160–1177.
[10] Korpi D, Tamminen J, Turunen M, *et al.* Full-Duplex Mobile Device: Pushing the Limits. IEEE Communications Magazine. 2016;54(9):80–87.
[11] Reiskarimian N, Dinc T, Zhou J, *et al.* One-Way Ramp to a Two-Way Highway: Integrated Magnetic-Free Nonreciprocal Antenna Interfaces for Full-Duplex Wireless. IEEE Microwave Magazine. 2019;20(2):56–75.
[12] Sabharwal A, Schniter P, Guo D, *et al.* In-Band Full-Duplex Wireless: Challenges and Opportunities. IEEE Journal on Selected Areas in Communications. 2014;32(9):1637–1652.
[13] Riihonen T, Werner S, and Wichman R. Mitigation of Loopback Self-Interference in Full-Duplex MIMO Relays. IEEE Transactions on Signal Processing. 2011;59(12):5983–5993.
[14] Everett E, Shepard C, Zhong L, *et al.* SoftNull: Many-Antenna Full-Duplex Wireless via Digital Beamforming. IEEE Transactions on Wireless Communications. 2016;15(12):8077–8092.
[15] Korpi D, Heino M, Icheln C, *et al.* Compact Inband Full-Duplex Relays With Beyond 100 dB Self-Interference Suppression: Enabling Techniques and Field Measurements. IEEE Transactions on Antennas and Propagation. 2017;65(2):960–965.
[16] Chung M, Sim MS, Kim DK, *et al.* Compact Full Duplex MIMO Radios in D2D Underlaid Cellular Networks: From System Design to Prototype Results. IEEE Access. 2017;5:16601–16617.
[17] Krishnaswamy H and Zussman G. A Full-Duplex Chip—One That Can Send and Receive Simultaneously—Could Double Phone-Network Data Capacity. IEEE Spectrum. 2016.
[18] Chen T, Dastjerdi MB, Zhou J, *et al.* Wideband full-duplex wireless via frequency-domain equalization: Design and experimentation. In: International Conference on Mobile Computing and Networking (ACM MobiCom); 2019, Article 3, pp. 1–16.

[19] Li H, Van Kerrebrouck J, Caytan O, *et al.* Self-Interference Cancellation Enabling High-Throughput Short-Reach Wireless Full-Duplex Communication. IEEE Transactions on Wireless Communications. 2018;17(10): 6475–6486.

[20] Dorsch J. Kumu Networks: Full Duplex On One Channel; 2019. [Online; accessed in 1-March-2019]. Available from: https://goo.gl/ierMm6.

[21] Thilina KM, Tabassum H, Hossain E, *et al.* Medium Access Control Design for Full Duplex Wireless Systems: Challenges and Approaches. IEEE Communications Magazine. 2015;53(5):112–120.

[22] Buzzi S, Chih-Lin I, Klein TE, *et al.* A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. IEEE Journal on Selected Areas in Communications. 2016;34(4):697–709.

[23] Nguyen D, Tran LN, Pirinen P, *et al.* Precoding for Full Duplex Multiuser MIMO Systems: Spectral and Energy Efficiency Maximization. IEEE Transactions on Signal Processing. 2013;61(16):4038–4050.

[24] Wei Z, Zhu X, Sun S, *et al.* Energy-Efficiency of Millimeter-Wave Full-Duplex Relaying Systems: Challenges and Solutions. IEEE Access. 2016;4: 4848–4860.

[25] Wen D, Yu G, Li R, *et al.* Results on Energy- and Spectral-Efficiency Tradeoff in Cellular Networks With Full-Duplex Enabled Base Stations. IEEE Transactions on Wireless Communications. 2017;16(3):1494–1507.

[26] Nguyen DN, Krunz M, and Dutkiewicz E. Full-Duplex MIMO Radios: A Greener Networking Solution. IEEE Transactions on Green Communications and Networking. 2018;2(3):652–665.

[27] Zhang Y, Xiao M, Han S, *et al.* On Precoding and Energy Efficiency of Full-Duplex Millimeter-Wave Relays. IEEE Transactions on Wireless Communications. 2019;18(3):1943–1956.

[28] Taghizadeh O, Neuhaus P, Mathar R, *et al.* Secrecy Energy Efficiency of MIMOME Wiretap Channels With Full-Duplex Jamming. IEEE Transactions on Communications. 2019;67(8):5588–5603.

[29] Psomas C, Mohammadi M, Krikidis I, *et al.* Impact of Directionality on Interference Mitigation in Full-Duplex Cellular Networks. IEEE Transactions on Wireless Communications. 2017;16(1):487–502.

[30] Marandi L, Naslcheraghi M, Ghorashi SA, *et al.* Delay Analysis in Full-Duplex Heterogeneous Cellular Networks. IEEE Transactions on Vehicular Technology. 2019;68(10):9713–9721.

[31] Goyal S, Galiotto C, Marchetti N, *et al.* Throughput and coverage for a mixed full and half duplex small cell network. In: IEEE International Conference on Communications (ICC). Kuala Lumpur: IEEE; 2016. pp. 1–7.

[32] Skouroumounis C, Psomas C, and Krikidis I. Heterogeneous FD-mm-Wave Cellular Networks With Cell Center/Edge Users. IEEE Transactions on Communications. 2019;67(1):791–806.

[33] Vermeulen T and Pollin S. Energy-delay analysis of full duplex wireless communication for sensor networks. In: IEEE Global Communications Conference; 2014. pp. 455–460.

[34] Al-Kadri MO, Aijaz A, and Nallanathan A. An Energy-Efficient Full-Duplex MAC Protocol for Distributed Wireless Networks. IEEE Wireless Communications Letters. 2016;5(1):44–47.

[35] Kobayashi M, Murakami R, Kizaki K, *et al.* Wireless Full-Duplex Medium Access Control for Enhancing Energy Efficiency. IEEE Transactions on Green Communications and Networking. 2018;2(1):205–221.

[36] Li D, Cheng J, and Leung VCM. Adaptive Spectrum Sharing for Half-Duplex and Full-Duplex Cognitive Radios: From the Energy Efficiency Perspective. IEEE Transactions on Communications. 2018;66(11):5067–5080.

[37] Askar R, Hamdan A, Keusgen W, *et al.* Analysis of utilizing lossless networks for self-interference cancellation purpose. In: IEEE Wireless Communications and Networking Conference (WCNC); 2018. pp. 1–6.

[38] Debaillie B, van den Broek D, Lavín C, *et al.* Analog/RF Solutions Enabling Compact Full-Duplex Radios. IEEE Journal on Selected Areas in Communications. 2014;32(9):1662–1673.

[39] Ju H and Zhang R. Optimal Resource Allocation in Full-Duplex Wireless-Powered Communication Network. IEEE Transactions on Communications. 2014;62(10):3528–3540.

[40] Kang X, Ho CK, and Sun S. Full-Duplex Wireless-Powered Communication Network With Energy Causality. IEEE Transactions on Communications. 2015;14(10):5539–5551.

[41] Mohammadi M, Chalise BK, Suraweera HA, *et al.* Throughput Analysis and Optimization of Wireless-Powered Multiple Antenna Full-Duplex Relay Systems. IEEE Transactions on Communications. 2016;64(4):1769–1785.

[42] Yadav A, Dobre OA, and Poor HV. Is Self-Interference in Full-Duplex Communications a Foe or a Friend? IEEE Signal Processing Letters. 2018;25(7): 951–955.

[43] Chu Z, Zhou F, Xiao P, *et al.* Resource Allocation for Secure Wireless Powered Integrated Multicast and Unicast Services With Full Duplex Self-Energy Recycling. IEEE Transactions on Wireless Communications. 2019;18(1): 620–636.

[44] da Silva Jr JMB, Fodor G, and Fischione C. Fast-Lipschitz Power Control and User-Frequency Assignment in Full-Duplex Cellular Networks. IEEE Transactions on Wireless Communications. 2017;16(10):6672–6687.

[45] Fischione C. Fast-Lipschitz Optimization With Wireless Sensor Networks Applications. IEEE Transactions on Automatic Control. 2011;56(10): 2319–2331.

[46] WINNER+. D5.3: WINNER+ Final Channel Models. Wireless World Initiative New Radio +; 2010. D5.3.

[47] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Further enhancements to LTE Time Division Duplex (TDD) for Downlink-Uplink (DL-UL) Interference Management and Traffic Adaptation. 3rd Generation Partnership Project (3GPP); 2012. 36.828.

[48] 3GPP. Base Station (BS) Radio Transmission and Reception (FDD). 3rd Generation Partnership Project (3GPP); 2017. 25.104.

[49] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) System Scenarios. 3rd Generation Partnership Project (3GPP); 2012. 36.942.

[50] Turgut G and Esma MC. Coverage in Heterogeneous Downlink Millimeter Wave Cellular Networks. IEEE Transactions on Communications. 2017;65(1):4463–4477.

[51] Błaszczyszyn B, Haenggi M, Keeler P, *et al.* Stochastic geometry analysis of cellular networks. Cambridge, United Kingdom; New York, NY, USA: Cambridge University Press; 2018.

[52] Atzeni I and Kountouris M. Full-Duplex MIMO Small-Cell Networks With Interference Cancellation. IEEE Transactions on Wireless Communications. 2017;16(12):8362–8376.

[53] Singh S, Zhang X, and Andrews JG. Joint Rate and SINR Coverage Analysis for Decoupled Uplink-Downlink Biased Cell Associations in HetNets. IEEE Transactions on Wireless Communications. 2015;14(10):5360–5373.

Chapter 13

Energy-efficient resource allocation design for NOMA systems

Zhiqiang Wei[1], Yuanxin Cai[1], Jun Li[2], Derrick Wing Kwan Ng[1] and Jinhong Yuan[1]

This chapter introduces the basic concepts on energy efficiency and non-orthogonal multiple access (NOMA) to unlock the potentials of future communication networks. The energy-efficient resource allocation design for NOMA systems is formulated as a non-convex optimization problem. Based on the fractional programming and successive convex approximation (SCA), a generic algorithm is proposed to achieve a suboptimal solution of the formulated problem. Simulation results are provided to verify the convergence of the proposed algorithm and to evaluate the system energy efficiency of the proposed design.

13.1 Introduction

13.1.1 Background

The fast development of wireless communications worldwide fuels the massive growth in the number of wireless communication devices and sensors for emerging applications such as smart logistics and transportation, environmental monitoring, energy management, safety management, and industry automation, just to name a few. In the Internet-of-Things era [1], it is expected that there will be 50 billion wireless communication devices connected worldwide with a connection density up to a million devices per km^2 [2,3]. The massive number of devices and associated explosive data traffic poses challenging requirements, such as massive connectivity [4] and ultra-high energy efficiency for next-generation wireless networks [5,6].

Recently, NOMA has drawn a lot of attentions as an important enabling technique to fulfill the challenging requirements of the next generation wireless networks, such as massive connectivity, high spectral efficiency, and improved energy

[1]School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia
[2]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China

efficiency [6–9]. In contrast to conventional orthogonal multiple access (OMA) schemes, NOMA allows multiple users to share the same degrees of freedom (DoF)* via power- and/or code-domain multiplexing at transmitter side and to perform multiuser detection (MUD) at receiver side to retrieve the messages of multiple users [7–9]. One can refer to various survey papers in [7–9] for more details. It is known that NOMA is beneficial to supporting a large number of connections to achieve efficient communications by introducing controllable symbol collisions in the same DoF. In recent years, NOMA has been extensively studied in the literature, e.g., [3,10–13], and it has been shown that NOMA can achieve a considerable performance gain over conventional OMA schemes in terms of spectral efficiency and energy efficiency. In fact, the concept of non-orthogonal transmissions can be dated back to the information theoretic results from the 1970s [14–16], which serves as a foundation for the recent development of NOMA [17–21]. However, NOMA has rekindled the interests of researchers, due to the benefit of the recent advances in signal processing and silicon technologies [22,23]. For example, the industrial community has proposed up to 16 various forms of NOMA as the potential multiple access schemes for the forthcoming 5G networks [24].

Meanwhile, energy-efficient communications have become an important focus in both academia and industry due to the growing demands of energy consumption and the arising environmental concerns around the world [25,26]. The escalating energy costs and the associated global carbon dioxide (CO_2) emission of information and communication technology devices have stimulated the interest of researchers in an emerging area of energy-efficient radio management. Therefore, studying energy-efficient radio management via exploiting limited system resources is critical to strike a balance between system energy consumption and throughput [27–34]. The optimization framework of sequential fractional programming for energy-efficient resource allocation design was reviewed in [29]. A complete framework for the optimization of energy efficiency is reviewed in [25], including fractional programming and generalized fractional programming. The energy-efficient resource allocation design problems of orthogonal frequency-division multiple access (OFDMA) communication systems have been addressed in [30–33]. In this chapter, we focus on the energy-efficient resource allocation design for NOMA communication systems.

13.1.2 Organization

The rest of the chapter is organized as follows. In Section 13.2, we discuss some basic concepts on NOMA, including the code-domain NOMA, the power-domain NOMA, the downlink NOMA, and the uplink NOMA. Then, the definition of the energy efficiency of NOMA systems is presented and the trade-off between the spectral efficiency and energy efficiency is revealed in Section 13.3. In Section 13.4, we formulate the energy-efficient resource allocation design in a NOMA system as a non-convex optimization problem. Also, we present a generic solution to solve the formulated problem via the fractional programming and the SCA methods. In Section 13.5, as an illustrative example, the energy-efficient power allocation and user scheduling

*Note that DoF is defined as the received signal space according to [14].

design is formulated for a downlink multicarrier NOMA (MC-NOMA) system and is solved according to our proposed generic solution. Section 13.6 provides some simulation results of the proposed algorithms for the considered MC-NOMA system. Then, Section 13.7 concludes this chapter.

13.1.3 Notations

Notations used in this chapter are as follows. Boldface capital and lower case letters are reserved for matrices and vectors, respectively. $\mathbb{C}^{M\times N}$ denotes the set of all $M \times N$ matrices with complex entries; $\mathbb{R}^{M\times N}$ denotes the set of all $M \times N$ matrices with real entries; $\mathbb{Z}^{M\times N}$ denotes the set of all $M \times N$ matrices with entries of integer number; and $\mathbb{B}^{M\times N}$ denotes the set of all $M \times N$ matrices with binary entries. $(\cdot)^{\mathrm{T}}$ denotes the transpose of a vector or a matrix; $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose of a vector or a matrix; $|\cdot|$ denotes the absolute value of a complex scalar; $\|\cdot\|$ denotes the Euclidean vector norm; $[\cdot]^{-1}$ denotes the inverse of a matrix; $\nabla_x f(x)$ denotes the gradient of $f(x)$ with respect to (w.r.t.) x; and diag$\{\mathbf{x}\}$ denotes a diagonal matrix with the entries of $\mathbf{x}$ on its diagonal. The circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu, \sigma^2)$.

13.2 Fundamentals of NOMA

13.2.1 From OMA to NOMA

Radio access technologies for cellular communications are characterized by multiple access schemes. Up to now, OMA has been adopted in previous generations of cellular networks, where different users are allocated to orthogonal resources in either time, frequency, or code-domain in order to mitigate multiple access interference. In particular, it is frequency-division multiple access for the first generation, time-division multiple access (TDMA) for the second generation (2G), code-division multiple access (CDMA) used by both 2G and the third generation, and OFDMA for current 4G. However, OMA schemes are not sufficient to support the upcoming wireless applications with massive connectivity and diverse quality of service (QoS) requirements [7–9]. In fact, due to the limited DoF, some users with better channel quality have a higher priority to be served, while other users with poor channel quality are suspended from service temporarily, which leads to unfairness and large latency in the system. Besides, the system resources are under-utilized when allocating some DoF only to users with poor channel quality from the overall system performance point of view.

In contrast to conventional OMA schemes, NOMA transmission techniques intend to share DoF among users and consequently need to employ MUD to separate interfered users sharing the same DoF, as illustrated in Figure 13.1. A comparative study of OMA and NOMA was carried out in [35]. By allowing multiple users to share the same DoF concurrently, NOMA mainly enjoys the following two advantages:

- **Supporting overload transmission**: The massive growth in the number of wireless communication devices and sensors creates a serious performance bottleneck in realizing reliable and ubiquitous wireless communication networks [3,4,36].

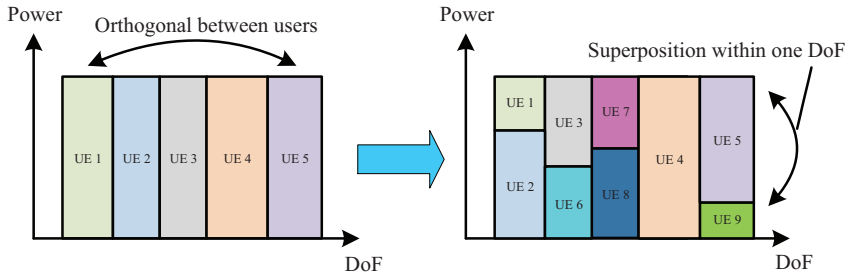


Figure 13.1 From OMA to NOMA via power-domain multiplexing the signals of user equipments (UEs)

It is highly desirable to have an innovative multiple access technique that is able to accommodate more users with limited system resource, i.e., time–frequency resource block, spatial subspace, and/or radio frequency chain. In general, a set of system resources can be interpreted as a kind of system DoF, which allows multiple data streams to be transmitted on each DoF without interfering each other. In other words, the system DoF is equal to the maximum number of the independently deliverable data streams. When the number of users is larger than the system DoF, an overload scenario occurs. In such a scenario, NOMA is an efficient multiple access scheme that paves the way to support overload transmission via DoF sharing. In particular, NOMA transmission technique allows multiple users to simultaneously share the same DoF by introducing controllable interference. Consequently, the system performance and user fairness can be improved by increasing the number of supported connections. Conceptually, NOMA is a generalization of OMA where the latter is more conservative, which does not allow DoF sharing among users. Despite the inherent DoF-sharing feature of NOMA, user scheduling and resource allocation design serve as the key to support overload transmission to achieve a spectral-efficient and/or energy-efficient performance, which is one of the main focuses of this chapter.

- **Improving flexibility in resource allocation**: In contrast to OMA scheme where a resource block is allocated exclusively to a single user, NOMA can utilize the limited spectrum more efficiently by allowing strong users to share the resource block of weak users without compromising much their performance. From the optimization point of view, NOMA relaxes the orthogonality constraint of OMA, which enables a more flexible management of radio resources and offers an efficient way to improve the spectral efficiency and/or energy efficiency via appropriate resource allocation [37]. In addition, it is well known that the potential spectral efficiency gain of NOMA over OMA is larger when channel conditions of the multiplexed users become more distinctive [38]. However, a higher transmit power is needed to satisfy the QoS requirement of the weak user if its channel condition is unsatisfactory. Therefore, there is a nontrivial trade-off between the spectral efficiency and energy efficiency that should be taken into account for resource allocation algorithm design for NOMA.

In general, according to the domain of multiplexing, we can divide the existing NOMA techniques into two categories [9], i.e., code-domain NOMA and power-domain NOMA.

13.2.2 Code-domain NOMA

The code-domain NOMA techniques, including low-density spreading (LDS) [39–42], sparse code multiple access (SCMA) [43–45], pattern division multiple access [46], etc., introduce redundancy via coding/spreading to facilitate users' signals separation at the receiver. For instance, SCMA is a multidimensional codebook-based non-orthogonal spreading technique, which enables SCMA to enjoy the shaping gain of multidimensional constellation compared to simple repetition code of LDS multiple access [43]. As the codewords of SCMA are sparse, it allows system overloading to enable massive connectivity while keeping moderate complexity of the receiver. The block diagram of a SCMA transmitter and a receiver in an uplink cellular system is illustrated in Figure 13.2.[†]

At the transmitter side, the SCMA encoder maps the encoded bits to a sparse complex vector via a constellation generation function and a sparse code mapper. The SCMA codeword of user m is given by

$$\mathbf{x}_m = \mathbf{V}_m \mathbf{g}_m(\mathbf{b}_m), \tag{13.1}$$

where $\mathbf{b}_m \in \mathbb{B}^{\log_2(L)}$ denotes the encoded information bits of length $\log_2(L) \in \mathbb{C}$. Function $\mathbf{g}_m : \mathbb{B}^{\log_2(L)} \rightarrow \mathcal{X}_m$ generates an N_m-dimensional constellation $\mathcal{X}_m$ of size L for user m, which maps $\log_2(L)$ information bits to an N_m-dimensional complex

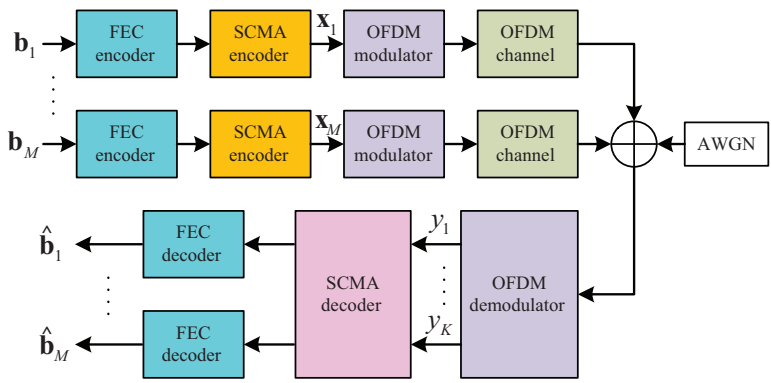


Figure 13.2 Block diagram of an uplink SCMA system

[†]In this chapter, we focus on the power domain NOMA [27] for the considered downlink communication scenario. Although the code-domain NOMA, such as SCMA, may outperform power-domain NOMA, SCMA is more suitable for the uplink communication where the reception complexity for information decoding is more affordable for base stations (BSs).

vector. Then the sparse mapping matrix $\mathbf{V}_m \in \mathbb{B}^{K \times N_m}$ maps the N_m-dimensional complex vector to K-dimensional sparse codewords $\mathbf{x}_m \in \mathbb{C}^K$. All $\mathbf{x}_m$ have the same positions of nonzero entries and constitute the codebook $\mathscr{C}_m$ for user m. Then, SCMA codewords of M users are multiplexed over K subcarriers.

At the receiver side, the received signal over all K subcarriers can be represented by

$$\mathbf{y} = \sum_{m=1}^{M} \mathrm{diag}(\mathbf{h}_m)\mathbf{x}_m + \mathbf{v}, \tag{13.2}$$

where $\mathbf{h}_m \in \mathbb{C}^K$ denotes the channel vector of user m on all subcarriers and $\mathbf{v}$ denotes the additive white Gaussian noise (AWGN) at the BS. Owing to the sparse property, there are only a few symbol collisions on each subcarrier. Therefore, similarly to LDS multiple access, some message passing algorithm-based MUD technique can be adopted to detect the multiplexed codewords with a moderate computational complexity.

13.2.3 Power-domain NOMA

The principle of power-domain NOMA is to exploit the users' power difference for multiuser multiplexing together with superposition coding (SC) at the transmitter, while applying successive interference cancelation (SIC) at the receivers for alleviating the inter-user interference (IUI) [9]. In fact, the concept of NOMA was originally proposed for downlink multiuser communications or so-called broadcast channels [14–16], where the two fundamental building blocks are SC and SIC. However, NOMA inherently exists in uplink communications, since the electromagnetic waves are naturally superimposed at a receiving BS and the implementation of SIC is more affordable at the BS than at user terminals. Therefore, the NOMA concept was recently generalized to uplink multiuser communications [47–50], i.e., multiple access channel, which will be detailed in the latter subsection. In this subsection, we introduce the concepts of SC and SIC in the context of downlink communications.

At the transmitter side, the coded signals intended for different users are superimposed with different power levels. Without loss of generality, the channel gains of users are assumed to be sorted with a particular ordering. In the literature [9,51,52], the user with a better channel quality is usually called a *strong user*, while the user with a worse channel quality is called a *weak user*. The transmit powers for the strong and weak users are allocated according to their channel gain order. To provide fairness and to facilitate the SIC decoding, the transmitter usually allocates more power to the weak user with a poor channel condition.[‡]

[‡]We note that allocating a higher power to the user with the worse channel is not necessarily required in NOMA, as shown in [53], especially when there is explicit minimum data rate requirement for each user.

At the receiver, SIC decoding is employed to exploit the heterogeneity in channel gains and transmit powers. For an SIC receiver, it first decodes other users' signals one by one based on a predefined decoding order before decoding its own signal. Usually, practical decoding order depends on the order of the received signal powers, where the user with a higher received power is decoded first during SIC decoding.[§] Upon finishing decoding one user's signal, the receiver subtracts it from its received signal. As a result, multiuser interference can be successively removed and the achievable data rate is improved. For downlink communications, users with better channel conditions can perform SIC to mitigate the IUI. Due to its advantages, SIC has been employed in practical systems such as CDMA [54] and vertical-bell laboratories layered space–time [55].

Compared to OMA schemes, SC combined with SIC can provide a comparable rate to the strong user, while achieving a performance close to the single-user upper bound for the weak user. Intuitively, the strong user, being at a high signal-to-noise ratio (SNR), is DoF limited and SC allows it to exploit the full DoF of the channel while being allocated only a small amount of transmit power, thus causing a small amount of interference to the weak user. In contrast, an orthogonal scheme has to allocate a significant fraction of the DoF to the weak user to achieve near single-user bound performance that causes a large degradation in the performance of the strong user. In fact, it has been proved that NOMA with SC and SIC is capable of achieving the capacities of general degraded broadcast channels [56].

In summary, the non-orthogonal feature can be introduced either in the power-domain only or in the hybrid code and power-domain. Although code-domain NOMA offers a the potential coding gain to improve spectral efficiency, power-domain NOMA is simpler since only minor changes are required in the physical layer operations at the transmitter side compared to current 4G technologies. In addition, power-domain NOMA paves the way for flexible resource allocation via relaxing the orthogonality requirement to improve the performance of NOMA, such as spectral efficiency [51,57,58], energy efficiency [59–62], and user fairness [37,50,63]. As a result, this chapter focuses on power-domain NOMA for illustration. In the following, the system models for uplink and downlink NOMA will be first presented and their multiuser capacity regions will be discussed then.

13.2.4 Downlink NOMA

In this subsection, to facilitate the presentation for the basic concepts of NOMA, we consider a simple single-carrier two-user downlink NOMA system. The generic system model for downlink NOMA is illustrated in Figure 13.3 with one BS and two users. The BS transmits the messages of both user 1 and user 2, i.e., s_1 and s_2, with different transmit powers, p_1 and p_2, on the same frequency band, respectively. The corresponding transmitted signal is represented by

$$x = \sqrt{p_1}s_1 + \sqrt{p_2}s_2, \tag{13.3}$$

[§]Note that the optimal SIC decoding order for uplink NOMA is different and will be detailed in Section 13.2.5.

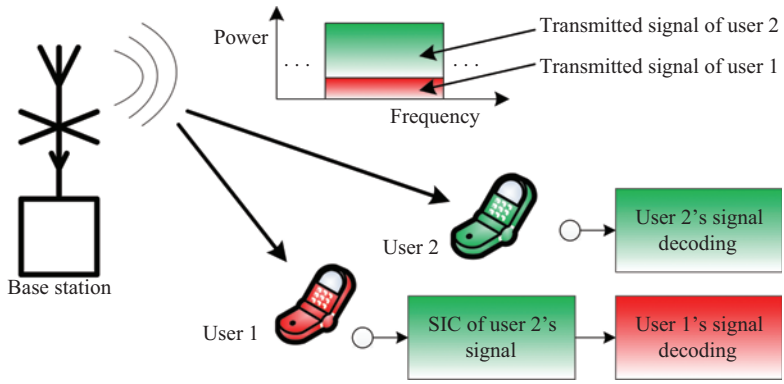


Figure 13.3 The downlink NOMA system model with one BS and two users

where transmit power is constrained by $p_1 + p_2 \leq p_{\max}$. The received signal at user k is given by

$$y_k = h_k x + n_k,\ k \in \{1, 2\}, \tag{13.4}$$

where $h_k \in \mathbb{C}$ denotes the channel coefficient between the BS and user k, including the joint effect of large- and small-scale fading. Variable n_k denotes the AWGN at user k with a noise power of σ^2, i.e., $n_k \sim \mathcal{CN}\left(0,\ \sigma^2\right)$. We assume that user 1 is the strong user with a better channel quality, while user 2 is the weak user with a worse channel quality, i.e., $|h_1|^2 \geq |h_2|^2$.

In downlink NOMA systems, the SIC decoding is implemented at the user side. The optimal SIC decoding order for maximizing the system sum-rate is with the descending order of channel gains normalized by noise [9]. It means that user 1 decodes s_2 first and removes the IUI of user 2 by subtracting s_2 from the received signal y_1 before decoding its own message s_1. On the other hand, user 2 does not perform interference cancelation and directly decodes its own message s_2 with interference from user 1. Since the considered downlink SISO-NOMA system model implies a degraded broadcast channel [56], we can easily obtain the following equations for characterizing the achievable rate region [51]:

$$R_{1,2} = \log_2\left(1 + \frac{p_2|h_1|^2}{p_1|h_1|^2 + \sigma^2}\right), \tag{13.5}$$

$$R_1 = \log_2\left(1 + \frac{p_1|h_1|^2}{\sigma^2}\right), \text{ and} \tag{13.6}$$

$$R_2 = \log_2\left(1 + \frac{p_2|h_2|^2}{p_1|h_2|^2 + \sigma^2}\right), \tag{13.7}$$

where $R_{1,2}$ denotes the achievable rate at user 1 to decode the message of user 2, R_1 denotes the achievable rate for user 1 to decode its own message after decoding and subtracting the signal of user 2, and R_2 denotes the achievable rate for user 2 to decode its own message. Note that the necessary condition for achieving the data rate R_1 in (13.6) is

$$R_{1,2} \geq R_2, \tag{13.8}$$

which ensures successful interference cancelation at user 1. Furthermore, we can observe that for any arbitrary power allocation p_1 and p_2, the condition in (13.8) can always be satisfied whenever $|h_1|^2 \geq |h_2|^2$. The system sum-rate for downlink NOMA can be obtained by

$$R_{\text{sum}}^{\text{DL}} = \log_2\left(1 + \frac{p_1|h_1|^2}{\sigma^2}\right) + \log_2\left(1 + \frac{p_2|h_2|^2}{p_1|h_2|^2 + \sigma^2}\right). \tag{13.9}$$

Note that SIC is not able to eliminate the interference caused by user 1 for user 2. Fortunately, if the power allocated to user 2 is significantly larger than that to user 1 in the aggregate received signal y_2, it does not introduce much performance degradation compared to allocating user 2 on this frequency band exclusively. The achievable rate

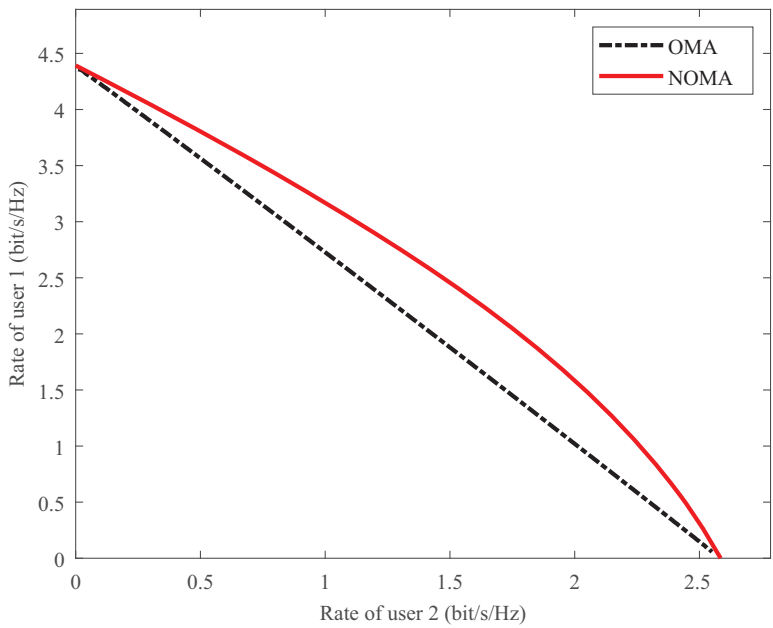


Figure 13.4 *The multiuser achievable rate region for downlink NOMA with one BS and two users. User 1 is the strong user with* $\frac{|h_1|^2}{\sigma^2} = 20$*, while user 2 is the weak user with* $\frac{|h_2|^2}{\sigma^2} = 5$

region of downlink NOMA is illustrated in Figure 13.4 in comparison with that of OMA. We can observe that the achievable rate region of OMA is only a subset of that of NOMA. As a result, NOMA provides a higher flexibility in resource allocation design for improving the system performance, especially considering the diverse QoS requirements of users.

13.2.5 Uplink NOMA

The generic system model for uplink NOMA is illustrated in Figure 13.5 with one BS and two users. Both users are transmitting their messages within the same frequency band with the same transmit power constraint, i.e., $p_1 \leq \frac{p_{\max}}{2}$ and $p_2 \leq \frac{p_{\max}}{2}$. The received signal at the BS is given by

$$y = \sqrt{p_1} h_1 s_1 + \sqrt{p_2} h_2 s_2 + n, \tag{13.10}$$

where $h_k \in \mathbb{C}$ denotes the channel coefficient between the BS and user k, s_k denotes the modulated symbol of user k, i.e., $k \in \{1, 2\}$, and $n \sim \mathcal{CN}(0, \sigma^2)$ denotes the AWGN at the BS. Without loss of generality, we assume that user 1 is the strong user and user 2 is the weak user, i.e.,$|h_1|^2 \geq |h_2|^2$. Due to their distinctive channel gains, the received signal power of user 1 is higher than that of the user 2, as shown in Figure 13.5.

In uplink NOMA, SIC decoding is performed at the BS to retrieve the messages s_1 and s_2 from the superimposed signal y. In particular, the BS first decodes the message of user 1 and then subtracts s_1 from the superimposed signal y. Then, the BS can decode

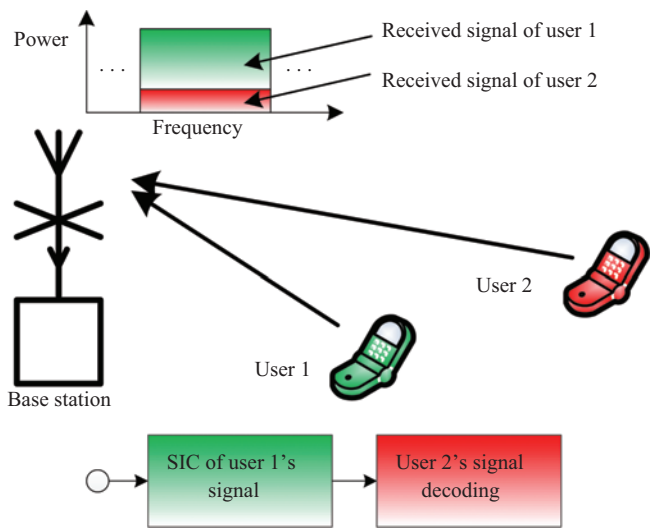


Figure 13.5 The uplink NOMA system model with one BS and two users

the message of user 2 without IUI. As a result, we can obtain the individual data rates for user 1 and user 2 as follows:

$$R_1 = \log_2\left(1 + \frac{p_1 |h_1|^2}{p_2|h_2|^2 + \sigma^2}\right) \text{ and} \tag{13.11}$$

$$R_2 = \log_2\left(1 + \frac{p_2 |h_2|^2}{\sigma^2}\right). \tag{13.12}$$

The system sum-rate for downlink NOMA can be obtained by

$$R_{\text{sum}}^{\text{UL}} = \log_2\left(1 + \frac{p_1|h_1|^2 + p_2|h_2|^2}{\sigma^2}\right). \tag{13.13}$$

From (13.13), we can observe that the system sum-rate is actually independent of the SIC decoding order. Different SIC decoding orders only change the individual data rates of both users in (13.11) and (13.12) without affecting the total system sum-rate, respectively. The achievable rate region of uplink NOMA is illustrated in Figure 13.6 in comparison with that of OMA. Similar to downlink NOMA, the achievable rate region of OMA is also a subset of that of NOMA, which provides a higher flexibility in resource allocation design.

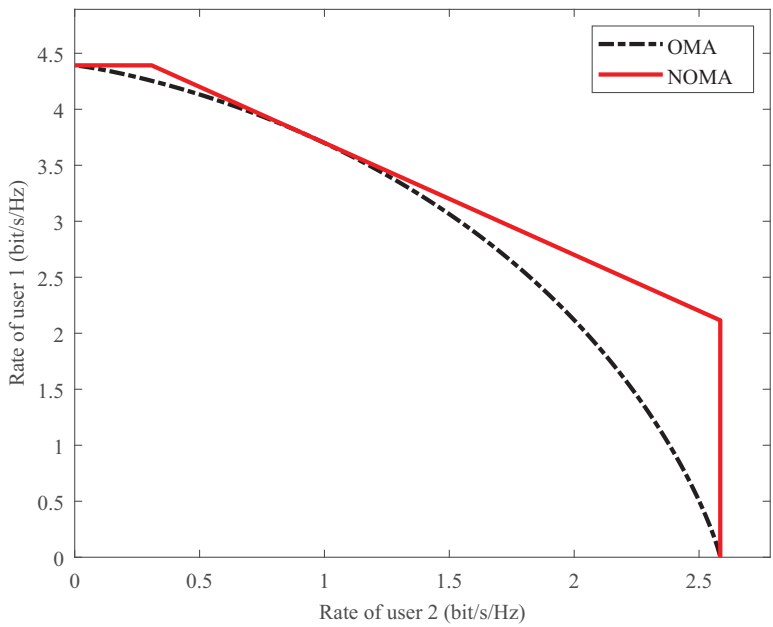


Figure 13.6 *The multiuser achievable rate region for uplink NOMA with one BS and two users. User 1 is the strong user with* $\frac{|h_1|^2}{\sigma^2} = 20$*, while user 2 is the weak user with* $\frac{|h_2|^2}{\sigma^2} = 5$

13.3 Energy efficiency of NOMA

Energy efficiency has emerged as a new prominent and fundamental figure of merit for wireless communication systems as the energy consumptions and related environment problems become a major issue currently. Based on the achievable sum-rate defined in the last section, we introduce the basic definition energy efficiency of NOMA adopted in this chapter. Then, we discuss the fundamental trade-off between the spectral efficiency and energy efficiency for NOMA systems.

13.3.1 Energy efficiency of downlink NOMA

In general, the energy efficiency is essentially in the form of a benefit–cost ratio[‖] to evaluate the amount of data delivered by utilizing the limited system energy resource (bits/J). Accordingly, the energy efficiency of downlink NOMA systems is defined as [30,31,64]

$$\begin{aligned} \mathrm{EE}^{\mathrm{DL}} &= \frac{W \cdot R_{\mathrm{sum}}^{\mathrm{DL}}}{\delta(p_1 + p_2) + P_{\mathrm{C}}} \\ &= \frac{W \cdot \left[\log_2\left(1 + \frac{p_1|h_1|^2}{\sigma^2}\right) + \log_2\left(1 + \frac{p_2|h_2|^2}{p_1|h_2|^2 + \sigma^2}\right)\right]}{\delta(p_1 + p_2) + P_{\mathrm{C}}}, \end{aligned} \tag{13.14}$$

where P_{C} denotes the static circuit power consumption associated with communications at both transceiver of downlink NOMA systems. Constant $\delta > 1$ captures the inefficiency of the transmit power amplifier. We assume that the power amplifier operates in its linear region and the hardware power consumption P_{C} is a constant. Here, we note that σ^2 denotes the effective noise power in the considered frequency band W.

Similarly, the energy efficiency of uplink NOMA systems is defined as

$$\begin{aligned} \mathrm{EE}^{\mathrm{UL}} &= \frac{W \cdot R_{\mathrm{sum}}^{\mathrm{UL}}}{\delta(p_1 + p_2) + P_{\mathrm{C}}} \\ &= \frac{W \cdot \log_2\left(1 + \frac{p_1|h_1|^2 + p_2|h_2|^2}{\sigma^2}\right)}{\delta(p_1 + p_2) + P_{\mathrm{C}}}. \end{aligned} \tag{13.15}$$

13.3.2 The trade-off between energy efficiency and spectral efficiency

To illustrate the fundamental trade-off between energy efficiency and spectral efficiency, we impose the same power allocation for both users in both downlink

[‖] We note that there are various types of energy efficiency definitions such as from facility level, equipment level, and network level, respectively [65], depending on the design of particular systems.

and uplink NOMA systems, i.e., $p_1 = p_2 = \frac{p_{\max}}{2}$. Then, the energy efficiencies of downlink NOMA and uplink NOMA are given by

$$\mathrm{EE}^{\mathrm{DL}} = \frac{W \cdot \left[\log_2\left(1 + \frac{p_{\max}|h_1|^2}{2\sigma^2}\right) + \log_2\left(1 + \frac{p_{\max}|h_2|^2}{p_{\max}|h_2|^2 + 2\sigma^2}\right)\right]}{\delta p_{\max} + P_{\mathrm{C}}} \text{ and} \tag{13.16}$$

$$\mathrm{EE}^{\mathrm{UL}} = \frac{W \cdot \log_2\left(1 + \frac{p_{\max}(|h_1|^2 + |h_2|^2)}{2\sigma^2}\right)}{\delta p_{\max} + P_{\mathrm{C}}}, \tag{13.17}$$

respectively. Furthermore, considering an OMA system, such as a TDMA system, we equally allocate the DoF to both users, and the system energy efficiency is given by

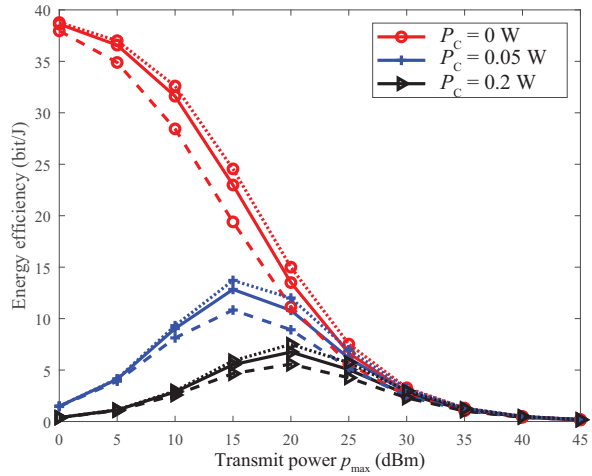
$$\mathrm{EE}^{\mathrm{OMA}} = \frac{W \cdot \left[\frac{1}{2}\log_2\left(1 + \frac{p_{\max}|h_1|^2}{\sigma^2}\right) + \frac{1}{2}\log_2\left(1 + \frac{p_{\max}|h_2|^2}{\sigma^2}\right)\right]}{\delta p_{\max} + P_{\mathrm{C}}}, \tag{13.18}$$

where both users only have half DoF for data transmission but enjoy the total transmit power and an interference-free link.

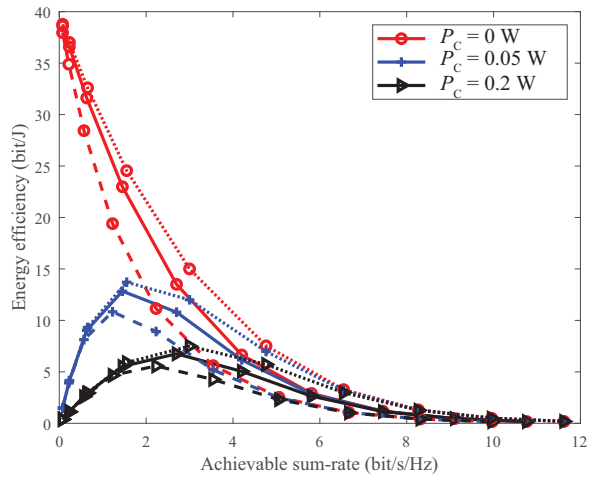
We can observe that the achievable sum-rate of uplink NOMA systems in the numerator of (13.17) increases logarithmically with the maximum transmit power $p_{\max}$, while its total power consumption in the denominator of (13.17) is an affine increasing function of $p_{\max}$. As a result, there is a nontrivial trade-off between the energy efficiency and achievable sum-rate that plays an important role in energy-efficient resource allocation design in uplink NOMA systems. More importantly, the trade-off between the energy efficiency and the achievable sum-rate is similar to that of the OMA case in (13.18), since (13.17) can be viewed as the data rate of a single user with a transmit power of $p_{\max}$ and a channel gain of $|h_1|^2 + |h_2|^2$. On the other hand, for downlink NOMA systems, it can be observed that the achievable sum-rate in the numerator of (13.16) also increases monotonically with $p_{\max}$ but becomes interference-limited when keeping increasing $p_{\max}$. As a result, the trade-off between the energy efficiency and the achievable sum-rate for downlink NOMA might be different from that of uplink NOMA.

Figure 13.7 illustrates the trade-off between the energy efficiency, transmit power, and achievable sum-rate for both OMA and NOMA systems. We can observe that all the curves of the three schemes have the same trend in both parts (a) and (b) of Figure 13.7. In particular, all the three schemes have the same performance in both low and high SNR regimes. For moderate transmit power, downlink NOMA offers a better trade-off than that of uplink NOMA, while uplink NOMA outperforms OMA schemes. When the total circuit power consumption is negligibly small, i.e., $P_{\mathrm{C}} = 0\,\mathrm{W}$, the system energy efficiency is monotonically decreasing w.r.t. both the total transmit power and the achievable sum-rate, as shown in Figure 13.7(a) and (b), respectively. In other words, transmission with an arbitrarily small power, i.e., $p_{\max} \to 0$, is the optimal operation point for maximizing the system energy efficiency and the obtained system energy efficiency is

$$\lim_{p_{\max} \to 0, P_{\mathrm{C}} = 0} \mathrm{EE}^{\mathrm{OMA}} = \mathrm{EE}^{\mathrm{DL}} = \mathrm{EE}^{\mathrm{UL}} = \frac{|h_1|^2 + |h_2|^2}{2\delta\sigma^2}. \tag{13.19}$$



(a)



(b)

Figure 13.7 An illustration of the trade-off between energy efficiency (bit/J) versus (a) transmit power (p_{max} (dBm)) and (b) achievable sum-rate (bit/s/Hz). The simulation setups are $W = 1$ Hz, $\frac{|h_1|^2}{\sigma^2} = 100$, $\frac{|h_2|^2}{\sigma^2} = 10$, $P_C = [0, 0.05, 0.2]$ W, $p_{max} = [0{:}5{:}45]$ dBm, and $\delta = 2$. The solid lines denote the performance of uplink NOMA systems, the dotted lines denote the performance of downlink NOMA systems, and the dash lines denote the performance of OMA systems

Furthermore, when $P_C > 0$, the system energy efficiency first increases with p_{max} and then decreases with p_{max}. In particular, when p_{max} is small, the system energy efficiency is mainly limited by the fixed circuit power consumption P_C, and the system achievable sum-rate scales almost linearly w.r.t. the total transmit power p_{max} in the low SNR regime. As a result, increasing the transmit power can effectively increase both the system achievable sum-rate and the system energy efficiency. In contrast, the total power consumption is dominated by the transmit power p_{max} in the high SNR regime, while there is only a marginal gain in system achievable sum-rate with increasing p_{max}. Therefore, after reaching the maximum system energy efficiency, further increasing the transmit power p_{max} would decrease the system energy efficiency rapidly as shown in Figure 13.7(a). Additionally, we can observe in Figure 13.7(a) that the optimal operation point is pushed toward the high SNR regime with increasing the circuit power consumption P_C. This is due to the fact that the larger the circuit power consumption, the higher transmit power is required to outweigh the impact of the circuit power consumption on the system energy efficiency. For a practical scenario of $P_C > 0$, finding the optimal operation point via resource allocation design to maximize the system energy efficiency has attracted significant attention in the literature in the past few years [32,33]. In the following, the energy-efficient resource allocation design for NOMA systems is formulated, and a generic solution is presented then.

13.4 Energy-efficient resource allocation design

In wireless communications, resource allocation is the concept of making the best use of limited communication resources based on the information available at the resource allocator to improve the system performance, as shown in Figure 13.8. Specifically, resource allocation designs rely on the application of the optimization theory to optimize the system performance with taking into account various QoS constraints. The available information at the resource allocator usually includes the channel state information (CSI), the transmit power budget, and the available bandwidth as well as time resource. The design objectives are usually maximizing the

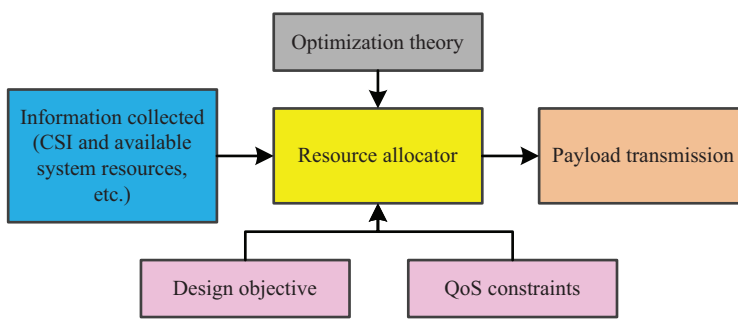


Figure 13.8 Optimization-based resource allocation design framework

system sum-rate [66–68], maximizing the system energy efficiency [26,27,30–34,64] or minimizing the system power consumption [12,69–74], etc. In this chapter, we focus on the problem formulation and solution for energy-efficient resource allocation design whose objective function is the system energy efficiency. The QoS requirements, such as the minimum data rate requirement and outage probability requirement, act as constraints in resource allocation optimization framework. Finally, based on the obtained resource allocation strategy, including power allocation, user scheduling, rate allocation, etc., the transmitter would transmit the payload accordingly. As shown in the literature [12,26,27,30–34,64,66–74], optimization theories, including convex [75] and non-convex optimization [76], play an important role for resource allocation design. Interested readers are referred to [75,76] for more details.

13.4.1 Design objectives

Considering a communication system with K users, the objective function (utility function) for maximizing the system energy efficiency is given by

$$\mathrm{EE}(\mathbf{x}) = \frac{\sum_{k=1}^{K} R_k(\mathbf{x})}{\sum_{k=1}^{K} (\delta p_k(\mathbf{x}) + P_{\mathrm{C},k})}, \tag{13.20}$$

where $\mathbf{x}$ denotes the optimization variable, and it may include the variables for power allocation, user scheduling, rate allocation, etc. The individual rate of user k, $R_k(\mathbf{x})$, is a function of the optimization variables. In the denominator, $P_{\mathrm{C},k}$ denotes the static circuit power consumption associated with user k. Additionally, $p_k(\mathbf{x})$ represents the transmit power for user k that is a function of other optimization variables, such as power allocation for other users, user scheduling, rate allocation, etc.

13.4.2 QoS constraint

To satisfy diverse QoS requirements for different applications, different types of QoS constraints can be incorporated in the problem formulation of resource allocation designs. In fact, QoS constraints combined with the system resource limitations span the feasible solution set $\mathscr{X}$ for the optimization variable, i.e., $\mathbf{x} \in \mathscr{X}$. In general, two kinds of QoS constraints commonly adopted in the literature are introduced in the following.

13.4.2.1 Minimum data rate requirement

As the name implies, the minimum data rate requirement is a constant minimum data rate $R_k^{\min}$ to support the application of communication user k, i.e.:

$$R_k(\mathbf{x}) \geq R_k^{\min}, \tag{13.21}$$

where $R_k^{\min}$ is usually predefined and obtained during the information collection phase in Figure 13.8. The minimum data rate requirement is imposed for resource allocation design to guarantee the QoS of each user [12]. In particular, when optimizing the system performance, some users with poor channel conditions may suffer from starvation since allocating resource to the weak user does not contribute so much gain

to the system performance. Hence, introducing a minimum data rate requirement can effectively balance the system performance and each user's QoS requirement.

13.4.2.2 Outage probability requirement

For a communication system with imperfect CSI at the transmitter side, there exists nonzero probability that the scheduled data rate exceeds the instantaneous channel capacity [77]. In this case, even applying powerful error correction coding cannot prevent packet error and thus an outage occurs. As a result, the outage probability for the communication link of user k can be defined as

$$\mathrm{P}_{\mathrm{out},k}(\mathbf{x}) = \Pr\{C_k(\mathbf{x}) < R_k(\mathbf{x})\}, \tag{13.22}$$

where $R_k(\mathbf{x})$ is the allocated data rate for user k and $C_k(\mathbf{x})$ is the channel capacity of the communication link for user k. They both depend on the resource allocation policy and thus the outage probability is a function of the resource allocation variables $\mathbf{x}$. The outage probability constraint has been employed for resource allocation design to enhance the communication reliability [12,78]. In particular, the outage probability of user k should be smaller than the maximum tolerable outage probability $\overline{\mathrm{P}}_{\mathrm{out},k}$, i.e.:

$$\mathrm{P}_{\mathrm{out},k}(\mathbf{x}) \leq \overline{\mathrm{P}}_{\mathrm{out},k}. \tag{13.23}$$

This probabilistic constraint takes the CSI imperfectness into consideration and hence is very useful for robust resource allocation in wireless communications. We note that the robust resource allocation design based on outage probability only needs to know the statistical CSI at the transmitter, rather than the instantaneous CSI. This makes the outage-constrained resource allocation design more practical since statistical CSI is usually available based on the long-term measurements and does not change so fast as the instantaneous CSI.

13.4.3 Fractional programming

The system-centric energy-efficient resource allocation design can be formulated as the following optimization problem:

A generic problem formulation for energy-efficient resource allocation design

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathrm{EE}(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{s.t.} \quad & \mathbf{x} \in \mathscr{X}, \end{aligned} \tag{13.24}$$

where $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{C}^n$ is the optimization variables and $\mathscr{X} \subseteq \mathbb{C}^n$ is the feasible solution set spanned by the system resource limitations and the QoS constraints. The numerator $f(\mathbf{x}) : \mathscr{X} \subseteq \mathbb{C}^n \to \mathbb{R}$ denotes the system data-rate produced by the resource allocation strategy $\mathbf{x}$. The denominator $g(\mathbf{x}) : \mathscr{X} \subseteq \mathbb{C}^n \to \mathbb{R}$ denotes the total system power consumption.

The formulated problem in (13.24) can be classified as fractional programming [79]. Without loss of generality, we define the maximum energy efficiency of the problem in (13.24) as follows:

$$\mathrm{EE}^* = \underset{\mathbf{x}}{\text{maximize}} \ \frac{f(\mathbf{x})}{g(\mathbf{x})}, \ \text{s.t.} \ \ \mathbf{x} \in \mathscr{X}. \tag{13.25}$$

Now, the following theorem can transform the fractional objective function in (13.24) to an equivalently subtractive form.

Theorem 1. Let $\mathbf{x}^*$ be a feasible solution of the problem in (13.24) and (13.26). The maximum energy efficiency EE^* is achieved if and only if:

$$\underset{\mathbf{x}\in\mathscr{X}}{\text{maximize}} \ f(\mathbf{x}) - \mathrm{EE}^* g(\mathbf{x}) = f(\mathbf{x}^*) - \mathrm{EE}^* g(\mathbf{x}^*) = 0, \tag{13.26}$$

for $f(\mathbf{x}) \geq 0$ and $g(\mathbf{x}) > 0$.

Proof. The proof can be found in Appendix A.1. □

In the literature, Dinkelbach's method has been proposed to find EE^* iteratively [79], as shown in Algorithm 13.1. In particular, in each iteration of the main loop, one needs to solve (13.26) for a given temporary $\mathrm{EE}^{\mathrm{iter}}$. The convergence of Dinkelbach's algorithm is stated in the following theorem.

Theorem 2. Dinkelbach's algorithm converges to the globally optimal solution of the problem in (13.24) if the problem in (13.26) can be solved optimally for a given EE.

Proof. The proof can be found in Appendix A.2. □

We note that Theorem 2 only requires that the problem in (13.26) can be solved but it does not impose any assumption on the convexity or concavity of the function $f(\mathbf{x})$, $g(\mathbf{x})$, and the feasible solution set $\mathscr{X}$. Dinkelbach's algorithm can be implemented with a lower computational complexity when $f(\mathbf{x})$ is a concave function, $g(\mathbf{x})$ is a convex function, and $\mathscr{X}$ is a compact convex set. Generally, the energy-efficient

Algorithm 13.1: Dinkelbach's algorithm

1: **Initialization**
Initialize the convergence tolerance $\varepsilon \to 0$, the maximum number of iterations $\mathrm{iter}_{\max}$, the iteration index $\mathrm{iter} = 1$ and the initial system energy efficiency $\mathrm{EE}^{\mathrm{iter}} = 0$
2: **repeat** {**Main loop**}
3: Solve (13.26) for the given temporary $\mathrm{EE}^{\mathrm{iter}}$ to obtain the resource allocation strategy $\mathbf{x}^{\mathrm{iter}}$
4: **if** $f\left(\mathbf{x}^{\mathrm{iter}}\right) - \mathrm{EE}^{\mathrm{iter}} g\left(\mathbf{x}^{\mathrm{iter}}\right) < \varepsilon$ **then**
5: Convergence = **true**
6: **return** $\mathbf{x}^* = \mathbf{x}^{\mathrm{iter}}$ and $\mathrm{EE}^* = \frac{f(\mathbf{x}^{\mathrm{iter}})}{g(\mathbf{x}^{\mathrm{iter}})}$
7: **else**
8: Set $\mathrm{EE}^{\mathrm{iter}+1} = \frac{f(\mathbf{x}^{\mathrm{iter}})}{g(\mathbf{x}^{\mathrm{iter}})}$ and $\mathrm{iter} = \mathrm{iter} + 1$
9: Convergence = **false**
10: **end if**
11: **until** Convergence = **true** or $\mathrm{iter} = \mathrm{iter}_{\max}$

resource allocation for uplink NOMA systems follows the problem formulation in (13.24) with a concave $f(\mathbf{x})$ and a convex $g(\mathbf{x})$ over a compact convex set $\mathscr{X}$, which can be efficiently solved by the Dinkelbach method. Unfortunately, the achievable sum-rate, $f(\mathbf{x})$, for downlink NOMA systems is usually non-convex due to the existence of IUI, as shown in (13.10). In this case, there is no systematic and computationally efficient approach to solve (13.26) globally. As a compromise approach, one should first convexify the associated function before applying the Dinkelbach method. In the following subsection, the SCA technique [80,81] can be utilized to convexify $f(\mathbf{x})$ iteratively in the original problem formulation in (13.24).

13.4.4 Successive convex approximation

As mentioned in Theorem 2, the Dinkelbach method can efficiently handle the fractional form objective function in (13.24) with a concave function over a convex function. When the non-convexity arises only from $f(\mathbf{x})$ in the numerator of the objective function in (13.24), before applying Dinkelbach's method in Algorithm 13.1, we perform an SCA to convexify $f(\mathbf{x})$ [80,81] iteratively, which forms an outer loop of the overall algorithm. To strike a balance between system performance and computational complexity, we propose a suboptimal iterative approach based on the SCA technique [82,83]. The method has a polynomial time computational complexity and usually achieves excellent performance. In particular, the achievable sum-rate $f(\mathbf{x})$ can be usually rewritten as the difference of two concave functions as follows:

$$f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}), \tag{13.27}$$

where $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are differentiable concave functions w.r.t. $\mathbf{x}$. For a given feasible solution $\mathbf{x}^{\text{outer}}$ in the outer loop, we can obtain its lower bound as follows:

$$f(\mathbf{x}) \geq f^{\text{outer}}(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}^{\text{outer}}) - \nabla_{\mathbf{x}} f_2(\mathbf{x}^{\text{outer}})^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^{\text{outer}}), \tag{13.28}$$

where $\nabla_{\mathbf{x}} f_2(\mathbf{x}^{\text{outer}}) = [\frac{\partial f_2(\mathbf{x})}{\partial x_1}|_{\mathbf{x}^{\text{outer}}}, \frac{\partial f_2(\mathbf{x})}{\partial x_2}|_{\mathbf{x}^{\text{outer}}}, \ldots, \frac{\partial f_2(\mathbf{x})}{\partial x_n}|_{\mathbf{p}^{\text{outer}}}]^{\mathrm{T}} \in \mathbb{C}^{n\times 1}$ denotes the gradient of $f_2(\cdot)$ w.r.t. $\mathbf{x}$.

Now, given a feasible solution $\mathbf{x}^{\text{outer}}$, a lower bound for the problem in (13.24) can be obtained by solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \text{EE}(\mathbf{x}) = \frac{f^{\text{outer}}(\mathbf{x})}{g(\mathbf{x})} \\ \text{s.t.} \quad & \mathbf{x} \in \mathscr{X}. \end{aligned} \tag{13.29}$$

To tighten the obtained lower bound, we employ an iterative outer loop to generate a sequence of feasible solution $\mathbf{x}^{\text{outer}}$ successively, as shown in Algorithm 13.2. In particular, the solution of (13.29) is denoted as $\mathbf{x}^{\text{outer}+1}$ and is further used for updating the problem (13.29). The convergence of the outer loop can be easily proved. In the outer-th iteration of outer loop, we can observe the following relationships:

$$\frac{f(\mathbf{x}^{\text{outer}})}{g(\mathbf{x}^{\text{outer}})} \overset{(a)}{=} \frac{f^{\text{outer}}(\mathbf{x}^{\text{outer}})}{g(\mathbf{x}^{\text{outer}})} \overset{(b)}{\leq} \frac{f^{\text{outer}}(\mathbf{x}^{\text{outer}+1})}{g(\mathbf{x}^{\text{outer}+1})} \overset{(c)}{\leq} \frac{f(\mathbf{x}^{\text{outer}+1})}{g(\mathbf{x}^{\text{outer}+1})}. \tag{13.30}$$

The equality in (a) can be easily obtained by substituting $\mathbf{x}^{\text{outer}}$ into (13.27) and (13.28). The inequality in (b) is because $\mathbf{x}^{\text{outer}+1}$ is the optimal solution for the problem in (13.29) and the inequality in (c) follows the system sum-rate lower bound in (13.28). With a compact feasible solution set and a continuous objective function, the outer loop can converge to a stationary suboptimal solution.

To solve the problem in (13.29) in each iteration of outer loop, an inner loop following the Dinkelbach algorithm in Algorithm 13.1 is employed. In each iteration of the inner loop (line 5 in Algorithm 13.2), we solve the following optimization problem for the given EE^{inner} and $\mathbf{x}^{\text{outer}}$:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & f^{\text{outer}}(\mathbf{x}) - \text{EE}^{\text{inner}} g(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathscr{X}. \end{aligned} \tag{13.31}$$

Note that the problem in (13.31) is a convex optimization, which can be solved efficiently by standard convex problem solvers, such as CVX [84]. The solution of (13.31) is denoted as $\widetilde{\mathbf{x}}^{\text{inner}}$. The detailed proof of convergence for the inner loop can be found in Appendix A.2. After the convergence of the inner loop, the feasible solution of the outer loop is updated on the basis of the output of the inner loop, i.e., $\mathbf{x}^{\text{outer}+1} = \widetilde{\mathbf{x}}^{\text{inner}}$, and the corresponding system energy efficiency is stored in $\text{EE}^{\text{outer}+1} = \frac{f^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}{g\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}$. The outer loop stops when the improvement of the system energy efficiency is smaller than a predefined threshold, as shown in line 15 of Algorithm 13.2. After the convergence of the outer loop, the obtained resource

Algorithm 13.2: Energy-efficient resource allocation algorithm

1: **Initialization**
Initialize the convergence tolerance ε, the maximum number of iterations for outer loop $\text{outer}_{\max}$, the maximum number of iterations for inner loop $\text{inner}_{\max}$, the initial iteration index $\text{outer} = \text{inner} = 1$, the initial feasible solution $\mathbf{x}^{\text{outer}}$, and the initial energy efficiency $\text{EE}^{\text{outer}} = 0$.
2: **repeat** {**Outer loop**}
3: Initialize the energy efficiency $\widetilde{\text{EE}}^{\text{inner}} = 0$.
4: **repeat** {**Inner loop**}
5: Solve (13.31) for a given $\mathbf{x}^{\text{outer}}$ and $\widetilde{\text{EE}}^{\text{inner}}$ to obtain the resource allocation solution $\widetilde{\mathbf{x}}^{\text{inner}}$.
6: **if** $f^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right) - \widetilde{\text{EE}}^{\text{inner}} g\left(\widetilde{\mathbf{x}}^{\text{inner}}\right) < \varepsilon$ **then**
7: Convergence = **true**
8: **update** $\mathbf{x}^{\text{outer}+1} = \widetilde{\mathbf{x}}^{\text{inner}}$ and $\text{EE}^{\text{outer}+1} = \frac{f^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}{g\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}$
9: Set outer = outer + 1
10: **else**
11: Convergence = **false**
12: Set $\widetilde{\text{EE}}^{\text{inner}+1} = \frac{f^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}{g\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)}$ and inner = inner + 1
13: **end if**
14: **until** Convergence = **true** or inner = $\text{inner}_{\max}$
15: **until** outer = $\text{outer}_{\max}$ or $\frac{\left|\text{EE}^{\text{outer}} - \text{EE}^{\text{outer}+1}\right|}{\text{EE}^{\text{outer}}} \leq \varepsilon$.
16: Return the resource allocation solution $\mathbf{x}^* = \mathbf{x}^{\text{outer}}$ and the energy efficiency $\text{EE}^* = \text{EE}^{\text{outer}}$

allocation solution and the corresponding system energy efficiency are returned, i.e., $\mathbf{x}^* = \mathbf{x}^{\text{outer}}$ and $\text{EE}^* = \text{EE}^{\text{outer}}$.

13.5 An illustrative example: energy-efficient design for multicarrier NOMA

In this section, we provide an illustrative example to demonstrate the energy-efficient resource allocation design for a downlink MC-NOMA system.

13.5.1 System model

We consider a single-cell downlink NOMA network with a cell radius of D meters, as shown in Figure 13.9. The BS is located at the cell center and is serving K mobile users randomly distributed within the cell. We consider a multicarrier system with N_{F} equally spacing subcarriers. We assume that $N_{\text{F}} \geq K$ such that both OFDMA and NOMA schemes can be utilized to multiplex multiple users, and we can compare their system energy efficiency. All transceivers are equipped with a single antenna. We assume that the CSI is perfectly known at the BS for energy-efficient resource allocation design.

Let the binary variable $u_{k,i}$ denote the user scheduling variable. In particular, $u_{k,i} = 1$ if subcarrier i is assigned to user k. Otherwise, $u_{k,i} = 0$. The transmit power allocated for user k on subcarrier $i \in \{1, \ldots, N_{\text{F}}\}$ is denoted as $p_{k,i} > 0$. For simplicity, we consider a normalized subcarrier spacing, i.e., $W_i = 1$ Hz, $\forall i$. The transmitted signal from the BS to all the users on subcarrier i is given by

$$x_i = \sum_{k=1}^{K} u_{k,i} \sqrt{p_{k,i}} s_{k,i}, \tag{13.32}$$

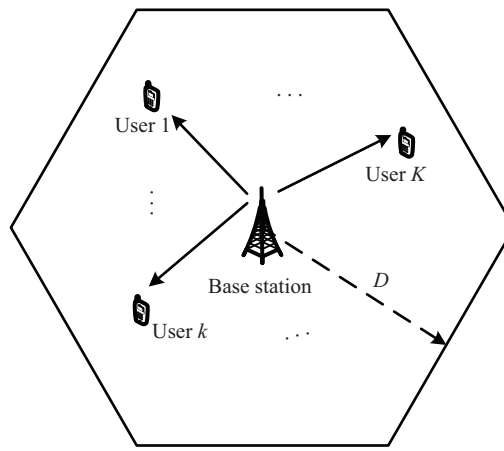


Figure 13.9 A single-cell downlink multicarrier NOMA system with a BS serving K users simultaneously

where $s_{k,i} \in \mathbb{C}$ denotes the modulated symbol for user k on subcarrier i. At the multicarrier demodulator output of user k, the received signal from the BS on subcarrier i is given by

$$y_{k,i} = h_{k,i} \sum_{k=1}^{K} u_{k,i} \sqrt{p_{k,i}} s_{k,i} + z_{k,i}, \tag{13.33}$$

where $z_{k,i} \sim \mathcal{CN}(0, \sigma^2)$ denotes the AWGN in subcarrier i at user k with a noise power of σ^2. The variable $h_{k,i} \in \mathbb{C}$ represents the frequency-domain channel coefficient between the BS and user k on subcarrier i. Without loss of generality, we assume $|h_{1,i}| > |h_{2,i}| > \cdots > |h_{K,i}|$. Note that on a different subcarrier, all the users may have different channel order. But this assumption can be easily maintained via a simple mapping relationship.

On each subcarrier, we assume:

$$\sum_{k=1}^{K} u_{k,i} \leq L, \forall i, \tag{13.34}$$

where L can be set such that an acceptable SIC decoding complexity and delay occurs at user side.¶ Furthermore, due to the SIC decoding, user k only faces the IUI from users $1, 2, \ldots, k-1$ on subcarrier i. Then, the achievable rate of user k on subcarrier i is obtained by

$$R_{k,i} = u_{k,i} \log_2 \left(1 + \frac{p_{k,i} |h_{k,i}|^2}{\sum_{k'=1}^{k-1} u_{k',i} p_{k',i} |h_{k,i}|^2 + \sigma^2} \right). \tag{13.35}$$

The individual data-rate of user k is given by

$$R_k(\mathbf{p}, \mathbf{u}) = \sum_{i=1}^{N_F} u_{k,i} \log_2 \left(1 + \frac{p_{k,i} |h_{k,i}|^2}{\sum_{k'=1}^{k-1} u_{k',i} p_{k',i} |h_{k,i}|^2 + \sigma^2} \right), \tag{13.36}$$

and the system sum-rate is given by

$$R_{\text{sum}}(\mathbf{p}, \mathbf{u}) = \sum_{k=1}^{K} \sum_{i=1}^{N_F} u_{k,i} \log_2 \left(1 + \frac{p_{k,i} |h_{k,i}|^2}{\sum_{k'=1}^{k-1} u_{k',i} p_{k',i} |h_{k,i}|^2 + \sigma^2} \right), \tag{13.37}$$

where $\mathbf{p} \in \mathbb{R}^{KN_F \times 1}$ and $\mathbf{u} \in \mathbb{B}^{KN_F \times 1}$ collect all the power allocation variables $p_{k,i}$ and user scheduling variables $u_{k,i}$, respectively.

The total system power consumption mainly consists of the transmit power and circuit power and it can be modeled by

$$U_p(\mathbf{p}, \mathbf{u}) = \delta \sum_{k=1}^{K} \sum_{i=1}^{N_F} u_{k,i} p_{k,i} + P_C, \tag{13.38}$$

¶Although optimizing L can further improve the system energy efficiency, it is beyond the scope of this chapter.

where P_{C} is the associated circuit power consumption at all the transceivers.** In addition, the total transmit power should satisfy

$$\sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}p_{k,i} \leq p_{\max}, \tag{13.39}$$

where $p_{\max}$ denotes system transmit power budget. Now, the system energy efficiency can be defined as follows:

$$\mathrm{EE}(\mathbf{p},\mathbf{u}) = \frac{R_{\mathrm{sum}}(\mathbf{p},\mathbf{u})}{U_{\mathrm{p}}(\mathbf{p},\mathbf{u})}. \tag{13.40}$$

13.5.2 Energy-efficient resource allocation design

To maximize the system energy efficiency of the considered MC-NOMA system, we formulate the energy-efficient resource allocation design as the following optimization problem:

$$\begin{aligned}
\underset{\mathbf{p},\mathbf{u}}{\text{maximize}} \quad & \mathrm{EE}(\mathbf{p},\mathbf{u}) = \frac{R_{\mathrm{sum}}(\mathbf{p},\mathbf{u})}{U_{\mathrm{p}}(\mathbf{p},\mathbf{u})} \\
\text{s.t.} \quad & \mathrm{C1}: u_{k,i} \in \{0,1\}, \forall k,i, \\
& \mathrm{C2}: p_{k,i} \geq 0, \forall k,i, \\
& \mathrm{C3}: \sum_{k=1}^{K} u_{k,i} \leq L, \forall i, \\
& \mathrm{C4}: \sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}p_{k,i} \leq p_{\max}, \\
& \mathrm{C5}: R_k(\mathbf{p},\mathbf{u}) \geq R_{\min}, \forall k,
\end{aligned} \tag{13.41}$$

where C1 and C2 in (13.41) are the definitions of $u_{k,i}$ and $p_{k,i}$, respectively. Constraints C3 and C4 in (13.41) follow (13.34) and (13.39), respectively. Constraint C5 in (13.41) imposes a minimum data rate requirement for each user.

The formulated problem is a mixed-integer non-convex optimization problem, which in general cannot be solved with a systematic and computationally efficient approach optimally. In particular, the non-convexity arises from the binary constraint C1 on the user scheduling variables and the coupling between the binary variables $u_{k,i}$ and continuous variables $p_{k,i}$. In addition, the remained IUI in the rate function in

**Here, we focus on the system energy efficiency maximization at the BS for the considered downlink communication system where the power consumption of receiver is negligible compared to that consumed by the BS and thus is not taken into consideration.

the numerator of the objective function is also an obstacle in designing the resource allocation efficiently. We note that the fractional form in the objective function suggests the user of the Dinkelbach method to handle this problem. However, as the rate expression is a non-convex function, following the idea of Algorithm 13.2, SCA is first applied to approximate it by a convex function successively.

To start with, we aim to handle the coupling between binary user scheduling variables $u_{k,i}$ and continuous power allocation variables $p_{k,i}$. We introduce an auxiliary variable $t_{k,i} = u_{k,i}p_{k,i}$ and adopt the big-M formulation [85] to transform the problem to its equivalent optimization problem as follows:

$$\begin{aligned}
\underset{\mathbf{p},\mathbf{t},\mathbf{u}}{\text{maximize}} \quad & \text{EE}(\mathbf{t},\mathbf{u}) = \frac{R_{\text{sum}}(\mathbf{t},\mathbf{u})}{U_{\text{p}}(\mathbf{t},\mathbf{u})} \qquad (13.42)\\
\text{s.t.} \quad & \text{C1, C2, C3,}\\
& \text{C4: } \sum_{k=1}^{K}\sum_{i=1}^{N_{\text{F}}} t_{k,i} \le p_{\max},\\
& \text{C5: } R_k(\mathbf{t},\mathbf{u}) \ge R_{\min}, \forall k,\\
& \text{C6: } t_{k,i} \ge 0, \ \forall k,i,\\
& \text{C7: } t_{k,i} \le p_{k,i}, \ \forall k,i,\\
& \text{C8: } t_{k,i} \le u_{k,i}p_{\max}, \ \forall k,i,\\
& \text{C9: } t_{k,i} \ge p_{k,i} - \left(1-u_{k,i}\right)p_{\max}, \ \forall k,i,
\end{aligned}$$

where new constraints C6–C9 are imposed following the big-M formulation [85] and $\mathbf{t} \in \mathbb{R}^{KN_{\text{F}}\times 1}$ collects all the auxiliary variables $t_{k,i}$. Furthermore, we rewrite the binary constraint C1 in its equivalent form:

$$\begin{aligned}
& \text{C1a: } 0 \le u_{k,i} \le 1, \ \forall k,i, \ \text{ and}\\
& \text{C1b: } \sum_{k=1}^{K}\sum_{i=1}^{N_{\text{F}}} u_{k,i} - \sum_{k=1}^{K}\sum_{i=1}^{N_{\text{F}}} u_{k,i}^2 \le 0. \qquad (13.43)
\end{aligned}$$

Then, we augment the constraint C1b into the objective with a penalty factor $\eta \gg 1$ [58,71], which results in the following optimization problem:

$$\begin{aligned}
\underset{\mathbf{p},\mathbf{t},\mathbf{u}}{\text{maximize}} \quad & \frac{R_{\text{sum}}(\mathbf{t},\mathbf{u}) - \eta\sum_{k=1}^{K}\sum_{i=1}^{N_{\text{F}}}\left(u_{k,i}-u_{k,i}^2\right)}{U_{\text{p}}(\mathbf{t},\mathbf{u})} \qquad (13.44)\\
\text{s.t.} \quad & \text{C1a, C2–C9.}
\end{aligned}$$

According to [58,71], the transformed problem in (13.44) is equivalent to the problem in (13.42). Then, with adopting the auxiliary variables $t_{k,i}$, the individual rate and the system sum-rate can be given by

$$\begin{aligned} R_k(\mathbf{t},\mathbf{u}) &= \sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right) \\ &\quad - \sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k-1} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right) \text{ and} \\ R_{\mathrm{sum}}(\mathbf{t},\mathbf{u}) &= \sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right) \\ &\quad - \sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k-1} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right), \end{aligned} \tag{13.45}$$

respectively.

Now, the problem in (13.44) can be rewritten as

$$\begin{aligned} \underset{\mathbf{p},\mathbf{t},\mathbf{u}}{\text{maximize}} \quad & \frac{\sum_{k=1}^{K}\left(F_k(\mathbf{t},\mathbf{u})-D_k(\mathbf{t},\mathbf{u})\right)-\eta\sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{F}}}\left(u_{k,i}-u_{k,i}^2\right)}{U_{\mathrm{p}}(\mathbf{t},\mathbf{u})} \\ \text{s.t.} \quad & \text{C1a, C2–C4, C6–C9,} \\ & \text{C5}: F_k(\mathbf{t},\mathbf{u})-D_k(\mathbf{t},\mathbf{u})\geq R_{\min},\forall k, \end{aligned} \tag{13.46}$$

where

$$\begin{aligned} F_k(\mathbf{t},\mathbf{u}) &= \sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right) \text{ and} \\ D_k(\mathbf{t},\mathbf{u}) &= \sum_{i=1}^{N_{\mathrm{F}}} u_{k,i}\log_2\left(\frac{\sum_{k'=1}^{k-1} t_{k',i}\left|h_{k,i}\right|^2}{u_{k,i}}+\sigma^2\right). \end{aligned} \tag{13.47}$$

Following the idea in Algorithm 13.2, before using the Dinkelbach method, we perform SCA to approximate the numerator of the objective function and the left-hand side of constraint C5 in (13.46). In particular, in each iteration of the outer loop, given a feasible solution $\left(\mathbf{p}^{\mathrm{outer}},\mathbf{t}^{\mathrm{outer}},\mathbf{u}^{\mathrm{outer}}\right)$, we can provide a lower bound for the problem in (13.46) via solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{p},\mathbf{t},\mathbf{u}}{\text{maximize}} \quad & \frac{Z^{\mathrm{outer}}(\mathbf{t},\mathbf{u})}{U_{\mathrm{p}}(\mathbf{t},\mathbf{u})} \\ \text{s.t.} \quad & \text{C1a, C2–C4, C6–C9,} \\ & \overline{\text{C5}}: F_k(\mathbf{t},\mathbf{u})-D_k^{\mathrm{outer}}(\mathbf{t},\mathbf{u})\geq R_{\min},\forall k, \end{aligned} \tag{13.48}$$

where

$$
\begin{aligned}
Z^{\text{outer}}(\mathbf{t},\mathbf{u}) &= \sum_{k=1}^{K}\left(F_k(\mathbf{t},\mathbf{u}) - D_k^{\text{outer}}(\mathbf{t},\mathbf{u})\right) \\
&\quad - \eta\sum_{k=1}^{K}\sum_{i=1}^{N_{\text{F}}}\left(u_{k,i} - 2u_{k,i}^{\text{outer}}u_{k,i} + (u_{k,i}^{\text{outer}})^2\right) \text{ and} \\
D_k^{\text{outer}}(\mathbf{t},\mathbf{u}) &= D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right) + \left[\nabla_{\mathbf{t}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)\right]^{\text{T}}\left(\mathbf{t}-\mathbf{t}^{\text{outer}}\right) \\
&\quad + \left[\nabla_{\mathbf{u}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)\right]^{\text{T}}\left(\mathbf{u}-\mathbf{u}^{\text{outer}}\right). \qquad (13.49)
\end{aligned}
$$

Vectors $\nabla_{\mathbf{t}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)$ and $\nabla_{\mathbf{u}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)$ denote the gradient vectors of $D_k(\mathbf{t},\mathbf{u})$ w.r.t. $\mathbf{t}$ and $\mathbf{u}$ at $\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)$, respectively. The second and third terms of (13.49) are given by

$$
\begin{aligned}
&\left[\nabla_{\mathbf{t}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)\right]^{\text{T}}\left(\mathbf{t}-\mathbf{t}^{\text{outer}}\right) \\
&= \frac{1}{\ln 2}\sum_{i=1}^{N_{\text{F}}}\frac{u_{k,i}^{\text{outer}}}{\sum_{k'=1}^{k-1}t_{k',i}^{\text{outer}}\left|h_{k,i}\right|^2 + u_{k,i}^{\text{outer}}\sigma^2}\left|h_{k,i}\right|^2\sum_{k'=1}^{k-1}\left(t_{k',i}-t_{k',i}^{\text{outer}}\right) \text{ and} \\
&\left[\nabla_{\mathbf{u}}D_k\left(\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)\right]^{\text{T}}\left(\mathbf{u}-\mathbf{u}^{\text{outer}}\right) \\
&= \frac{1}{\ln 2}\sum_{i=1}^{N_{\text{F}}}\sum_{k=1}^{K}\ln\left(\frac{\sum_{k'=1}^{k-1}t_{k',i}^{\text{outer}}\left|h_{k,i}\right|^2}{u_{k,i}^{\text{outer}}}+\sigma^2\right)\left(u_{k,i}-u_{k,i}^{\text{outer}}\right) \\
&\quad - \frac{1}{\ln 2}\sum_{i=1}^{N_{\text{F}}}\sum_{k=1}^{K}\frac{\sum_{k'=1}^{k-1}t_{k',i}^{\text{outer}}\left|h_{k,i}\right|^2}{\sum_{k'=1}^{k-1}t_{k',i}^{\text{outer}}\left|h_{k,i}\right|^2 + u_{k,i}^{\text{outer}}\sigma^2}\left(u_{k,i}-u_{k,i}^{\text{outer}}\right), \qquad (13.50)
\end{aligned}
$$

respectively. Note that the approximation in constraint $\overline{\text{C5}}$ in (13.48) results in a smaller feasible solution set, and solving (13.48) yields a lower bound for the problem in (13.46). Thus, following the similar idea in (13.30), we can prove the convergence of the SCA in the outer loop.

According to the Dinkelbach method in Algorithm 13.1, the problem in (13.48) can be solved by solving the problem as follows iteratively:

$$
\begin{aligned}
\underset{\mathbf{p},\mathbf{t},\mathbf{u}}{\text{maximize}} \quad & Z^{\text{outer}}(\mathbf{t},\mathbf{u}) - \widetilde{\text{EE}}^{\text{inner}}U_{\text{p}}(\mathbf{t},\mathbf{u}) \qquad (13.51)\\
\text{s.t.} \quad & \text{C1a, C2–C4, } \overline{\text{C5}}\text{, C6–C9.}
\end{aligned}
$$

In each iteration of the inner loop, the obtained solution of the problem in (13.52) is denoted as $\left(\widetilde{\mathbf{p}}^{\text{inner}},\widetilde{\mathbf{t}}^{\text{inner}},\widetilde{\mathbf{u}}^{\text{inner}}\right)$, and the intermediate parameter EE_{inner} is updated by

$$
\widetilde{\text{EE}}^{\text{inner}+1} = \frac{Z^{\text{outer}}\left(\widetilde{\mathbf{t}}^{\text{inner}},\widetilde{\mathbf{u}}^{\text{inner}}\right)}{U_{\text{p}}\left(\widetilde{\mathbf{t}}^{\text{inner}},\widetilde{\mathbf{u}}^{\text{inner}}\right)}. \qquad (13.52)
$$

Now, Algorithm 13.2 can be employed to solve the problem in (13.44) with substituting $\mathbf{x}^{\text{outer}} = \left(\mathbf{p}^{\text{outer}},\mathbf{t}^{\text{outer}},\mathbf{u}^{\text{outer}}\right)$, $\widetilde{\mathbf{x}}^{\text{inner}} = \left(\widetilde{\mathbf{p}}^{\text{inner}},\widetilde{\mathbf{t}}^{\text{inner}},\widetilde{\mathbf{u}}^{\text{inner}}\right)$, $f^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right) = Z^{\text{outer}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)$, and $g\left(\widetilde{\mathbf{x}}^{\text{inner}}\right) = U_{\text{p}}\left(\widetilde{\mathbf{x}}^{\text{inner}}\right)$.

13.6 Simulation results and discussions

In this section, simulations are performed to evaluate the performance of the proposed energy-efficient resource allocation design algorithms based on the illustrative example in Section 13.5. In addition, we unveil some interesting insights about energy-efficient resource allocation design. Unless specified otherwise, the simulation setting is given as follows. A single-cell system with a cell size $D = 500$ m is considered. A single-antenna BS is located at the center of the cell. All the $K = 8$ single-antenna users are uniformly distributed in the cell. The propagation path loss from the BS to each user follows the path-loss model in [86] with a path-loss exponent 3.6. The number of subcarriers in the considered MC-NOMA system is $N_{\mathrm{F}} = 6$, each with an equal subcarrier spacing of $W = 15$ kHz. As the number of subcarriers is smaller than the number of users, i.e., $N_{\mathrm{F}} < K$, NOMA is utilized to handle this overload scenario. However, to obtain an acceptable SIC decoding complexity and delay at user side, we assume $L = 2$ for the considered MC-NOMA system. In the power consumption model in (13.14), the fixed circuit power consumption is $P_{\mathrm{C}} = 1$ W and $\delta = 2$. The maximum transmit power of the considered MC-NOMA system $p_{\max}$ ranges from 20 to 45 dBm. The noise power spectral density is $N_0 = -130$ dBm/Hz and hence the noise power in each subcarrier is $\sigma^2 = -88.2$ dBm. In addition, the requested minimum data rate $R_{\min}$ ranges from 0.25 to 1 bit/s/Hz for each user. The system energy efficiencies obtained in the following simulation cases are averaged over multiple channel realizations.

13.6.1 Convergence of the proposed algorithms

In this section, we verify the convergence of the proposed algorithms developed in Algorithms 13.1 and 13.2. For initialization, we consider an equal resource allocation strategy in the first iteration of outer loop, i.e., $u_{k,i} = \frac{L}{K}$ and $t_{k,i} = \frac{p_{\max}}{KN_{\mathrm{F}}}$. In these simulations, we consider $p_{\max} = 40$ dBm and $R_{\min} = [0.25, 0.5, 1]$ bit/s/Hz.

Figure 13.10 illustrates the convergence of the Dinkelbach algorithm in the first round of the outer loop in Algorithm 13.1. We can observe that the system energy efficiency monotonically increases with the number of inner loop iterations. In addition, the Dinkelbach algorithm enjoys both quick convergence and low complexity. Figure 13.11 illustrates the convergence of the outer loop of the proposed energy-efficient resource allocation algorithm in Algorithm 13.2. We can observe that the system energy efficiency also monotonically increases with the number of outer loop iterations and quickly converges to a stationary point. Besides, we can observe that the higher the required minimum data rate, the smaller the converged system energy efficiency. This is due to the fact that a more stringent data requirement, $R_{\min}$, results in a smaller feasible solution set for the problem in (13.41) and thus leads to a smaller system energy efficiency.

13.6.2 System energy efficiency versus the total transmit power

Figure 13.12 illustrates the system energy efficiency versus the maximum transmit power $p_{\max}$ with $R_{\min} = [0.25, 0.5, 1]$ bit/s/Hz. It can be seen that the system energy

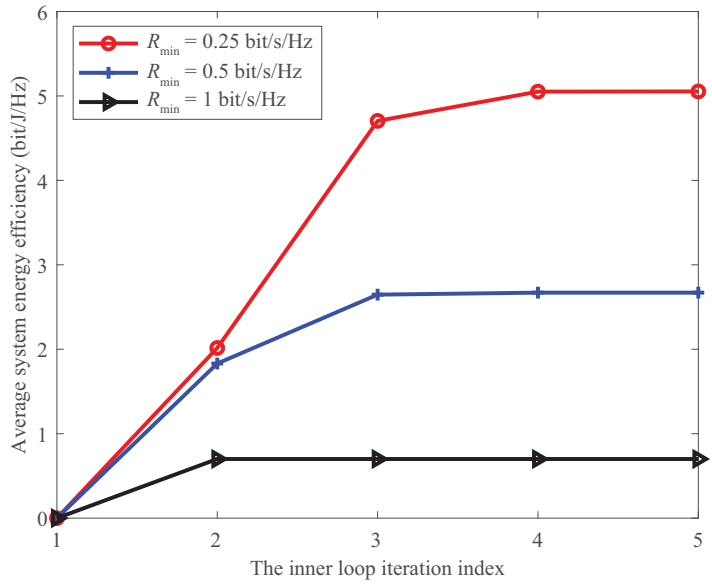


Figure 13.10 The convergence of the Dinkelbach algorithm

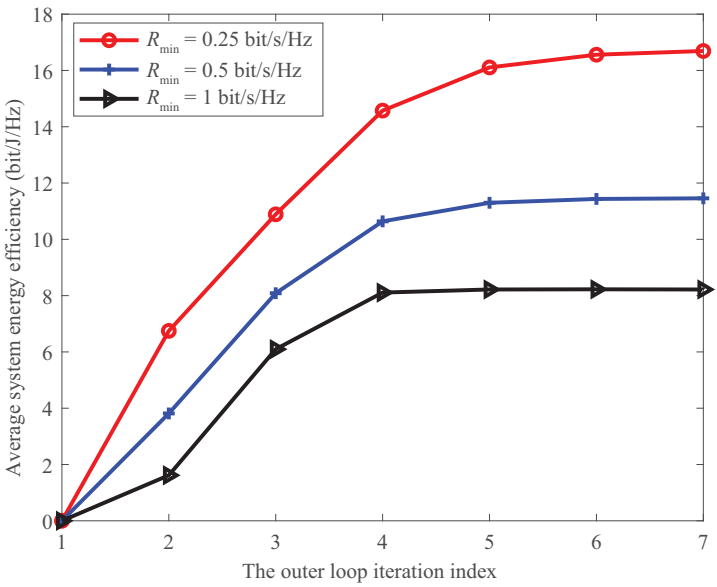


Figure 13.11 The convergence of the outer loop

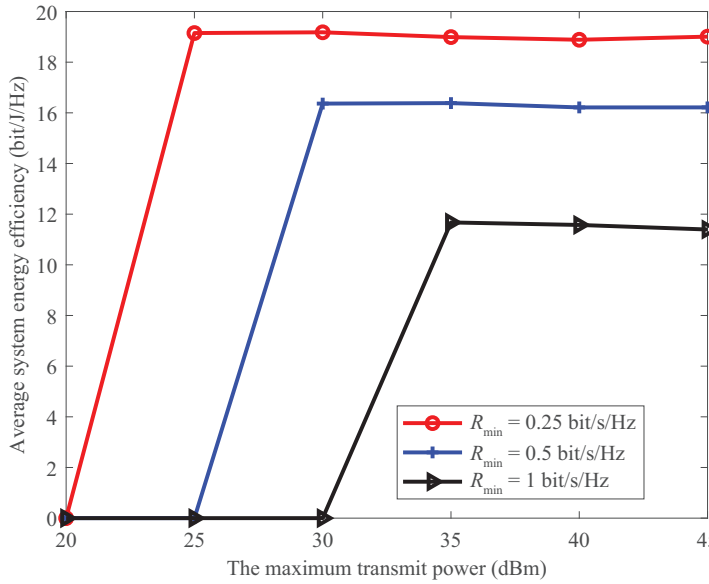


Figure 13.12 The average system energy efficiency versus the maximum transmit power

efficiency first increases with $p_{\max}$ and then saturates in the high SNR regime. In fact, as demonstrated in Figure 13.7(a), in the low SNR regime, the system energy efficiency is limited by the achievable sum-rate. Therefore, the energy-efficient resource allocator fully exploits the increased transmit power in the low SNR regime. However, there is a diminishing return in the spectral efficiency when allocating more transmit power in the high SNR regime. Hence, the system power consumption would outweigh the spectral efficiency gain when allocating more transmit power in the high SNR regime. Therefore, in the high SNR regime, the energy-efficient resource allocator only utilizes the right enough amount of the total power budget and further increasing $p_{\max}$ cannot increase the system energy efficiency. It can be seen that for small transmit power, the system energy efficiency is zero since the optimization problem in (13.41) is infeasible due to an insufficient power to satisfy the data rate constraint C5.

13.7 Conclusions

This chapter has introduced the fundamentals of NOMA and energy efficiency and presented the energy-efficient resource allocation for NOMA systems. In particular, the trade-off between energy efficiency and spectral efficiency for NOMA systems has been revealed, which provides some insights for practical implementations of NOMA. A generic problem formulation for energy-efficient resource allocation design for

NOMA systems was proposed, which is in the form of fractional programming. We proposed an iterative energy-efficient resource allocation algorithm, where an SCA-based outer loop handles the non-convexity arising from the IUI in NOMA systems and the Dinkelbach method serves as the inner loop to solve the transformed problem. The proposed can achieve a stationary point with a polynomial computational complexity. As an illustrative example, we studied the energy-efficient resource allocation design for an MC-NOMA system. The energy-efficient power allocation and user scheduling were formulated as an optimization problem with taking into account the minimum data rate requirements. The proposed energy-efficient resource allocation algorithm was adopted to solve the formulated problem to strike a balance between the system performance and the computational complexity. Our simulation results have demonstrated the fast convergence of both the inner loop and outer loop in our proposed algorithms. Excellent system energy efficiency can be achieved via proposed resource allocation design.

Appendices

A.1 Proof of Theorem 1

We start the proof by verifying the forward via a similar approach in [79]. Without loss of generality, let $\mathbf{x}^*$ be a solution of the problem in (13.24), i.e.:

$$\mathrm{EE}^* = \frac{f(\mathbf{x}^*)}{g(\mathbf{x}^*)} \geq \frac{f(\mathbf{x})}{g(\mathbf{x})}, \mathbf{x} \in \mathscr{X}. \tag{A.1}$$

Hence, we have

$$f(\mathbf{x}) - \mathrm{EE}^* g(\mathbf{x}) \leq 0, \mathbf{x} \in \mathscr{X}, \tag{A.2}$$

$$f(\mathbf{x}^*) - \mathrm{EE}^* g(\mathbf{x}^*) = 0. \tag{A.3}$$

Comparing (A.2) and (A.3), we can observe that $\mathbf{x}^*$ is the optimal solution of the problem in (13.26).

Now, we prove the converse. Let $\mathbf{x}^*$ be a solution of the problem in (13.26), i.e.,

$$f(\mathbf{x}) - \mathrm{EE}^* g(\mathbf{x}) \leq f(\mathbf{x}^*) - \mathrm{EE}^* g(\mathbf{x}^*) = 0, \mathbf{x} \in \mathscr{X}. \tag{A.4}$$

Hence, we have

$$\mathrm{EE}^* \geq \frac{f(\mathbf{x})}{g(\mathbf{x})}, \mathbf{x} \in \mathscr{X}, \tag{A.5}$$

$$\mathrm{EE}^* = \frac{f(\mathbf{x}^*)}{g(\mathbf{x}^*)}. \tag{A.6}$$

Comparing (A.5) and (A.6), we can observe that $\mathbf{x}^*$ is the optimal solution of the problem in (13.24) and it results in the optimal value EE^*. It completes the proof of Theorem 1.

A.2 Proof of Theorem 2

Let us define the optimal value of the problem in (13.26) given the parameter EE as

$$F(\mathrm{EE}) = \underset{\mathbf{x}\in\mathscr{X}}{\text{maximize}}\ f(\mathbf{x}) - \mathrm{EE}g(\mathbf{x})\,. \tag{A.7}$$

For any feasible solution $\widetilde{\mathbf{x}} \in \mathscr{X}$ and $\widetilde{\mathrm{EE}} = \frac{f(\widetilde{\mathbf{x}})}{g(\widetilde{\mathbf{x}})}$, we can observe that

$$F\big(\widetilde{\mathrm{EE}}\big) = \underset{\mathbf{x}\in\mathscr{X}}{\text{maximize}}\ f(\mathbf{x}) - \frac{f(\widetilde{\mathbf{x}})}{g(\widetilde{\mathbf{x}})}g(\mathbf{x}) \geq f(\widetilde{\mathbf{x}}) - \frac{f(\widetilde{\mathbf{x}})}{g(\widetilde{\mathbf{x}})}g(\widetilde{\mathbf{x}}) = 0. \tag{A.8}$$

Therefore, following the updating rule in line 8 of the Dinkelbach algorithm in Algorithm 13.1, we have

$$F\big(\mathrm{EE}^{\mathrm{iter}}\big) = f\big(\mathbf{x}^{\mathrm{iter}}\big) - \mathrm{EE}^{\mathrm{iter}}g\big(\mathbf{x}^{\mathrm{iter}}\big) = \big(\mathrm{EE}^{\mathrm{iter}+1} - \mathrm{EE}^{\mathrm{iter}}\big)\,g\big(\mathbf{x}^{\mathrm{iter}}\big) \geq 0. \tag{A.9}$$

Since $g(\mathbf{x}) > 0$, we have $\mathrm{EE}^{\mathrm{iter}+1} \geq \mathrm{EE}^{\mathrm{iter}}$. It implies that the system energy efficiency EE monotonically increases with the number of iterations in the Dinkelbach algorithm. Besides, we can observe that

$$\begin{aligned} F\big(\mathrm{EE}^{\mathrm{iter}+1}\big) &= f\big(\mathbf{x}^{\mathrm{iter}+1}\big) - \mathrm{EE}^{\mathrm{iter}+1}g\big(\mathbf{x}^{\mathrm{iter}+1}\big) \\ &< f\big(\mathbf{x}^{\mathrm{iter}+1}\big) - \mathrm{EE}^{\mathrm{iter}}g\big(\mathbf{x}^{\mathrm{iter}+1}\big) \\ &\leq \underset{\mathbf{x}\in\mathscr{X}}{\text{maximize}}\ f(\mathbf{x}) - \mathrm{EE}^{\mathrm{iter}}g(\mathbf{x}) \\ &= F\big(\mathrm{EE}^{\mathrm{iter}}\big)\,, \end{aligned} \tag{A.10}$$

which implies the monotonically decreasing feature of $F(\mathrm{EE})$ with the proceeding of iterations in the Dinkelbach algorithm. It completes the proof of the convergence of the Dinkelbach algorithm.

In the following, we prove the optimality of the Dinkelbach algorithm by contradiction. Let EE^* denote the optimal value of the problem in (13.24). According to Theorem 1, we have

$$F(\mathrm{EE}^*) = 0. \tag{A.11}$$

Assuming the convergence point $\big\{\mathbf{x}^{\mathrm{iter}}, \mathrm{EE}^{\mathrm{iter}}\big\}$ is not the optimal point, i.e., $\mathrm{EE}^{\mathrm{iter}} < \mathrm{EE}^*$ and $\mathbf{x}^{\mathrm{iter}} = \arg\underset{\mathbf{x}\in\mathscr{X}}{\text{maximize}}\ f(\mathbf{x}) - \mathrm{EE}^{\mathrm{iter}}g(\mathbf{x})$. Thus we have $F\big(\mathrm{EE}^{\mathrm{iter}}\big) = 0$. However, according to (A.10), we have $F(\mathrm{EE}^*) < F\big(\mathrm{EE}^{\mathrm{iter}}\big) = 0$, which contradicts to (A.11).

Note that the proof shown before is based on the condition that the problem in (13.24) can be solved optimally.

References

[1] Zorzi M, Gluhak A, Lange S, *et al.* From Today's INTRAnet of Things to a Future INTERnet of Things: A Wireless- and Mobility-Related View. IEEE Wireless Commun. 2010;17(6):44–51.

[2] The 5G Unified Air Interface: Scalable to an extreme variation of requirements. Qualcomm Technologies Inc.; 2015.

[3] Sun Z, Xie Y, Yuan J, *et al.* Coded Slotted ALOHA for Erasure Channels: Design and Throughput Analysis. IEEE Trans Commun. 2017; 65(11): 4817–4830.

[4] Sun Z, Wei Z, Yang L, *et al.* Joint user identification and channel estimation in massive connectivity with transmission control. In: Proc. IEEE Intern. Sympos. on Turbo Codes Iterative Information Process; 2018. pp. 1–5.

[5] Andrews JG, Buzzi S, Choi W, *et al.* What Will 5G Be? IEEE J Sel Areas Commun. 2014;32(6):1065–1082.

[6] Wong VW, Schober R, Ng DWK, *et al.* Key technologies for 5G wireless systems. Cambridge University Press; 2017.

[7] Dai L, Wang B, Yuan Y, *et al.* Non-Orthogonal Multiple Large-Scale Underlay Access for 5G: Solutions, Challenges, Opportunities, and Future Research Trends. IEEE Commun Mag. 2015;53(9):74–81.

[8] Ding Z, Liu Y, Choi J, *et al.* Application of Non-Orthogonal Multiple Access in LTE and 5G Networks. IEEE Commun Mag. 2017;55(2):185–191.

[9] Wei Z, Jinhong Y, Ng DWK, *et al.* A Survey of Downlink Non-Orthogonal Multiple Access for 5G Wireless Communication Networks. ZTE Commun. 2016;14(4):17–25.

[10] Sun Z, Wei Z, Yang L, Yuan J, Cheng X, and Wan L. Exploiting Transmission Control for Joint User Identification and Channel Estimation in Massive Connectivity. IEEE Trans Commun. 2019;67(9):6311–6326.

[11] Liu B, Wei Z, Yuan J, and Pajovic M. Deep Learning Assisted User Identification in Massive Machine-Type Communications. In: 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA; 2019. pp. 1–6.

[12] Wei Z, Ng DWK, Yuan J, *et al.* Optimal Resource Allocation for Power-Efficient MC-NOMA with Imperfect Channel State Information. IEEE Trans Commun. 2017;65(9):3944–3961.

[13] Wei Z, Ng DWK, and Yuan J. Joint Pilot and Payload Power Control for Uplink MIMO-NOMA with MRC-SIC Receivers. IEEE Commun Lett. 2018;22(4):692–695.

[14] Tse D and Viswanath P. Fundamentals of wireless communication. Cambridge University Press; 2005.

[15] Cover T. Broadcast Channels. IEEE Trans Inf Theory. 1972;18(1):2–14.

[16] Cover TM and Thomas JA. Elements of information theory. John Wiley & Sons, Inc.; 1991.

[17] Qiu M, Huang Y, Yuan J, *et al.* Lattice-Partition-Based Downlink Non-Orthogonal Multiple Access Without SIC for Slow Fading Channels. IEEE Trans Commun. 2019;67(2):1166–1181.

[18] Wei Z, Zhao L, Guo J, *et al.* Multi-Beam NOMA for Hybrid mmWave Systems. IEEE Trans Commun. 2019;67(2):1705–1719.
[19] Qiu M, Huang Y, Shieh S, *et al.* A Lattice-Partition Framework of Downlink Non-Orthogonal Multiple Access Without SIC. IEEE Trans Commun. 2018;66(6):2532–2546.
[20] Wei Z, Kwan Ng DW, and Yuan J. NOMA for Hybrid MmWave Communication Systems With Beamwidth Control. IEEE J Sel Top Signal Process. 2019;13(3):567–583.
[21] Wei Z, Yang L, Ng DWK, Yuan J, and Hanzo L. On the Performance Gain of NOMA Over OMA in Uplink Communication Systems. IEEE Trans Commun. 2020;68(1):536–568.
[22] 3GPP TR 36.859. Study on Downlink Multiuser Superstition Transmission (MUST) for LTE (Release 13); 2015.
[23] DOCOMO, MediaTek. DOCOMO and MediaTek Achieve World's First Successful 5G Trial Using Smartphone-Sized NOMA Chipset-Embedded Device to Increase Spectral Efficiency; 2017. Available from: https://www.nttdocomo.co.jp/english/info/media_center/pr/2017/1102_02.html.
[24] Chen Y, Bayesteh A, Wu Y, *et al.* Toward the Standardization of Non-Orthogonal Multiple Access for Next Generation Wireless Networks. IEEE Commun Mag. 2018;56(3):19–27.
[25] Zappone A, Jorswieck E, *et al.* Energy Efficiency in Wireless Networks via Fractional Programming Theory. Found Trends Commun Inf Theory. 2015;11(3–4):185–396.
[26] Wu Q, Li GY, Chen W, *et al.* An Overview of Sustainable Green 5G Networks. IEEE Wireless Commun. 2017;24(4):72–80.
[27] Wu Q, Chen W, Ng DWK, *et al.* Spectral and Energy-Efficient Wireless Powered IoT Networks: NOMA or TDMA? IEEE Trans Veh Technol. 2018;67(7):6663–6667.
[28] Zappone A, Sanguinetti L, Bacci G, *et al.* Energy-Efficient Power Control: A Look at 5G Wireless Technologies. IEEE Trans Signal Process. 2015;64(7):1668–1683.
[29] Zappone A and Jorswieck EA. Energy-Efficient Resource Allocation in Future Wireless Networks by Sequential Fractional Programming. Digital Signal Process. 2017;60:324–337.
[30] Ng DWK, Lo ES, and Schober R. Energy-Efficient Resource Allocation in OFDMA Systems With Large Numbers of Base Station Antennas. IEEE Trans Wireless Commun. 2012;11(9):3292–3304.
[31] Ng DWK, Lo ES, and Schober R. Wireless Information and Power Transfer: Energy Efficiency Optimization in OFDMA Systems. IEEE Trans Wireless Commun. 2013;12(12):6352–6370.
[32] Ng DWK, Lo ES, and Schober R. Energy-Efficient Resource Allocation in Multi-Cell OFDMA Systems With Limited Backhaul Capacity. IEEE Trans Wireless Commun. 2012;11(10):3618–3631.
[33] Ng DWK, Lo ES, and Schober R. Energy-Efficient Resource Allocation for Secure OFDMA Systems. IEEE Trans Veh Technol. 2012;61(6):2572–2585.

[34] Cai Y, Wei Z, Li R, *et al.* Energy-efficient resource allocation for secure UAV communication systems. In: Proc. IEEE Wireless Commun. and Networking Conf.; 2019. pp. 1–8.
[35] Wang P, Xiao J, and Li P. Comparison of Orthogonal and Non-Orthogonal Approaches to Future Wireless Cellular Systems. IEEE Veh Technol Mag. 2006;1(3):4–11.
[36] Sun Z, Yang L, Yuan J, *et al.* Physical-Layer Network Coding Based Decoding Scheme for Random Access. IEEE Trans Veh Technol. 2019;68(4): 3550–3564.
[37] Timotheou S and Krikidis I. Fairness for Non-Orthogonal Multiple Access in 5G Systems. IEEE Signal Process Lett. 2015;22(10):1647–1651.
[38] Ding Z, Fan P, and Poor HV. Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions. IEEE Trans Veh Technol. 2016;65(8):6010–6023.
[39] Hoshyar R, Wathan FP, and Tafazolli R. Novel Low-Density Signature for Synchronous CDMA Systems Over AWGN Channel. IEEE Trans Signal Process. 2008;56(4):1616–1626.
[40] Hoshyar R, Razavi R, and Al-Imari M. LDS-OFDM an efficient multiple access technique. In: Proc. IEEE Veh. Techn. Conf.; 2010. pp. 1–5.
[41] Razavi R, Al-Imari M, Imran MA, *et al.* On Receiver Design for Uplink Low Density Signature OFDM (LDS-OFDM). IEEE Trans Commun. 2012;60(11):3499–3508.
[42] Huang T, Yuan J, Cheng X, *et al.* Design of degrees of distribution of LDS-OFDM. In: Proc. IEEE Intern. Conf. on Signal Process. and Commun. Syst.; 2015. pp. 1–6.
[43] Nikopour H and Baligh H. Sparse code multiple access. In: Proc. IEEE Personal, Indoor and Mobile Radio Commun. Sympos.; 2013. pp. 332–336.
[44] Yuan W, Wu N, Guo Q, *et al.* Iterative Receivers for Downlink MIMO-SCMA: Message Passing and Distributed Cooperative Detection. IEEE Trans Wireless Commun. 2018;17(5):3444–3458.
[45] Yuan W, Wu N, Yan C, *et al.* A Low-Complexity Energy-Minimization-Based SCMA Detector and Its Convergence Analysis. IEEE Trans Veh Technol. 2018;67(12):12398–12403.
[46] Dai X, Chen S, Sun S, *et al.* Successive interference cancelation amenable multiple access (SAMA) for future wireless communications. In: Proc. IEEE Intern. Commun. Conf.; 2014. pp. 222–226.
[47] Al-Imari M, Xiao P, Imran MA, *et al.* Uplink non-orthogonal multiple access for 5G wireless networks. In: Proc. IEEE Intern. Sympos. on Wireless Commun. Systems; 2014. pp. 781–785.
[48] Al-Imari M, Xiao P, and Imran MA. Receiver and resource allocation optimization for uplink NOMA in 5G wireless networks. In: Proc. IEEE Intern. Sympos. on Wireless Commun. Systems; 2015. pp. 151–155.
[49] Yang Z, Ding Z, Fan P, *et al.* A General Power Allocation Scheme to Guarantee Quality of Service in Downlink and Uplink NOMA Systems. IEEE Trans Wireless Commun. 2016;15(11):7244–7257.

[50] Wei Z, Guo J, Ng DWK, *et al.* Fairness comparison of uplink NOMA and OMA. In: Proc. IEEE Veh. Techn. Conf.; 2017. pp. 1–6.
[51] Ding Z, Yang Z, Fan P, *et al.* On the Performance of Non-Orthogonal Multiple Access in 5G Systems With Randomly Deployed Users. IEEE Signal Process Lett. 2014;21(12):1501–1505.
[52] Liu Y, Ding Z, Elkashlan M, *et al.* Nonorthogonal Multiple Access in Large-Scale Underlay Cognitive Radio Networks. IEEE Trans Veh Technol. 2016;65(12):10152–10157.
[53] Vaezi M, Schober R, Ding Z, and Poor HV. Non-Orthogonal Multiple Access: Common Myths and Critical Questions. IEEE Wireless Commun. 2019;26(5):174–180.
[54] Patel P and Holtzman J. Analysis of a Simple Successive Interference Cancelation Scheme in a DS/CDMA System. IEEE J Sel Areas Commun. 1994;12(5):796–807.
[55] Wolniansky PW, Foschini GJ, Golden GD, *et al.* V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. In: URSI Symp. on Signals, Syst, and Electronics; 1998. pp. 295–300.
[56] Cover TM and Thomas JA. Elements of information theory. New York, NY, USA: Wiley-Interscience; 2006.
[57] Hanif MF, Ding Z, Ratnarajah T, *et al.* A Minorization-Maximization Method for Optimizing Sum Rate in the Downlink of Non-Orthogonal Multiple Access Systems. IEEE Trans Signal Process. 2016;64(1):76–88.
[58] Sun Y, Ng DWK, Ding Z, *et al.* Optimal joint power and subcarrier allocation for MC-NOMA systems. In: Proc. IEEE Global Commun. Conf.; 2016. pp. 1–6.
[59] Sun Q, Han S, Chin-Lin I, *et al.* Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems. In: Proc. IEEE Intern. Commun. Conf.; 2015. pp. 2668–2673.
[60] Zhang Y, Wang HM, Zheng TX, *et al.* Energy-Efficient Transmission Design in Non-Orthogonal Multiple Access. IEEE Trans Veh Technol. 2016;66(3):2852–2857.
[61] Zhang H, Fang F, Cheng J, *et al.* Energy-Efficient Resource Allocation in NOMA Heterogeneous Networks. IEEE Wireless Commun. 2018;25(2): 48–53.
[62] Zhou F, Wu Y, Hu RQ, *et al.* Energy-Efficient NOMA Enabled Heterogeneous Cloud Radio Access Networks. IEEE Network. 2018;32(2):152–160.
[63] Liu Y, Elkashlan M, Ding Z, *et al.* Fairness of User Clustering in MIMO Non-Orthogonal Multiple Access Systems. IEEE Commun Lett. 2016;20(7):1465–1468.
[64] Ng DWK, Lo ES, and Schober R. Robust Beamforming for Secure Communication in Systems With Wireless Information and Power Transfer. IEEE Trans Wireless Commun. 2014;13(8):4599–4615.
[65] Hasan Z, Boostanimehr H, and Bhargava VK. Green Cellular Networks: A Survey, Some Research Issues and Challenges. IEEE Commun Surv Tutor Mag. 2011;13(4):524–540.

[66] Ng DWK and Schober R. Cross-Layer Scheduling for OFDMA Amplify-and-Forward Relay Networks. IEEE Trans Veh Technol. 2010;59(3):1443–1458.
[67] Ng DWK and Schober R. Resource Allocation and Scheduling in Multi-Cell OFDMA Systems With Decode-and-Forward Relaying. IEEE Trans Wireless Commun. 2011;10(7):2246–2258.
[68] Li R, Wei Z, Yang L, *et al.* Joint trajectory and resource allocation design for UAV communication systems. In: Proc. IEEE Global Commun. Conf.; 2018. pp. 1–7.
[69] Leng S, Ng DWK, and Schober R. Power efficient and secure multiuser communication systems with wireless information and power transfer. In: Proc. IEEE Intern. Commun. Conf.; 2014. pp. 800–806.
[70] Ng DWK, Sun Y, and Schober R. Power efficient and secure full-duplex wireless communication systems. In: Proc. IEEE Conf. on Commun. and Network Security; 2015. pp. 1–6.
[71] Ng DWK, Wu Y, and Schober R. Power Efficient Resource Allocation for Full-Duplex Radio Distributed Antenna Networks. IEEE Trans Wireless Commun. 2016;15(4):2896–2911.
[72] Wei Z, Ng DWK, and Yuan J. Power-efficient resource allocation for MC-NOMA with statistical channel state information. In: Proc. IEEE Global Commun. Conf.; 2016. pp. 1–7.
[73] Sun Y, Ng DWK, Zhu J, *et al.* Multi-Objective Optimization for Robust Power Efficient and Secure Full-Duplex Wireless Communication Systems. IEEE Trans Wireless Commun. 2016;15(8):5511–5526.
[74] Boshkovska E, Ng DWK, Dai L, *et al.* Power-Efficient and Secure WPCNs With Hardware Impairments and Non-Linear EH Circuit. IEEE Trans Commun. 2018;66(6):2642–2657.
[75] Boyd S and Vandenberghe L. Convex optimization. Cambridge University Press; 2004.
[76] Horst R and Tuy H. Global optimization: Deterministic approaches. Springer Science & Business Media; 2013.
[77] Wang K, So AM, Chang T, *et al.* Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks: Tractable Approximations by Conic Optimization. IEEE Trans Signal Process. 2014;62(21):5690–5705.
[78] Zhang X, Palomar DP, and Ottersten B. Statistically Robust Design of Linear MIMO Transceivers. IEEE Trans Signal Process. 2008;56(8):3678–3689.
[79] Dinkelbach W. On Nonlinear Fractional Programming. Manage Sci. 1967;13(7):492–498.
[80] Marks BR and Wright GP. A General Inner Approximation Algorithm for Nonconvex Mathematical Programs. Oper Res. 1978;26(4):681–683.
[81] Zappone A, Björnson E, Sanguinetti L, *et al.* Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks. IEEE Trans Signal Process. 2017;65(11):2844–2859.
[82] Dinh QT and Diehl M. Local convergence of sequential convex programming for nonconvex optimization. In: Recent Advances in Optimization and its Applications in Engineering. Springer; 2010. pp. 93–102.

[83] Vucic N, Shi S, and Schubert M. DC programming approach for resource allocation in wireless networks. In: Proc. Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw.; 2010. pp. 380–386.
[84] Grant M and Boyd S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1; 2014. http://cvxr.com/cvx.
[85] Lee J and Leyffer S. Mixed integer nonlinear programming. vol. 154. Springer Science & Business Media; 2011.
[86] 3GPP TR 36.814. Evolved Universal Terrestrial Radio Access: Further Advancements for E-UTRA Physical Layer Aspects; 2010.

This page intentionally left blank

Chapter 14

Energy-efficient illumination toward green communications

Hany Elgala[1], Ahmed F. Hussein[1] and Monette H. Khadr[1]

The advancement in white light-emitting diodes (LEDs) technology makes it the most preferred highly efficient lighting solution. Not only LEDs consume less energy and reduce carbon emissions, but also their average life expectancy is above 10 years, i.e., 50,000 h. Achieving more than 75% of energy savings has encouraged the widespread use of LEDs for indoor and outdoor applications, as well. As a consequence of the huge investments in the LED-based lighting industry, another emerging technology has grown, which is visible light communications (VLC). For instance, LEDs can switch to various light intensity levels at an extremely fast rate, i.e., imperceivable by a human eye, which allows data to be modulated through light, enabling wireless communications [1]. Recent research discusses how LEDs can be used for communication, positioning, and sensing [2]. Adopting VLC enables the use of an ultrawide range of unregulated visible light, offering 10,000 times more bandwidth capacity than radio frequency (RF)-based technologies. Rates of over 10 Gbps have already been demonstrated using LEDs, and an enticing rate of 100 Gbps was reported using laser diodes [3]. This chapter discusses state-of-the-art VLC modulation techniques, potential indoor scenarios, and associated challenges. In the first section, advancements in modulation schemes that are compatible with illumination requirements are highlighted. Such schemes enable the simultaneous offering of multiple wireless services, including communication, sensing, and security, and will even promote more efficient VLC systems. Then, this chapter discusses the possibility of coexisting VLC with RF technologies, followed by recent advancements in VLC-based multiple-input and multiple-output (MIMO) techniques. Finally, the potential of applying deep learning (DL) algorithms to improve the performance of VLC systems is investigated.

14.1 Introduction

Internet applications are currently predominantly wireless, with cellular and wireless indoor networks, such as Wi-Fi, leading the stage of wireless access. Smart devices

[1]Electrical and Computer Engineering Department, University at Albany – State University of NY, Albany, United States

such as handheld mobile computers, cellphones, and sensors are providing additional capabilities aiming to increase the quality of life for users. According to Cisco, 1.5 mobile devices per capita are expected by 2020, with more than three-fifths of all the devices connected will be smart [4]. Today, over 80% of data traffic is being consumed indoors [5], resulting in a tremendous increase in energy consumption [6]. According to a CNN report, Internet energy consumption in the United States was larger than that of the automotive industry and about half that of the chemical industry in the year 2011. Imagine the number of available Wi-Fi access points around the world and the amount of power they consume even while being idle. The Center for Energy-Efficient Telecommunications at the University of Melbourne dedicated a white paper to study the energy consumption of wireless cloud-based services [7]. Their findings showed that energy calculations climbed from 9.2 TWh in the year 2012 to 43 TWh in 2015, which is a 460% increase within only 3 years. As a result, there was an increase in carbon footprint from 6 megatonnes of CO_2 in 2012 to 30 megatonnes of CO_2 in 2015, which is equivalent of adding 4.9 million cars to the roads. The demand for Internet connectivity is expected to grow even further in the next decade; however, this trend is alarming as the environmental costs of this growth will be detrimental. Hence, alternative greener technologies are needed to off-load wireless traffic and VLC is one of the strongest contenders [8].

14.2 Novel modulation techniques

14.2.1 Mixed-carrier communications

After a decade from introducing mobile devices, the age of untethered cloud virtual reality headsets, and always-connected wireless medical devices are forthcoming. However, the capabilities of current wireless technologies cannot meet the demand of future high degree of heterogeneity in terms of services, device classes, deployment environments, and mobility levels [9]. The present approach to deal with this issue is a portfolio of wide-/local-/personal-/body-area connectivity solutions based on standardized or proprietary technologies that have different types of transceiver front ends, modulating waveforms, medium access protocols, and network features [10–12]. Debates on the benefits of single-carrier modulation (SCM) and multi-carrier modulation (MCM) schemes have led to the choice of different binary and M-ary schemes depending on the targeted application, aiming for better spectral efficiency, power efficiency, and more [13–16]. For instance, current IEEE 802.11 wireless local area network (WLAN) (Wi-Fi) standards are based on the choice of specific SCM and/or MCM schemes.

Most of the research efforts to evolve or revolutionize existing networks do not consider a holistic approach to simultaneously support multiservice communication and sensing for devices having different resource capabilities. They do not assess composite waveforms offering parallel high and low bitstreams and simultaneous transmission of SCM and MCM schemes with integrated beacons needed for sensing. The same waveform could be processed by high-end and low-end devices. Research efforts to enable the multiservice capability of VLC waveforms are very limited.

For example, data communications and dimming are considered in reverse polarity optical orthogonal frequency division multiplexing (RPO-OFDM) [17] to enable high-speed and illumination compatible VLC links. In this technique, OFDM samples are conditioned to form a pulse-width modulated (PWM)-like envelope through the reverse polarity concept, i.e., resembling PWM cycles for digital dimming control. Accordingly, the signal-to-noise ratio (SNR) of the OFDM signal is independent of the brightness level over a wide dimming range.

14.2.1.1 Binary-level transmission

In the lighting industry, pulsed dimming is favored over analog dimming, where a PWM forward current is used to control the light source intensity to avoid color shifts [18]. Based on the set duty cycle of the pulse and the proper pulse duration (T_{PWM}), an LED is driven to operate into "ON" and "OFF" intervals. For example, turning on the light for half the period sets the LED to 50% brightness. The longer the "ON" interval relative to the "OFF" interval, the brighter the LED becomes. The PWM-dimming relies upon the human eye's ability to assimilate the overall average of the emitted light pulses. The rate of a few hundred to hundreds of thousands of pulses per second ensures that flickering is unnoticed [19].

In mixed-carrier communications (MCC), beaconing is realized based on the PWM digital-to-analog converter (PWM-DAC) concept, where each light source is assigned an identity (ID), i.e., a unique frequency value. Specifically, every LED transmits a unique beacon frequency, i.e., a DC-biased sinusoidal signal with V_b, which is sampled by PWM cycles. Thus, the modulating waveform is a series of pulses with a fixed amplitude ratio $(1/k)$ of the peak-to-peak voltage (V_{pp}) of the beacon waveform (i.e., V_{pp}/k) and varying duty cycles that are proportional to the sampling amplitudes of the analog sinusoidal signal as shown in Figure 14.1(a). Based on the PWM-sampled beacons received power from individual light sources, localization algorithms can precisely determine the location of the receiver.

Any analog signal can be mapped into a PWM-like format. Accordingly, such binary-level transmission aims at generating a PWM-sampled layer of beacons, SCM or MCM signals. A 1-bit sigma–delta DAC is a potential candidate being a simple and

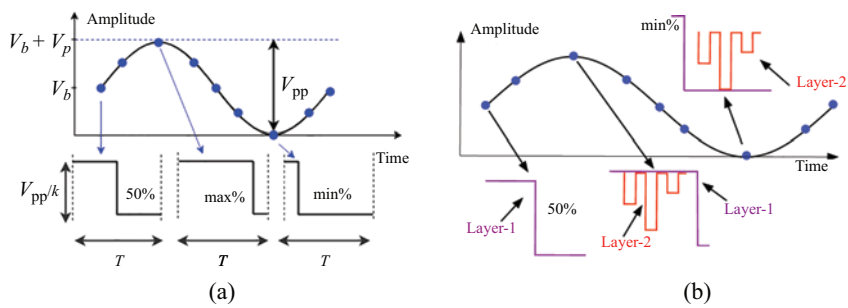


Figure 14.1 (a) Individual PWM-sampled sinusoidal beacon, (b) a Layer-1 PWM of a sinusoidal beacon and multilevel Layer-2 values

theoretically linear approach to realize such 1-bit output [20]. Standard demodulation could still be adopted after the regeneration of the signal at the receiver. Besides a sinusoidal beacon-to-PWM mapping, PWM can be used to linearly represent the bipolar OFDM signal [21]. It is worth mentioning that this approach can greatly contribute to addressing the issues of high power consumption that DACs require for high-order SCM quadrature-amplitude modulation (QAM) and associated high peak-to-average-power ratio (PAPR) in the case of OFDM [22]. In contrast to high PAPR associated with OFDM, the PAPR of such binary-level signals only depends on the symbol length. However, practically, this approach trades off bandwidth efficiency for power efficiency. Further mapping schemes can be optimized and evaluated based on theoretical models and results from practical measurements in real-world scenarios.

14.2.1.2 Multilevel transmission

In Figure 14.1(b), the binary-level PWM mapping/transmission is referred to as Layer-1 transmission. A cycle of a beaconing sinusoid having a certain magnitude and frequency will be mapped into a string of PWM-like periods with varying duty cycles representing magnitudes of individual samples. Toward integrating parallel data streams using multiple SCM and MCM, Layer-1 transmission can form an envelope for an additional Layer-2 transmission to realize an additional stream of bits [23,24]. Layer-2 could be formed based on discrete-level SCM such as pulse-amplitude modulation or continuous MCM signal such as OFDM (see Figure 14.1(b)). It is worth mentioning that the addition of layers is not a superposition in the analog domain, rather a pre-designed multilevel waveform in the digital domain. Different frame structures to integrate different Layer-1 and Layer-2 modulation schemes, as well as beaconing signals, are possible.

14.2.1.3 Frame structure

In this section, the multilevel waveform design and frame structure are presented. The proposed design enables multiservice and device-independent capabilities based on MCC. For instance, it offers two different paths for high-capacity and low-capacity links to serve both high-speed wireless streaming and IoT connectivity. Also, indoor localization and tracking applications based on VLC beacons can be realized. Furthermore, the frame enables dimming control capability that is considered as an essential feature for energy conservation and extending the lifetime of LEDs.

Inspired by SCM phase-shift keying (PSK) and pulse-position modulation (PPM), a transmission layer could be realized by controlling the phase and position of the beacon within the frame. In other words, both position and phase of PWM-sampled beacons can be manipulated to modulate low-capacity bits for IoT applications using low-end devices. Inspired by PSK, PWM cycles with duration T_{PWM} are rearranged based on the target phase to modulate a stream of bits in what is called beacon PSK (BnPSK). As shown in Figure 14.2, a beacon is manipulated to represent different phases that correspond to modulated bits based on quadrature PSK (QPSK) modulation. As a result, demodulating the beacons in Figure 14.2(a)–(d) results in a bit pattern of "00", "01", "10" or "11". In order to maintain dimming control capability, the PWM-sampled beacons are allocated within a frame structure. The frame duration

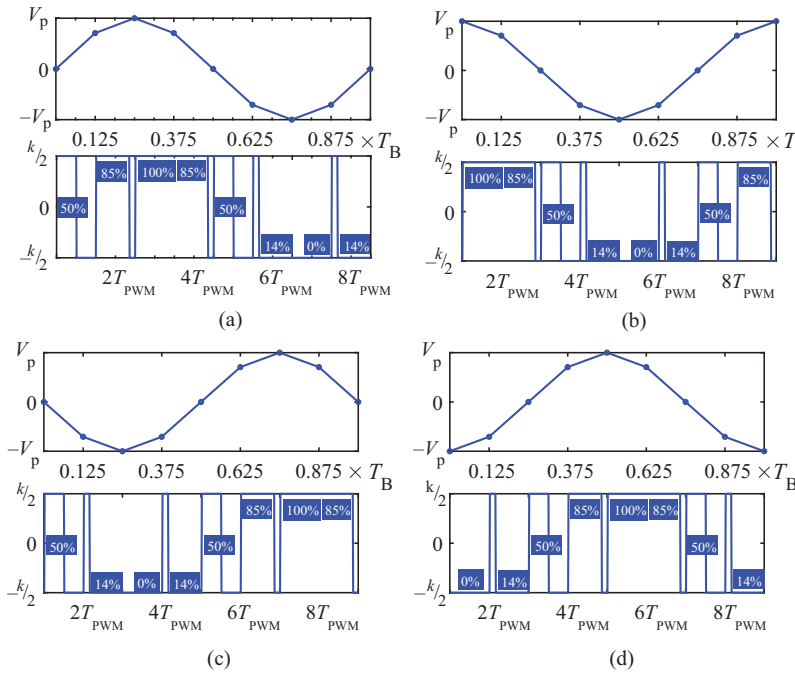


Figure 14.2 A PWM-encoded beacon showing corresponding duty cycles to beacon samples. In parts (a)–(d), beacon's phase control from 0° to 270° is shown to represent Quadrature BnPSK modulation as an example of convenient modulation techniques to low-end receivers

T_f is decided based on the beacon refresh rate, i.e., periodicity of T_B, required by the localization and tracking application. The added PWM-like envelope after and/or before the beacon is controlled to result in the target frame brightness ratio (FBR). FBR is defined to be the effective brightness ratio after averaging the PWM-like envelope, where the relation between the human eye response to light and the designed FBR is nonlinear, as shown in (14.1), due to the eye pupil wideness at low brightness levels [17].

$$\text{Perceived light (\%)} = 100 \times \sqrt{\frac{\text{FBR (\%)}}{100}} \tag{14.1}$$

Figure 14.3(a) illustrates three cases for the dimming control process by varying the duty cycles of the PWM-like envelope, while the PWM-sampled beacon looks the same for every case. In Figure 14.3(b), another layer of SCM is added by manipulating the position of the beacon within the full frame. This beacon-position modulation (BPM) is inspired by the PPM concept. In this figure, 8-BPM is illustrated for N frames, while every individual frame can be controlled to provide a different brightness level.

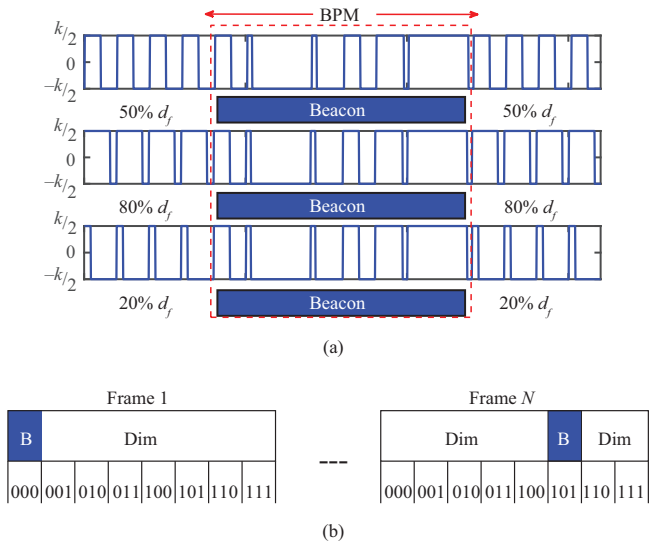


Figure 14.3 *(a) A PWM-encoded beacon placed within different frame structures with varying duty cycles to generate waveforms with varying brightness levels. (b) BPM illustration with eight different time slots per frame, as an example, to modulate the transmitted bits based on BPM concept*

In order to add another layer of MCM to serve reliable broadband links, OFDM samples are conditioned to form the PWM-like envelope required to serve the previous applications, i.e., dimming control, beaconing, BnPSK, and BPM. Thus, the SNR is independent of the brightness over a wide dimming range. Different formats of VLC-compatible OFDM techniques have been introduced in literature limited by the intensity modulation/direct detection (IM/DD) requirements of transmitting positive and real signals. For example, one of the well-known techniques that trades off power efficiency for the sake of higher spectral efficiency is DC-biased optical (DCO)-OFDM. Another technique that trades off spectral efficiency and extended bandwidth for a higher power efficiency is asymmetrically clipped optical (ACO)-OFDM. The stream of bits can be modulated by any of the unipolar optical OFDM techniques such as ACO-OFDM or as introduced in literature, DCO-OFDM can be used with different settings [25–27]. This is shown in Figure 14.4 where the ACO-OFDM samples, shown in blue, comprise a PWM-like envelope for the beacon and the dimming control duty cycles, shown in black.

14.2.1.4 Spectrum management and interference analysis

Due to the nature of the proposed MCC, a thorough design procedure has to be in process to avoid any possible interference between services. This arises from improper spectrum management among the different streams and lack of interference analysis,

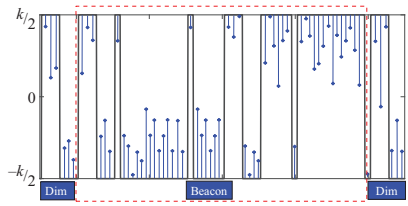


Figure 14.4 ACO-OFDM samples are conditioned to formulate a PWM-like sampled beacon within a PWM-like dimming control frame

leading to the degradation of the received signal quality. In this section, an analytical framework is presented to investigate the spectral synthesis of MCC signals to achieve proper spectrum management. The composite signal is genuinely investigated from a novel perspective compared to the existing approach in literature, where analytical investigation of baseband DCO-OFDM is considered and followed by the spectral analysis of the composite signal.

Proper spectrum management is controlled by the design key parameters such as the number of OFDM symbols per PWM-like cycle, the PWM-like duty cycle, and the frequency components of the PWM-like envelope as well as the OFDM subcarriers. In order to understand the design key parameters affecting the performance of MCC transmission, the following equations represent the time-domain shape of the MCC signal and its spectral profile. In (14.2), $x_h(t)$ represents the hth time-domain baseband DCO-OFDM symbol within a frame of OFDM symbols based on [28,29]. In (14.3), a PWM-like envelope of consecutive 50% duty cycles, $E_{\text{PWM}}(t)$, represents the dimming control envelope of the composite signal. The 50% duty cycle is initially chosen to demonstrate the analysis in a simplified form before being generalized for various duty cycles. The samples of H OFDM symbols are reformed to take the PWM-like envelope shape represented in (14.3). The samples of the OFDM symbols are transmitted consecutively within this PWM-like envelope, where the samples that form the odd PWM-like cycles are DC-biased with amplitude A after being reversed, while the samples that form the even PWM-like cycles are not changed. This is detailed in (14.4) that results in the generation of the dimming control time-domain composite signal, $w_{\text{d}}(t)$, as shown in (14.5).

$$x_h(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} a_{k,h}\, e^{j\,2\pi f_k (t - h\,T_{\text{OFDM}})} \tag{14.2}$$

$$E_{\text{PWM}}(t) = \sum_{h=1}^{H} rect\left(\frac{t - (2h-1)\frac{T_{\text{OFDM}}}{2}}{T_{\text{OFDM}}}\right)(-1)^h \tag{14.3}$$

$$E_{\text{DC}}(t) = \sum_{h=1,3,5,\ldots}^{H} A\, rect\left(\frac{t - (2h-1)\frac{T_{\text{OFDM}}}{2}}{T_{\text{OFDM}}}\right) \tag{14.4}$$

$$w_{\mathrm{d}}(t) = \left[E_{\mathrm{PWM}}(t) \times \sum_{h=1}^{H} x_h(t) \right] + E_{\mathrm{DC}}(t) \tag{14.5}$$

where N is the length of the fast Fourier transform (FFT) and the inverse FFT (IFFT), i.e., number of subcarriers, $a_{k,h}$ is the data symbol modulating the kth subcarrier in the hth OFDM symbol period, f_k is the frequency of the kth subcarrier, and T_{OFDM} is the OFDM symbol duration. In the previous set of equations, the generation process of the composite MCC signal in time-domain is clearly shown. The same set of equations is used to synthesize the spectral profile of the designed waveform. The following set of equations clarifies how the signal is presented in the frequency-domain. In (14.6), $X_h(f)$ represents the spectrum of a single OFDM time symbol. In (14.7), $W_{\mathrm{d}}(f)$ represents the spectrum of the 50% dimming control composite signal, $w_{\mathrm{d}}(t)$, previously shown in (14.5). This set of equations highlights the spectral profile of a composite signal with a 50% PWM-like envelope of one OFDM symbol per PWM-like cycle. However, as previously explained the practical scenario requires more flexibility over the duty cycles for dimming control, as well as the number of OFDM symbols per PWM-like cycle associated with a larger bandwidth.

$$X_h(f) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \frac{a_{k,h}}{\sqrt{N}} T_{\mathrm{OFDM}} (-1)^k \frac{\sin(\pi(f T_{\mathrm{OFDM}} - k))}{\pi(f T_{\mathrm{OFDM}} - k)} \tag{14.6}$$

$$W_{\mathrm{d}}(f) = \frac{1}{H} \left\{ \sum_{h=1}^{H} X_h(f) e^{-j\phi_h} (-1)^h + A \frac{\sin(\pi f T_{\mathrm{OFDM}})}{\pi f T_{\mathrm{OFDM}}} \sum_{h=1,3,5,\ldots}^{i} e^{-j\phi_h} \right\} \tag{14.7}$$

where $\phi_h = 2\pi f(2h-1)\frac{T_{\mathrm{OFDM}}}{2}$. In (14.8), $W_{\mathrm{dim}}(f)$ illustrates the spectrum of an RPO-OFDM signal with more generalized design parameters, for instance, varying PWM-like duty cycles, D, and varying number of OFDM symbols per PWM-like period, T_{PWM}:

$$W_{\mathrm{dim}}(f) = W_{\mathrm{dim1}}(f) + W_{\mathrm{dim2}}(f) \tag{14.8}$$

$$W_{\mathrm{dim1}}(f) = \sum_{l=1}^{L} \left[\sum_{h=1}^{C_1} X_{l,h}(f) e^{-j[\phi_h + \phi_l + \pi]} - \sum_{h=C_1+1}^{C_2} X_{l,h}(f) e^{-j[\phi_h + \phi_l + \pi]} \right] \tag{14.9}$$

$$W_{\mathrm{dim2}}(f) = \frac{AC_1 T_{\mathrm{OFDM}}}{LT_{\mathrm{PWM}}} \frac{\sin(\pi f C_1 T_{\mathrm{OFDM}})}{\pi f C_1 T_{\mathrm{OFDM}}} e^{-j[\frac{\phi_h C_1}{(2h-1)} + \phi_l]} \tag{14.10}$$

where $\phi_l = 2\pi f(l-1)T_{\mathrm{PWM}}$, L is the number of PWM cycles per frame, $X_{l,h}(f)$ represents the spectrum of the hth OFDM symbol in the lth PWM cycle in the frame, the number of OFDM symbols per T_{PWM} is represented by $C_2 = T_{\mathrm{PWM}}/T_{\mathrm{OFDM}}$ and $C_1 = DC_2$.

In order to have a better insight on how the previous analysis could help in the signal synthesis, MCC time-domain signals and frequency-domain profiles are shown when C_2 is equal to 2, i.e., two OFDM symbols per PWM-like cycle, in Figure 14.5(a)

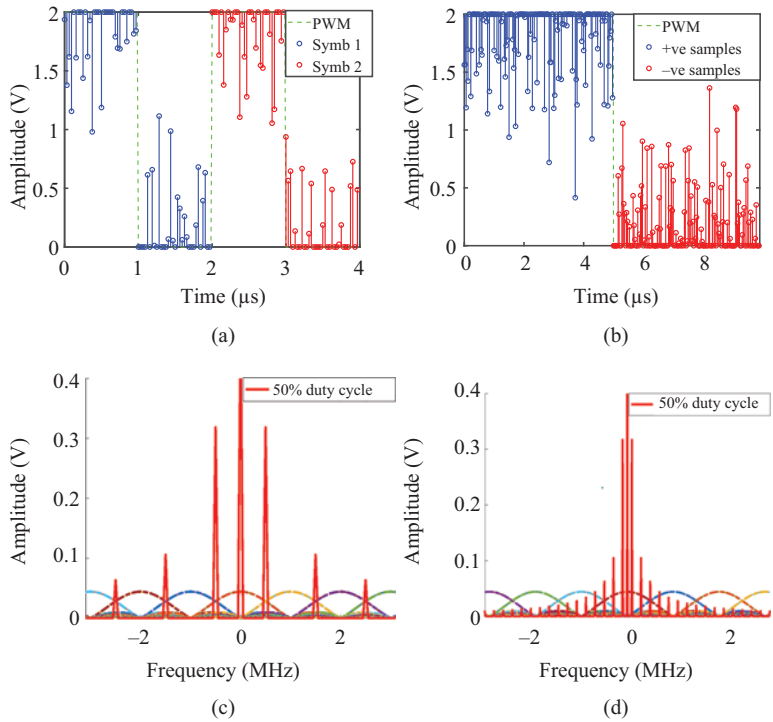


Figure 14.5 Two cycles of mixed carrier time-domain waveform with a 1 μs OFDM symbol duration and a 2 μs PWM-like envelope in (a) and one cycle showing a PWM-like envelope duration of 10 μs transmitting five OFDM symbols of 1 μs duration each in (b). The spectrum for both cases is shown in (c) and (d), respectively, illustrating nonoverlapping spectrum of OFDM subcarriers, i.e., multi-colored, and 50% PWM-like harmonics, i.e., red

and (c), and when C_2 is equal to 10 in Figure 14.5(b) and (d). The signal conditioning process differs in both the cases. For instance, in the former case each OFDM symbol with duration, i.e., T_{OFDM}, of 1 μs fits T_{PWM} of 2 μs. The positive and negative samples are conditioned to shape a PWM-like envelope of consecutive positive and negative cycles. In this case $C_2 = 2$, the orthogonality between the OFDM subcarriers and the PWM-like harmonics is not lost. This is highlighted in Figure 14.5(c), where there is no overlap between the PWM-like harmonics and the OFDM subcarriers. An interesting fact is that increasing the number of OFDM symbols per PWM-like cycle can still maintain orthogonality. This is shown in Figure 14.5(b), where the positive samples of five OFDM symbols of duration 1 μs each are shaping the ON period of a PWM-like cycle of duration 10 μs, while the negative samples are forming the OFF period. As the PWM-like cycle duration increases from 2 to 10 μs, the fundamental frequency is reduced and so as the harmonics, as shown in Figure 14.5(d). Note that

the figures are adjusted, i.e., zoomed in, to precisely show the relation between the frequency components, while the total number of subcarriers is 16.

Unlike the previous cases, Figure 14.6(a) illustrates the impact of the number of OFDM symbols per T_{PWM}, i.e., C_2, on the orthogonality loss between PWM-like harmonics and OFDM subcarriers. This is clearly represented by the overlapping frequencies in Figure 14.6(a), where $C_2 = 3$. This scenario arises from improper design considerations resulting into an overlap between PWM-like harmonics and some of the OFDM subcarriers, for instance, the odd subcarriers. However, there is no impact of varying the duty cycle, i.e., D, while maintaining an integer value for C_1, on losing the orthogonality. As seen in Figure 14.6(b), the spectral profile shows no overlap between OFDM subcarriers and PWM-like harmonics for three different cases of varying D, while $C_2 = 10$. This is clearly reflected on the energy saving, where the communication quality is still maintained, while D can be at its minimum. Both figures are properly zoomed in the same way as in Figure 14.5(c) and (d).

14.2.1.5 Performance and discussion

In this section, the performance of MCC is evaluated in terms of the reliability of the services it can provide. One of the main services that the proposed approach should not affect is the quality of illumination. Thus, the dimming control property is affected by the PWM-like envelope quality, which is directly affected by the OFDM samples. In order to evaluate and characterize the mixed carrier performance, a MATLAB® simulation model is realized, in addition to experimental validation.

In this simulation, ACO-OFDM signal format is considered [26] to demonstrate the proposed approach; however, the design can be generalized to be integrated with all other forms of unipolar OFDM, for instance DCO-OFDM. The FFT length, i.e., N, of 128 is used in the simulation with 1 MHz separation between OFDM subcarriers. The beacon and the PWM-like envelope are generated at frequencies of 20 and 100 kHz, respectively. An LOS additive white Gaussian noise (AWGN) channel is considered. In Figure 14.7, the trade-off between communication and illumination quality is investigated. The necessity for such investigation is representing how the illumination efficiency, i.e., dimming control range, is affected by increasing the communication throughput. In Figure 14.7(a), the effect of conditioning OFDM samples to form a mixed carrier envelope is illustrated. The x-axis represents the PWM-like envelope without OFDM samples, while the y-axis represents the FBR of OFDM samples conditioned to form a mixed carrier envelope. In this simulation, the beacon comprises 10% of the whole envelope in both the cases, where the effective brightness of a PWM-sampled beacon will always be 50% due to the symmetry nature of a sinusoidal waveform. This results in reducing the dimming control range, i.e., FBR, on x-axis to be between 5% and 95%. Moreover, as the power of the OFDM samples increases, the FBR range is reduced. In Figure 14.7(a), as the OFDM SNR increases to 24 dB, the FBR becomes 23% rather than 5% and the maximum FBR is reduced from 95% to 77%. This leads to reduction in the dimming control range to be 54% of the capability of the light source and reduces the illumination efficiency, as shown in Figure 14.7(b). However, reducing the SNR to 12 dB provides better control, i.e., 80%, over the dimming range as shown in the same figure. There is a clear trade-off

between the OFDM signal quality and the dimming range (i.e., SNR and dimming range are inversely proportional). The maximum achievable dimming control range depends on getting the minimum SNR for a BPSK modulation (i.e., 1-bit per symbol constellation).

However, the bit error rate (BER) performance is still maintained at varying brightness levels. This is illustrated in Figure 14.8 showing BER curves for the OFDM data stream obtained at different brightness levels (5%, 50%, and 95% FBR) with SNR ranging from 12 to 20 dB. These curves are generated for 256 kbits frame size to achieve a BER confidence level better than 95% [30]. As shown in Figure 14.8, BER values below 10^{-3} are obtained for SNR values above 18 dB, which is below

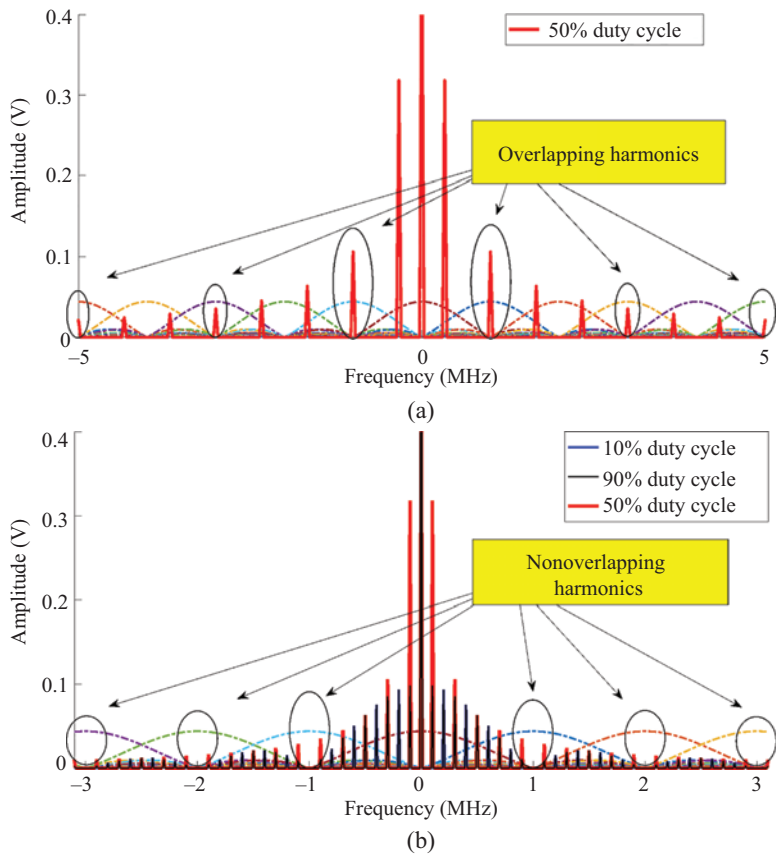


Figure 14.6 *(a) Spectrum of a 50% mixed-carrier waveform with overlapping PWM-like harmonics and odd OFDM subcarriers, i.e., $C_2 = 3$, while (b) shows the spectrum of a waveform with nonoverlapping PWM-like harmonics and OFDM subcarriers at different duty cycles, i.e., $C_2 = 10$*

the forward-error-correction (FEC) limit of 3.8×10^{-3} [31], and the dimming level has almost no effect on the bit-error performance of the system. This highlights the amount of energy saving can be achieved by reducing the brightness level to 5% while preserving the same BER performance.

The value of spectral analysis of a designed mixed carrier waveform emerges from the proper spectrum and interference management it enables. For instance, a mixed carrier waveform provides multiple services of dimming control, broadband Internet access on top of sensing and positioning applications, and thus interservice interference is to be avoided through spectrum management. Moreover, interference analysis highlights the impact of nonlinearity in front ends, resulting in an overlap between OFDM subcarriers and PWM-like harmonics of 25% duty cycle, i.e., $D = 25\%$, on signal quality and BER performance. In addition, the analysis highlights the impact of external sources of interference, i.e., surrounding light fixtures, whether being only used for lighting or incorporating more services.

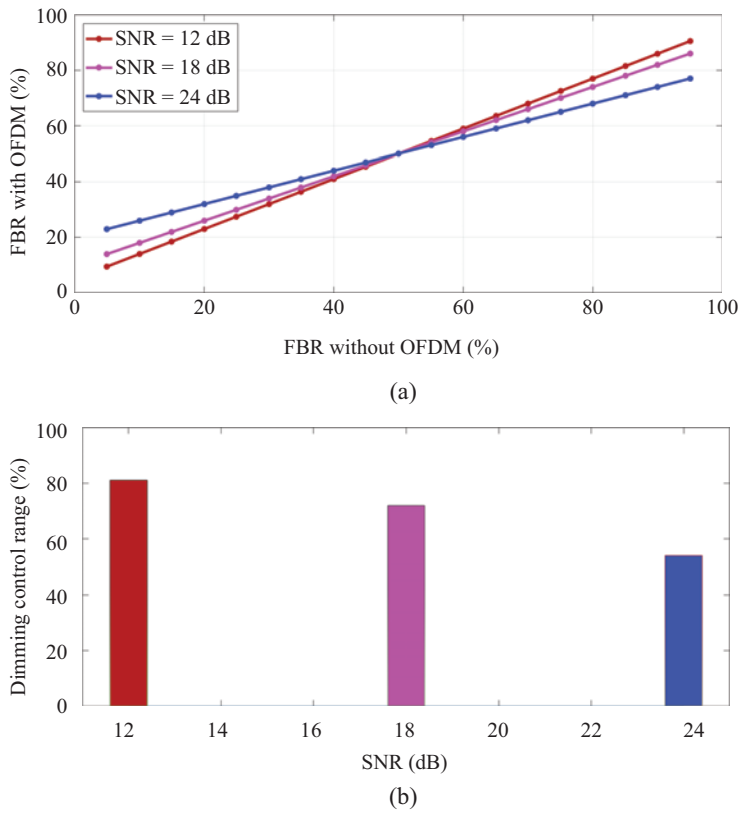


Figure 14.7 Effect of OFDM SNR on dimming control illustrating (a) FBR with vs. without OFDM and (b) dimming control range vs. SNR

In order to support the validity of the analytical work, the BER performance of a 64-QAM mixed carrier waveform is simulated under different interference scenarios in Figure 14.9. The investigated sources of interference are the ones resulting from light fixtures that offer PWM-based lighting, in addition to those that transmit beacon signals needed for sensing and positioning applications. The deviations in hardware may lead to overlap between OFDM subcarriers and PWM-like harmonics. This overlap has different impacts on the BER performance for different interference scenarios. BER for a mixed carrier waveform without overlapping harmonics and without being exposed to external interference, i.e., in red, is showing the best performance, as expected. Being exposed to an external beacon signal that interferes with the PWM-like harmonics only, i.e., no overlap between OFDM and PWM-like harmonics, a mixed carrier waveform still maintains the best BER performance. However, being exposed to the same interfering signal while PWM-like harmonics and OFDM subcarriers overlap, i.e., in gray, requires at least 5 dB increase in signal-to-interference-plus-noise ratio (SINR) in order to target BER $\geq 10^{-3}$. It is essential to emphasize that the amplitude of the interfering signal is fixed and independent from the main signal amplitude. Another source of external interference on mixed carrier waveform with overlapping harmonics is the surrounding light fixtures only used for illumination, i.e., PWM dimming control. For the blue curve, the PWM frequency of the interfering waveform is 500 kHz, which is the same as the PWM-like envelope of the mixed carrier waveform. However, D of the former is 25%, while the same of the latter is 50%. The BER performance is comparable to the scenario of not being subject to external interference. However, varying the PWM frequency of the interfering signal results in more degradation of the BER performance, as shown by the

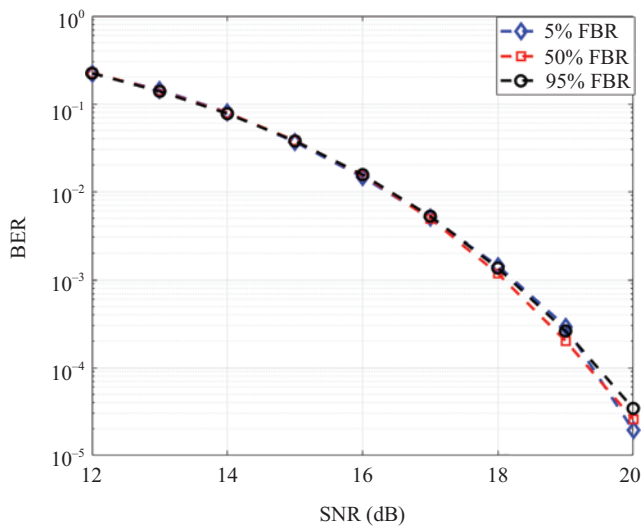


Figure 14.8 BER performance of the proposed system using 16-QAM

magenta curve. A gap of 2–3 dB, from the no interference scenario, is obvious at BER of $\geq 10^{-3}$. The constellations in Figure 14.9(b) and (c) represent the received signal in the same interference scenario, i.e., a random beacon signal with the same frequency as one of the OFDM subcarriers, at SINR of 26 dB, with and without harmonics overlap, consecutively. The interference effect is observed on the real axis of the constellation due to the real nature of the interfering signal.

14.2.2 Lightweight MCC

In this section, a new concept for modulating multiple light sources to realize a lightweight version of OFDM communication chain is discussed. In the proposed

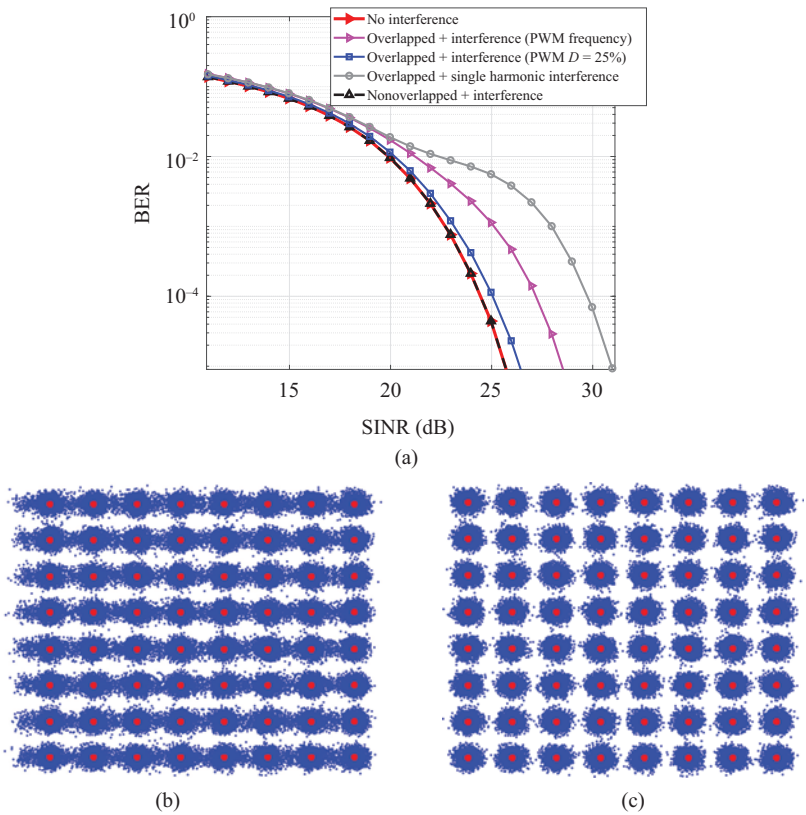


Figure 14.9 (a) BER performance for 64-QAM mixed-carrier waveform under different interference scenarios, whereas (b) and (c), respectively, show received constellations for a received mixed-carrier waveform exposed to an interfering beacon signal at SINR of 26 dB, with and without harmonics overlap, i.e., $C_2 = 3$ and $C_2 = 10$

scheme, different SCM streams are synchronously transmitted from a cluster of different LEDs carrying the encoded OFDM time-domain samples, thus enabling the realization of the Fourier transformation in the optical domain. Accordingly, the FFT operation required for the demodulation at the receiver side is eliminated, which is crucial for resource-limited IoT devices. Moreover, this scheme enables the MCC vision by generating sinusoidal streams that are responsible for SCM, while other sinusoidal waveforms are added up over-the-channel to realize the subcarriers of a wideband MCM. Thus, the proposed scheme enables parallel transmission of high-speed and low-speed streams to serve multiple device capability. Finally, the proposed concept (1) offers the same spectral efficiency as ACO-OFDM, (2) reduces the bandwidth requirement of individual light sources, and (3) reduces the PAPR of the signal formed and transmitted over the optical channel.

14.2.2.1 FFT-less concept

The structure of a transceiver supporting the proposed lightweight optical OFDM (LwO-OFDM) system is shown in Figure 14.10. For the transmitter, comparing with the conventional ACO-OFDM system using a single LED, the building blocks to generate the time-domain samples $X[n]$ are identical (i.e., in blue). After the IFFT block, due to the anti-symmetry of an ACO-OFDM signal, the proposed LwO-OFDM only requires the transmission of the first $\frac{N}{2}$ time-domain samples. These samples are used to scale the amplitudes of different sinusoidal signals. The frequencies of these signals are set based on the subcarrier frequencies and subcarrier separation of the system. These sinusoidal signals are used to modulate the intensities of an array of LEDs. To transmit a single OFDM symbol of period T, the $\frac{N}{2}$ samples are transmitted over two symbols, where $X_a(t)$ is transmitted at T_a and $X_b(t)$ is transmitted at T_b. During T_a and T_b (i.e., normally equal) samples are carried over cosine and sine signals, respectively. A 180° is added to ensure the proper integration of the negative sign before the imaginary component as shown later in the following analysis in (14.13). A bank of $\frac{N}{2} - 1$ oscillators is phase controlled by a 270° phase-shift controller to generate cosine and negative sine signals with proportional amplitudes to the IFFT time vector $x[n]$. There is no need for an extra oscillator to represent the 0th subcarrier that represents the DC-bias of the signal, as it can be replaced by proper DC-biasing of the remaining LEDs. Each sinusoidal waveform modulates one of the LEDs forming the array and the optical power from all LEDs is summed on air. As the signal is generated with the previous considerations, the summation process resembles the FFT process in conventional ACO-OFDM receivers (i.e., FFT-on-air). Unlike RF, VLC channel does not experience severe multipath fading as the size of the photo detectors used in the receivers is typically millions of square wavelengths. Therefore, as shown in the receiver block diagram, the equalization process is not highlighted to reduce the system complexity [32]. In order to have better insight over the transmitted signal shape, the digital samples $x[n]$ after an IFFT of length N on the transmitter side can be represented by (14.11).

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k]\, e^{\frac{j2\pi kn}{N}}, \qquad \text{for } 0 \le n \le N-1 \tag{14.11}$$

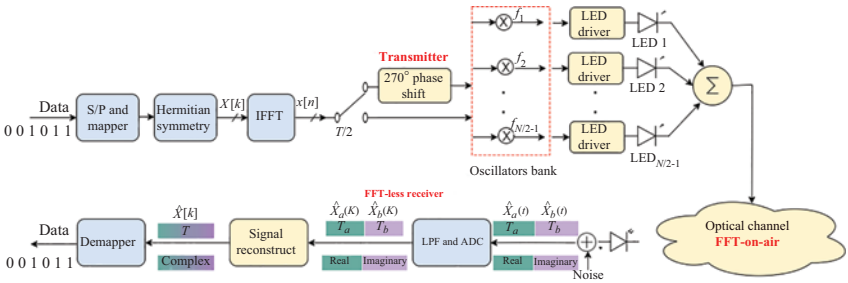


Figure 14.10 Proposed LwO-OFDM block diagram with common blocks of ACO-OFDM in blue and presenting FFT-on-air process and FFT-less receiver design

where the first $\frac{N}{2}$ samples of the vector X_k represent the output samples of the mapper (i.e., QAM-modulator) and the zero-forced odd subcarriers as in conventional ACO-OFDM. The following $\frac{N}{2}$ samples are complex conjugates of the first $\frac{N}{2}$ samples in order to ensure real (i.e., noncomplex)-valued IFFT output samples. Conventionally, an OFDM receiver needs to perform FFT to get IQ samples for a QAM demodulator. The $\hat{X}[k]$ vector is obtained from the FFT process shown in the following equation:

$$\hat{X}[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n]\, e^{\frac{-j2\pi kn}{N}}, \qquad \text{for } 0 \le k \le N-1 \tag{14.12}$$

In this context, $\hat{X}[k]$ can be rewritten as

$$\hat{X}[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n]\left(\cos\frac{2\pi kn}{N} - j\sin\frac{2\pi kn}{N}\right), \quad \text{for } 0 \le k \le N-1 \tag{14.13}$$

As clearly shown by (14.13), using the bank of oscillators in LwO-OFDM enables applying FFT-on-air. This enables the reception of an OFDM signal with a simple FFT-less receiver. The transmitted signals follow what is indicted in Figure 14.10, where analog signals $\hat{X}_a(t)$ and $\hat{X}_b(t)$ of durations T_a and T_b are detected by the PD to represent the transmitted real and imaginary coefficients of $\hat{X}[k]$ (i.e., should be detected in a conventional ACO-OFDM system), respectively. These analog signals are expressed in the following equations:

$$\hat{X}_a(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{\frac{N}{2}-1} x[n]\cos(2\pi f_n t), \qquad \text{for } 0 \le t \le \frac{T}{2} \tag{14.14}$$

$$\hat{X}_b(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{\frac{N}{2}-1} -x[n]\sin(2\pi f_n t), \qquad \text{for } \frac{T}{2} \le t \le T \tag{14.15}$$

where T is the LwO-OFDM symbol duration and T_a and T_b are both equal to $\frac{T}{2}$. The received signals are then digitized as in a conventional ACO-OFDM receiver using

an ADC to convert $\hat{X}_a(t)$ and $\hat{X}_b(t)$ into $\hat{X}_a[k]$ and $\hat{X}_b[k]$, respectively. After that, $\hat{X}[k]$ is reconstructed as in the following equation:

$$\hat{X}[k] = \hat{X}_a[k] + j\,\hat{X}_b[k], \qquad \text{for } 0 \leq k \leq N-1 \tag{14.16}$$

FFT-on-air process can be shown as described in Figure 14.11. From the previous equations, it is noted that LwO-OFDM is designed to have a symbol duration T. However, in order to transmit the real and imaginary coefficients on two different time slots, the total symbol time should be doubled that of a conventional ACO-OFDM. This will reduce the spectral efficiency to half of ACO-OFDM. An interesting property of the proposed scheme can be illustrated in Figure 14.11(a) and (b). In Figure 14.11(a), the obtained digital signal (i.e., $\hat{X}_a[k]$) from the spatial summation of the sinusoidal (i.e., cosine) waveforms is symmetric. In other words, the second $\frac{N}{2}$ samples are the image of the first $\frac{N}{2}$. This is also the case in Figure 14.11(b) but with the image inverted. In fact, it is waste of resources to send these redundant images, so the technique is adjusted to send only half of the time samples as in Figure 14.11(c) and (d). As a result, the whole symbol duration of LwO-OFDM is the same as a conventional ACO-OFDM.

As discussed in Section 14.2, an ACO-OFDM signal has an anti-symmetry shape. In a conventional ACO-OFDM system, the redundant negative values are clipped in order to obtain higher power efficiency. This is illustrated in Figure 14.12(a) in blue. In LwO-OFDM, the oscillators are phase controlled that enable the generation of

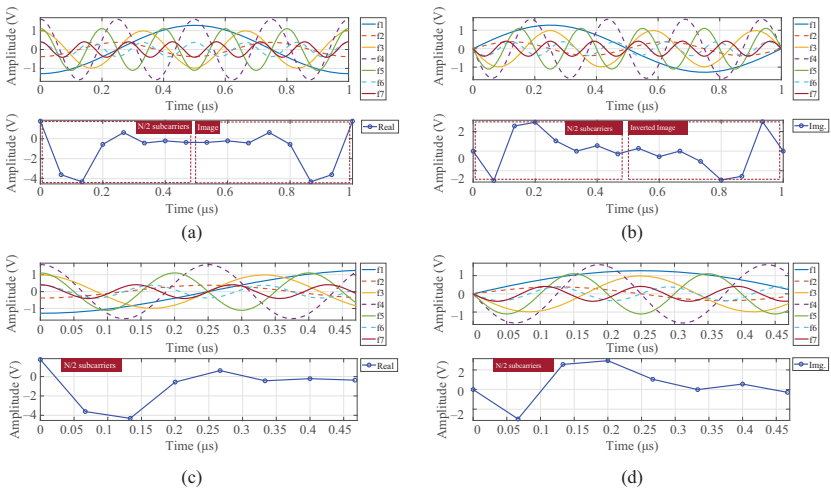


Figure 14.11 LwO-OFDM sinusoidal waveforms summation in optical domain with full symbol duration (i.e., T) in the first rows of (a) and (b), and half duration (i.e., $\frac{T}{2}$) in the first rows of (c) and (d). In the second rows of (a) and (b), the digitized vectors $\hat{X}_a[k]$ and $\hat{X}_b[k]$ are represented with full duration (i.e., T) but half duration (i.e., $\frac{T}{2}$) in (c) and (d)

sinusoidal waveforms representing the negative values from the IFFT, as clear in Figure 14.12(a) in red. Therefore, the zero-clipping process can be eliminated and replaced by sending half of the subcarrier samples only (i.e., $\frac{N}{2}$) to avoid redundancy. This has a direct benefit on the consumed bandwidth by each single LED, which becomes half of the needed spectrum by an LED transmitting a conventional ACO-OFDM signal that carries the same data. This is clearly presented in Figure 14.12(b), with a simulated LwO-OFDM consuming 4 MHz, while ACO-OFDM that carries the same data consumes 8 MHz. The simulation parameters to generate this curve are detailed in the following section.

From the earlier, it is clear that LwO-OFDM transmits one symbol over two time slots (i.e., T_a and T_b) each with half the number of IFFT coefficients (i.e., $\frac{N}{2}$). Interestingly, this results in obtaining a symbol duration equal to that of an ACO-OFDM and the same number of active and total subcarriers, as well (i.e., $\frac{N}{4}$ out of N). As expected, this results in the same spectral efficiency of a conventional ACO-OFDM as shown in (14.17). Although the cyclic prefix is not discussed, it is still a valid system parameter:

$$R_{\text{LwO}} = \left(\frac{\frac{N}{4}}{N + N_{cyc}}\right)\log_2 M \quad \text{bits/s/Hz} \tag{14.17}$$

14.2.2.2 Performance evaluation

A simulation model is realized in MATLAB to demonstrate the performance of the proposed scheme compared to the conventional ACO-OFDM systems. However, the proposed scheme can be adopted using other forms of optical OFDM techniques. An IFFT of length ($N = 16$) is exploited in the simulation model with a subcarrier separation of 1 MHz. The sampling frequency (f_s) complies with the Nyquist rate (i.e., double the maximum frequency) with a value of 16 MHz. On account of the fact that the complex conjugate amplitudes after the IFFT are not transmitted, only seven light sources and oscillators (i.e., $\frac{N}{2} - 1$) are needed on the transmitter side. This number can be increased to eight, while the first LED will only be responsible for DC-bias

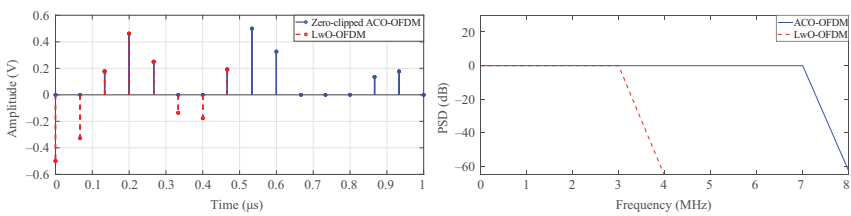


Figure 14.12 (a) Time-domain coefficients of zero-clipped conventional ACO-OFDM (i.e., in blue) and LwO-OFDM subcarriers (i.e., in red), and (b) reduced bandwidth requirements for LwO-OFDM compared to conventional ACO-OFDM due to sending half of the subcarriers ($\frac{N}{2}$)

Table 14.1 Simulation and system design parameters

System design specifications	
OFDM technique	ACO-OFDM
IFFT length (N)	16
Subcarrier separation	1 MHz
Sampling frequency (f_s)	16 MHz
QAM order	32, 64, 128, 256
Symbol duration (T)	1 μ
Number of oscillators/light sources	7
Simulation parameters	
LED dynamic range	0.5 V
Noise level	−10 dBm
SNR range	10:40 dB
Number of symbols	10,000

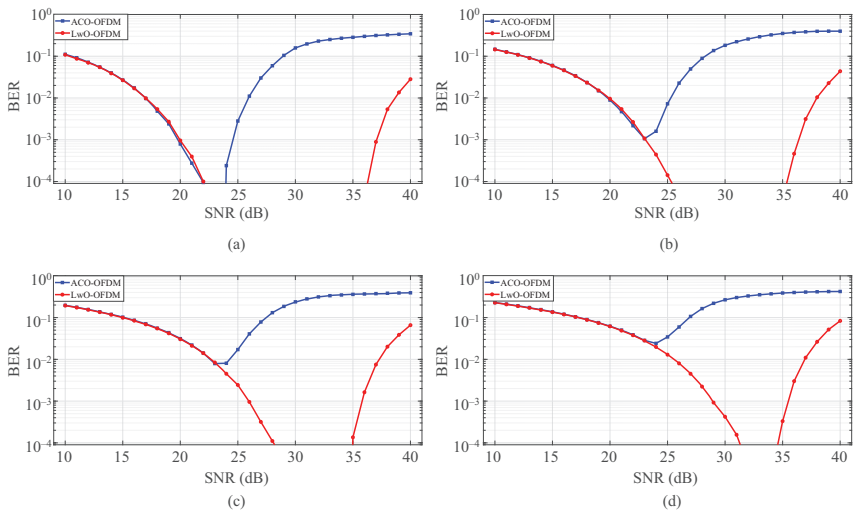


Figure 14.13 BER of conventional ACO-OFDM (i.e., blue) and LwO-OFDM (i.e., red) at different SNR levels and showing clipping effects on the signal performance for (a) 32-QAM, (b) 64-QAM, (c) 128-QAM, and (d) 256-QAM

(i.e., 0th subcarrier), but in our case it can be replaced by biasing all the light sources in the LED array with the appropriate DC-bias. BER is simulated for conventional ACO-OFDM and LwO-OFDM using different modulation orders (32–256 QAM) in order to study the effects of SNR and clipping on the transmitter side on the system performance. Higher QAM orders are targeted to compensate for the low spectral efficiency. Table 14.1 summarizes the simulation and system design parameters.

In Figure 14.13, there is a clear enhancement in BER performance when LwO-OFDM is compared to conventional ACO-OFDM. For low SNR values till about 23 dB, the LwO-OFDM performance is nearly the same as the ACO-OFDM. For a fair comparison, the power of an ACO-OFDM symbol is divided on both real and imaginary spatially summed symbols of LwO-OFDM. As expected, when the QAM order increases the performance degrades and higher SNR levels are needed to achieve the same BER as in lower QAM orders. Both curves of LwO-OFDM and ACO-OFDM approach a BER of nearly 10^{-4} at a range of 23–32 dB for 32–256 QAM orders. The regions where the performance of both systems is matching are the regions where there is nearly no clipping or clipping effects on the transmitted signals is negligible. As the SNR exceeds 23 dB, signal clipping starts to degrade the performance due to the increased signal power above the LEDs dynamic range. Unlike LwO-OFDM, the clipping effect is severe and quick for ACO-OFDM signals due to relatively higher PAPR. For instance, there is a gain of 10–12 dB in different modulation orders for the favor of the proposed scheme. There is a good agreement between the obtained results and the signal generation procedure for both techniques. In an ACO-OFDM system, a single LED emits an OFDM symbol with high PAPR characteristics, while in LwO-OFDM, multiple LEDs are responsible for generating sinusoidal waveforms with fixed PAPR of nearly 3 dB that sum up on air. These results indicate that LwO-OFDM signal is more resilient to light source clipping. The PAPR comparison between both techniques is given by the complementary cumulative distribution function (CCDF) of PAPR that denotes the probability that the emitted signals from LEDs exceed a certain threshold PAPR_0 and is given by

$$\text{CCDF} = P(\text{PAPR} > \text{PAPR}_0) \tag{14.18}$$

where a general expression for PAPR of a transmitted signal $x[n]$ can be deduced from:

$$\text{PAPR} = \frac{\max|x[n]|^2}{E[|x[n]|^2]} \tag{14.19}$$

As previously stated, the LwO-OFDM symbol is mainly formed from the summation of the optical power of sinusoidal waveforms on air. Therefore, in (14.18) $x[n]$ is replaced by $x[n]\cos(2\pi f_n t + \phi)$. As a result, the proposed scheme shows 5 dB PAPR reduction (i.e., fixed PAPR of nearly 3 dB) compared to ACO-OFDM at CCDF of 10^{-2}, which is clearly illustrated in Figure 14.14.

14.3 State-of-the-art VLC topics

As previously mentioned, LEDs are energy-efficient light sources, and thus, using VLC can result in designing communication systems with high energy efficiency [33]. The high energy efficiency is achieved by realizing illumination and data transmission at the same time. LED technology has been significantly successful in terms of market share not only because of its highly energy efficiency, but also because it has a long operational lifetime and it enables a more comfortable lighting system for

humans [34]. For these reasons, LED lighting is expected to dominate future lighting markets, and VLC systems can become a de facto solution for indoor and short-range communications, as the energy used for the communication would not increase the energy costs [35]. With the growing interest in VLC systems, new research topics have emerged in recent years. This section is dedicated to highlighting some of the state-of-the-art topics discussed in VLC research and presents the benefits and pitfalls of the techniques presented in literature.

14.3.1 Security of coexistence with RF technologies

Instead of replacing the legacy RF technology entirely, other technologies can be used concurrently with RF to meet the needs of present and future applications. Heterogeneous networks, or HetNets for short, are networks that adopt two, or more, technologies. Thus, a diverse span of the spectrum can be utilized to provide high quality-of-service for network users. Employing multiple radio interfaces in communication systems has proven to enhance energy efficiency in RF system [36]. VLC, due to its high energy efficiency, is one of the most promising technologies to be integrated with preexisting RF-based networks. The integration of both technologies improves the system's overall energy consumption, reliability, and coverage area. Moreover, as the available spectrum increases, contention decreases and the stringent requirements of the medium access protocols can be relaxed; hence, unprecedented data rates can be offered [37]. Thus, attention needs to be forwarded toward common protocols designed for these integrated systems. One of the interesting research topics in hybrid RF/VLC systems is security.

Due to the nature of wireless radio propagation, the PHY, i.e., the lowest layer of the protocol stack, is vulnerable to eavesdroppers [38]. However, security primitives are usually implemented in the upper layers of the protocol stack, while lower layer security functions are usually ignored and are system dependent. Thus, applications that send sensitive and private data on-the-fly such as health-care sensor nodes, which cannot perform the cryptographic methodologies' expensive mathematical calculations, suffer and are prone to data hacking. Moreover, attackers can exploit the medium

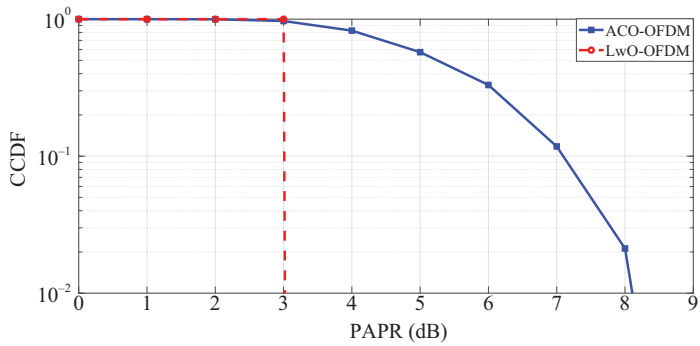


Figure 14.14 PAPR CCDF of conventional ACO-OFDM (i.e., blue) and LwO-OFDM (i.e., red)

access control header, which is sent in plain text, and reveal side-channel information, including source and destination addresses, the mapping scheme, and synchronization information, which can be used to perform various attacks [39]. Consequently, PHY security is an emanating requirement for wireless communication systems.

There are numerous research efforts dedicated for PHY security enhancements in RF-based systems, and they can be divided into two approaches: physical-layer security (PLS) and physical-layer encryption (PLE). PLS techniques rely on the unpredictability of the wireless multipath channel to defend the transmission [40]. However, unlike the RF channel, the optical channel is not rich scattering; hence, it lacks the unpredictability criterion. On the other hand, PLE schemes aim to protect the entire PHY packet by encrypting the data flow in the PHY modulation stages [41]. Then, decryption on the receiver's end is performed using a pre-shared key between the transmitter and the receiver. All cryptographic measures rely on the assumption that it is computationally infeasible for an eavesdropper to decipher the data without secret key knowledge, which is a statement that has not yet been mathematically proven if the eavesdropper has sufficient time and computation resources [42]. For VLC, approaches such as null steering and artificial noise insertion to combat eavesdropping were investigated in [43,44]. Eavesdropping mitigation using cooperative jamming is another popular approach, adopted in [45], among others. In [46], a random time reversal scheme that makes the signal focused on the legitimate receiver while interfering with the eavesdropper's channel with VLC multipath redundancy, time reversal, and random choice technique was presented. Key-based approaches were also investigated such as the work done in [47].

Despite of aforementioned techniques, there is a lack in research that is dedicated for securing hybrid RF/VLC transmission. The subsequent subsection depicts the challenges of designing a technology-independent technique using a state-of-the-art physical-layer (PHY) security technique known as security-aware OFDM (SA-OFDM) [48] as an example. SA-OFDM combines a PHY key-based approach, i.e., symbol-level encryption key, along with a novel keyless approach that artificially degrades the eavesdropper's channel by redesigning the time-domain samples of the OFDM symbols, which causes the signal characteristics to be completely masked. Thus, even if the eavesdropper is able to intercept the PHY encryption key by exhaustive computation, a distorted signal will be received instead, and the secrecy of the transmission is preserved. However, as will be discussed in the forthcoming subsections, each technology has its own unique characteristics which makes designing a technology-independent technique troublesome. SA-OFDM offers a unified framework with only minor modifications required in data processing; however, the transceiver architecture is common and only the front ends are varied based on the utilized technology. Before discussing the details of SA-OFDM, a brief overview of OFDM in VLC is presented.

14.3.1.1 OFDM in VLC

The de facto technique utilized in most recent wireless standards, i.e., IEEE 802.11 protocol suite and LTE, is OFDM. OFDM is used due to its high spectral efficiency and robustness to inter-symbol interference, as well as fading. Thus, the analysis

focuses on OFDM due to its wide adoption; however, the proposed scheme is generic and can be applied to non-OFDM systems. The principle of OFDM is modulating data sequences by applying N-point IFFT onto a series of orthogonal subcarriers; producing N complex IQ samples corresponding to N subcarriers. Conventionally, an N-point complex baseband time-domain OFDM symbol can be represented as

$$\begin{aligned} x_n &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} X_i e^{j2\pi \frac{ni}{N}} \\ &= Re\left\{ \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} X_i e^{j2\pi \frac{ni}{N}} \right\} + jIm\left\{ \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} X_i e^{j2\pi \frac{ni}{N}} \right\} \\ &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \{(\Re_i + j\Im_i)(\cos(ni\theta_N) + j\sin(ni\theta_N))\} \end{aligned} \tag{14.20}$$

where n is the time index of the signal and X_i is the ith subcarrier-modulated signal and $\theta_N = 2\pi/N$. Additionally, X_i can be expressed as $\Re_i$ and $\Im_i$ representing its real and imaginary parts, respectively, and $e^{j2\pi \frac{ni}{N}}$ as $\cos(ni\theta_N) + j\sin(ni\theta_N)$. Unlike RF systems, VLC systems mostly use IM/DD. The baseband signal, $x(t)$, does not modulate the amplitude and phase of the optical carrier; it modulates its intensity instead. Thus, $x(t)$ must be positive and real. Conventional OFDM techniques cannot be directly used in VLC, instead, at the transmitter's side of IM/DD-based OFDM systems, the input vector to IFFT is designed to have Hermitian symmetry. As a result; the time-domain OFDM signal becomes real. Then, it becomes positive either by adding a DC bias, as in DCO-OFDM, or by clipping at zero, as in ACO-OFDM. In both schemes, the DC and the $N/2$ subcarriers are set to zero to ensure that the output is real [1]. Moreover, in ACO-OFDM, the data is carried only on the odd subcarriers; thus, only $N/4 - 1$ input values are fed to the N point IFFT in ACO-OFDM. On the other hand, DCO-OFDM has $N/2 - 1$ inputs, making it more spectrally efficient than ACO-OFDM. Thus, DCO-OFDM is adopted due to its spectral efficiency for the optical link, while conventional OFDM is used for the RF link.

14.3.1.2 SA-OFDM transmission

First, the time-domain samples of the OFDM symbols, i.e., after IFFT, are divided into groups. The amplitude of the IQ samples are manipulated by a parameter α. This operation is equivalent to performing a nonlinear masking of the information symbols in the frequency domain. Then, the modified samples are scrambled based on a pre-shared key. Thus, the transmitted signal will be significantly varied over the frequency spectrum and will no longer exhibit OFDM characteristics.

The OFDM symbol in (14.20) can also be represented as follows:

$$x^T = \left(F_{\cos}^{-1}\Re^T - F_{\sin}^{-1}\Im^T\right) + j\left(F_{\sin}^{-1}\Re^T + F_{\cos}^{-1}\Im^T\right) \tag{14.21}$$

where $F_{\cos}^{-1}$ equals:

$$\frac{1}{\sqrt{N}}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \cos(\theta_N) & \dots & \cos((N-1)\theta_N) \\ 1 & \cos(2\theta_N) & \dots & \cos(2(N-1)\theta_N) \\ \vdots & \vdots & & \vdots \\ 1 & \cos(i\theta_N) & \dots & \cos(i(N-1)\theta_N) \\ \vdots & \vdots & & \vdots \\ 1 & \cos((N-1)\theta_N) & \dots & \cos((N-1)^2\theta_N) \end{pmatrix} \tag{14.22}$$

with $[\cdot]^T$ denoting the transpose operation and $F_{\sin}^{-1}$ is identical to $F_{\cos}^{-1}$ but the cosine functions are replaced by sines. An SA-OFDM-modified signal, s, can be presented by

$$s^T = \left(F_{\cos}^{-1}\Re^T - F_{\sin}^{-1}\Im^T + \alpha a^T\right) + j\left(F_{\sin}^{-1}\Re^T + F_{\cos}^{-1}\Im^T + \alpha b^T\right) \tag{14.23}$$

where a and b are two vectors of length N, and their entries are equiprobable and $\in [-1, 1]$ to average the signal power increase. The parameter α is adjusted to maintain the PAPR of the OFDM symbol while preserving the security gain. Then, using an $N \times N$ scrambling matrix, M_Z, s is scrambled; mapping the ith element into the jth location. The scrambling function Z_n of N elements determines the scrambling sequence. The scrambling matrix can be expressed as $M_Z^T = \left(e_{Z_0}, e_{Z_1}, \dots, e_{Z_{N-1}}\right)$, where e_{Z_n} denotes an N-dimension row vector with only one element equal to one in the designated position and zero otherwise. An SA-OFDM transmitted signal, $\tilde{s}$, can be presented by its transpose as $\tilde{s}^T = M_Z s^T$. Thus:

$$\tilde{s}^T = M_Z\left[\left(F_{\cos}^{-1}\Re^T - F_{\sin}^{-1}\Im^T + \alpha a^T\right) + j\left(F_{\sin}^{-1}\Re^T + F_{\cos}^{-1}\Im^T + \alpha b^T\right)\right] \tag{14.24}$$

The block diagram of the proposed system is shown in Figure 14.15, showing the stages to generate $\tilde{s}$. For RF transmission, the front ends are antennas on the transmitting and receiving sides, while in VLC the transmitting front end is a light source and the receiving front end is a photodetector.

14.3.1.3 SA-OFDM reception

At the receiving end, since the legitimate user has the descrambling sequence based on the pre-shared key, the scrambling operation can be reverted. Then, using an optimal

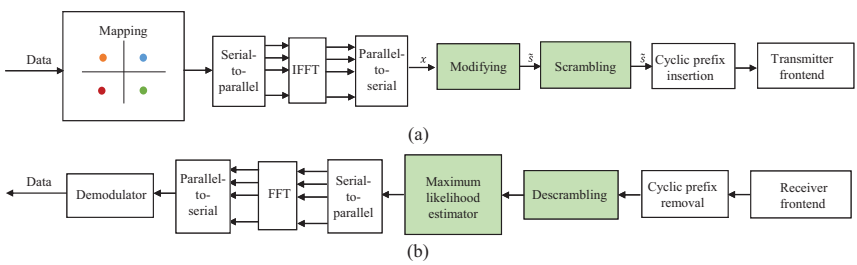


Figure 14.15 SA-OFDM system block diagram with (a) transmitter and (b) receiver

detector, i.e., the maximum likelihood (ML) decoding, the variables a and b are estimated per OFDM symbol restoring back the signal to its original time-domain form. The proceeding steps are common for any OFDM receiver, the frequency-domain signal, X, is obtained after the FFT operation. Then, the signal is passed to QAM demodulator and the transmitted data can be obtained, as depicted in Figure 14.16(b). On the other hand, if the eavesdropper does not know the key, i.e., lacks the knowledge of the scrambling matrix M_z, FFT will be directly performed on the received signal to recover the message $\hat{S}$:

$$\hat{S}^T = (F_{\cos} + jF_{\sin})\tilde{s}^T = M_Z(F_{\cos} + jF_{\sin})s^T \tag{14.25}$$

where $F_{\cos} = F_{\cos}^{-1} = \cos(ik\theta_N)_{0\le i,k<N}$ and $F_{\sin} = -F_{\sin}^{-1} = \sin(-ik\theta_N)_{0\le i,k<N}$. Then,

$$\begin{aligned}\hat{S}^T &= M_Z(F_{\cos} + jF_{\sin}) \\ &\quad \times \left[\left(F_{\cos}^{-1}\Re^T - F_{\sin}^{-1}\Im^T + \alpha a^T\right) + j\left(F_{\sin}^{-1}\Re^T + F_{\cos}^{-1}\Im^T + \alpha b^T\right) \right]\end{aligned} \tag{14.26}$$

can be simplified as

$$\begin{aligned}\hat{S}^T &= M_Z[F_{\cos}^2\Re^T + F_{\sin}^2\Re^T + jF_{\cos}^2\Im^T + jF_{\sin}^2\Im^T + \alpha F_{\cos}a^T + \alpha F_{\cos}b \\ &\quad + j\alpha F_{\sin}a^T - \alpha F_{\sin}b^T] \\ &= M_Z[\Re^T + j\Im^T + \alpha F_{\cos}(a+b)^T + \alpha F_{\sin}(ja-b)^T] \\ &= M_Z[X^T + \alpha F_{\cos}(a+b)^T + \alpha F_{\sin}(ja-b)^T]\end{aligned} \tag{14.27}$$

given that $F_{\cos}F_{\sin} = 0$ and $F_{\cos}^2 + F_{\sin}^2 = I$ where I is an $N \times N$ identity matrix. As observed in (14.27), the modification and scrambling operations cause constellation rotation when viewed in the frequency domain. As a result, the eavesdropper's constellation is significantly distorted even at an SNR of 60 dB as shown in Figure14.16(b). Even if the eavesdropper is able to obtain the scrambling key, the constellation will be distorted by the components $\alpha F_{\cos}(a+b)^T + \alpha F_{\sin}(ja-b)^T$, causing the constellations given in Figure14.16(c). The derivations listed earlier are analogous for VLC transmission, with the difference of the lack of the imaginary component and hence the constellations in VLC will only vary on the real axis as portrayed in Figure14.16(d).

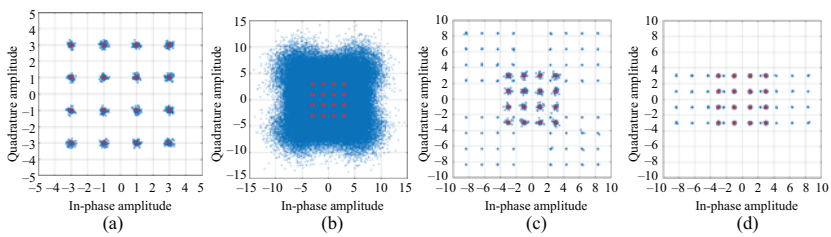


Figure 14.16 Constellations of (a) the legitimate user and (b) eavesdropper with scrambling, while (c) eavesdropper without scrambling for RF and (d) eavesdropper without scrambling in VLC at 16-QAM with SNR = 60 dB and under Rayleigh channel for RF and AWGN for VLC

While the legitimate receiver, based on the system design, has the ability to accurately estimate the parameters a and b, the eavesdropper requires 2^{2N} operations per OFDM symbol to guess them. For most wireless standards, the least number of OFDM subcarriers is 64, i.e., $N = 64$, which means that the eavesdropper requires 3.4×10^{38} operations per OFDM symbol for every possible value of α in order to restore the signal back to its original form in the time-domain.

14.3.1.4 SA-OFDM performance

SA-OFDM is first tested for RF transmission under flat fading Rayleigh channel model with added AWGN noise, the results can be observed in Figure14.17. The secrecy BER, which is the BER of the eavesdropper under the proposed technique, is evaluated given the premises of the eavesdropper having knowledge of the secret key and without. The eavesdropper's BER without key-knowledge has a constant value of 0.5, i.e., equivalent to random guessing. However, even if the key is intercepted, assuming the signal is normalized to have a maximum of 1 and at $\alpha = 0.05$, the eavesdropper's BER saturates within the range of 10^{-2}. Since, this BER performance is below the threshold for FEC, the security of the system can be maintained. On the other hand, the legitimate user's BER remains unaffected and is consistent with the simulated BER of conventional OFDM, as depicted in Figure 14.17(a). As another metric to quantify the security gain of the system, secrecy capacity is used [42]. Secrecy capacity, C_s, can be defined as the difference between the channel capacities of the legitimate user C_u and the eavesdropper C_e as

$$C_s = C_u - C_e = \log\left(1 + \frac{P}{N_u}\right) - \log\left(1 + \frac{P}{N_e}\right) \tag{14.28}$$

where N_u and N_e are the user's and eavesdropper noise power, respectively, and P is the signal power. The eavesdropper's channel is degraded by the signal manipulation process described in the preceding subsections, which can be seen as artificial noise insertion; thus, it becomes of lower quality when compared to that of the legitimate user. For an OFDM system, it can be redefined as

$$C_s = \sum_{n=0}^{N-1} \left(\log\left(1 + \frac{P_n}{N_{nu}}\right) - \log\left(1 + \frac{P_n}{N_{ne}}\right)\right) \tag{14.29}$$

where $\frac{P_n}{N_{nu}}$ and $\frac{P_n}{N_{ne}}$ are the user and eavesdropper SNR on the nth subcarrier, respectively. By definition, a positive C_s denotes that the system is able to avoid the data being intercepted by the eavesdropper. The PHY security gain of SA-OFDM is also depicted in Figure 14.17(b), showing the secrecy capacity of SA-OFDM with and without the eavesdropper's knowledge of the utilized key. SA-OFDM excels without key knowledge, however, even with key knowledge the secrecy capacity always yields a positive value which is intuitively proportional with SNR.

The performance of the SA-OFDM is also evaluated for the VLC link. To quantify the effect of the parameter α on the eavesdropper's BER performance, Figure 14.18(a) shows the BER performance vs. SNR for the (VLC) link at 16-QAM DCO-OFDM at different values of α. As shown, even with key knowledge, the BER of the eavesdropper is less than 10^{-2} and is proportional to α, consistent with the results

obtained in the RF link. In Figure 14.18(a), the eavesdropper's BER at $\alpha = 0.15$ is 1.3×10^{-2}, while that at $\alpha = 0.05$ is 0.4×10^{-2} at SNR = 24 dBm, where the BER remains constant. Given the eavesdropper has key knowledge, the system's secrecy capacity is calculated for the VLC link and is presented in Figure 14.18(b). Intuitively, the secrecy capacity C_s improves with α. Yet, even at $\alpha = 0.05$ the secrecy capacity persistently has a positive value, which naturally increases as the SNR value increases. In summary, efficient techniques can be proposed for deployment in multiple technologies; however, the techniques have to be designed in a manner that accommodates the specific traits of the utilized technologies.

14.3.2 Augmented MIMO in VLC

MIMO techniques are one of the most renowned methods for multiplying the capacity of a communication system using multiple transmission and receiving elements to

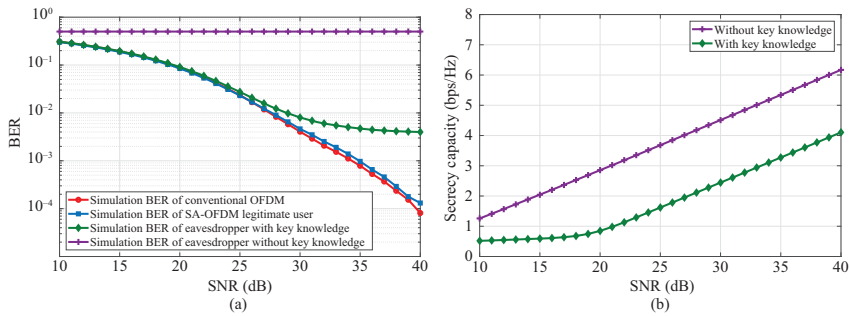


Figure 14.17 Performance of the 16-QAM RF link under a flat fading Rayleigh channel with a maximum doppler shift of 3 Hz, a delay spread of 20 ns, and a channel bandwidth of 20 MHz: (a) BER and (b) secrecy capacity

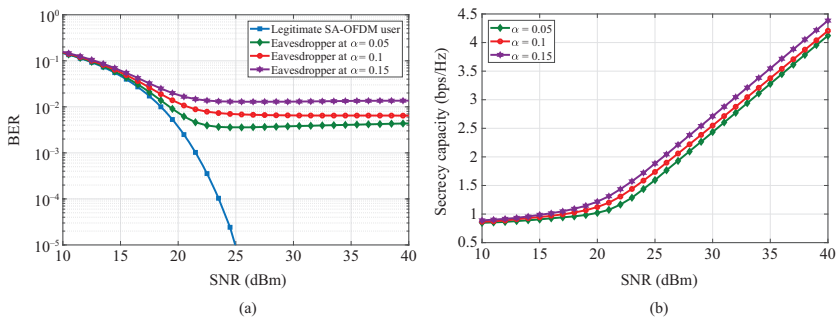


Figure 14.18 Performance of the VLC link for various α values for 16-QAM DCO-OFDM under AWGN channel model assuming the eavesdropper has key knowledge: (a) BER and (b) secrecy capacity

exploit multipath propagation. MIMO has become an essential element of RF wireless communication standards, including the IEEE 802.11 protocol suite and LTE. More recently, MIMO has been applied to VLC, however, adopting MIMO in VLC has its challenges. In VLC, due to the propagation properties of light, the channel variations between transmitting elements in close proximity are insufficient for accurate MIMO spatial identification [49]. Nevertheless, MIMO techniques rely on channel variations to provide spectral efficiency gains [50]. MIMO in VLC has had extensive attention from the research community. A performance comparison of the various MIMO techniques proposed for indoor environments utilizing optical wireless communications (including VLC) was given in [51]. Experimental demonstrations of VLC-MIMO systems were given in [52–54]. Extensive research was dedicated to designing receivers optimized for MIMO-VLC systems such as in [55–57]. Additionally, various publications focused on interference cancellation between different users in multiuser MIMO (MU-MIMO) settings, as in [58,59]. A thorough survey on MU-MIMO in VLC can be found in [60] and a comprehensive investigation of both the illumination and communication coverage of an indoor MIMO-VLC system in [61].

The simplest form of MIMO is repetition coding that simultaneously sends the same signal over multiple transmitters. It has the advantage of transmit-diversity, making it a technique that is resilient to noise and transmitter-receiver misalignment [51]. However, it does not provide any spectral multiplexing gains, as the same data is transmitted from all transmitters. In spatial multiplexing (SMP), parallel data streams are emitted from the transmitters, enabling high data rates by harnessing SMP gains [62]. To provide these gains, low channel correlation is required. Spatial modulation (SM) comes as a compromise between the two previous techniques, by setting one transmitter ON at a time. Hence, it eradicates interchannel interference while improving the system's complexity and spectral efficiency [51]. The concept behind SM is extracting additional bits from the spatial dimension by identifying the transmitting element from which the information was sent. Conventional techniques require accurate channel state information (CSI) and complex precoding for spatial identification, which is infeasible in low-complexity systems [62]. Accordingly, MIMO spectral efficiency gains in low-complexity VLC-based systems are arduous. A VLC MIMO communication system acceding IM and DD can be modeled as [63],

$$y = Hx + w \tag{14.30}$$

where w represents the noise, H is the MIMO channel matrix, and x is the transmitted signal. The noise in VLC is a result of the ambient shot light and thermal noise with a distribution modeled as AWGN with zero mean and a variance of σ^2 [32,64]. The channel in the optical domain is presented as real attenuation coefficients, relying heavily on the transmitter–receiver alignment. The line-of-sight channel gain of each propagation link can be calculated by [63]

$$h_{ij} = \begin{cases} \frac{(m+1)A}{2\pi d_{ij}^2} \cos^m \phi_{ij} \cos \varphi_{ij}; & 0 \leq \varphi \leq \varphi_{1/2} \\ 0; & \varphi > \varphi_{1/2} \end{cases} \tag{14.31}$$

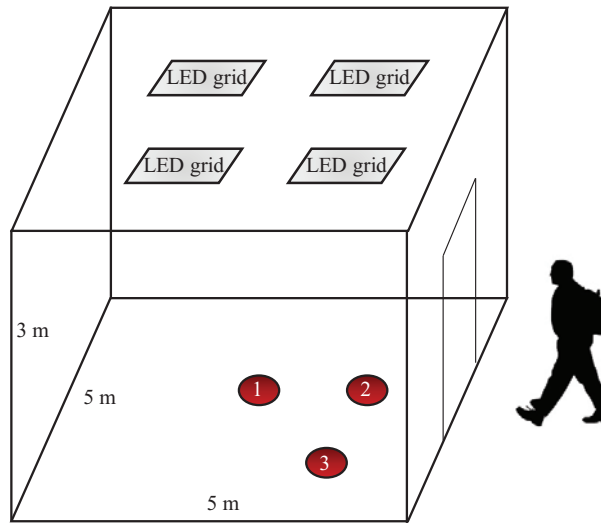


Figure 14.19 The simulated room modeled at different user location scenarios with location 1 denoting (0, 0), location 2 (0, 1.25), and location 3 (−1.25, 1.25)

where the angle of irradiance is ϕ, φ is the angle of incidence, and d is the separating distance between the transmitter i and the receiver j. Furthermore, $\varphi_{1/2}$ denotes the field-of-view (FOV) semi-angle at the receiver, m is the Lambertian emission, and A is the photodetector area. The channel gain of the first reflection path is [32]

$$h_{rij} = \begin{cases} \frac{(m+1)A}{2(\pi d_{1ij} d_{2ij})^2} \cos^m \phi_{ij} \cos \varphi_{ij} \cos \alpha_{ij} \cos \beta_{ij}; & 0 \leq \varphi \leq \varphi_{1/2} \\ 0; & \varphi > \varphi_{1/2} \end{cases} \tag{14.32}$$

According to [51], there are distinct channel transfer factors between each transmitter and receiver pair; hence, the receiver can track the transmitting source and detect spatial symbols. To verify the practicality of the previous statement, the channel gains at three random locations shown in Figure 14.19 are calculated. A 4×1 configuration is considered, i.e., 4 transmitters and 1 receiver setup, also known as an MISO (multiple-input–single-output) setting. The room dimensions are $5 \times 5 \times 3$ m^3 and the LEDs are located on the ceiling with a spacing of 2.15 m from the receiving plane. Further used simulation parameters are listed in Table 14.2. Applying (14.31) into this scenario, the values of the channel gains are

$$\begin{aligned} H_{\text{loc}_1} &\approx 10^{-5} \times (0.2211 \;\; 0.2211 \;\; 0.2211 \;\; 0.2211) \\ H_{\text{loc}_2} &\approx 10^{-5} \times (0.0933 \;\; 0.0933 \;\; 0.3333 \;\; 0.3333) \\ H_{\text{loc}_3} &\approx 10^{-5} \times (0.1192 \;\; 0.0521 \;\; 0.5667 \;\; 0.1192) \end{aligned} \tag{14.33}$$

Based on (14.33), spatial identification in the center of the room (location 1) is almost impossible to achieve, as all the channel gains are equal. Moreover, location

Table 14.2 System parameters for the VLC link

	Parameters	**Values**
Transmitter	Number of LED grids	4
	Transmitted power per LED	20 mW
	Number of LEDs per array	$60 \times 60 = 3{,}600$
Receiver	Active area	1 cm^2
	Half-angle FOV	70°
	Height w.r.t. the floor	0.85 m

two and three, shown in Figure 14.19, also suffer from having equal channel gains. Furthermore, the channel gains resulting from the first reflected path in the previously given scenario are calculated, the values of the sum of first reflection channel gains from the four walls are

$$\begin{aligned} H_{r\text{loc}_1} &\approx 10^{-6} \times (0.1489 \;\; 0.1489 \;\; 0.1489 \;\; 0.1489) \\ H_{r\text{loc}_2} &\approx 10^{-6} \times (0.1144 \;\; 0.1144 \;\; 0.3420 \;\; 0.3420) \\ H_{r\text{loc}_3} &\approx 10^{-6} \times (0.1988 \;\; 0.0734 \;\; 0.5616 \;\; 0.1470) \end{aligned} \tag{14.34}$$

From (14.34), it is evident that even by considering the first reflection paths, spatial variations remain very sensitive to receiver location and hence spatial identification requires perfect and extremely sensitive CSI. Approaches such as link blockage and transmitted power imbalance are proposed in [51]; however, their effect on illumination, which is the primary task of the LEDs, is not studied. As a result, a reliable technique that is capable of identifying the transmitting element without relying on CSI and is independent to transmitter–receiver alignment is needed. A novel technique known as augmented SM (ASM) [65] addresses the aforementioned predicament and will be explained in the forthcoming subsection. The perception of ASM is combining two data streams on a single waveform. One stream supports a lower data rate, while the second stream supports higher speeds. A possible utilization of such unique ASM feature could be sending the identity of the transmitting element, i.e., transmitter PHY ID, on the low-data rate stream while using the high-data stream to carry the information. In ASM, the transmitted signal varies depending on which transmitter is sending the signal. For illustration, Figure 14.20(a) depicts a 16-sample ACO-OFDM signal. As an example, when only two transmitters are assumed, Figure 14.20(b) shows the signal if transmitted by the first transmitter, while Figure 14.20(c) represents the signal if transmitted by the second. ASM is built upon ACO-OFDM for its higher power efficiency, which makes it apt for a multitude of applications, including IoT applications. A transmitted ACO-OFDM signal after inverse fast Fourier IFFT is given by (14.20). The complex bipolar symbols, $X = [0, X_1, \ldots, X_{N/2-1}, 0, X^*_{N/2-1}, \ldots, X^*_1]^T$, are converted into real non-negative intensity signals, $s = [s_0, \ldots, s_{N-1}]^T$. The superscripts $[\cdot]^T$ and $[\cdot]^*$ denote the vector transpose and complex conjugate, respectively.

14.3.2.1 ASM system model

Figure 14.21 shows the block diagram of the transmitter and receiver for the proposed ASM VLC system. In ASM, the time-domain samples of the ACO-OFDM signal, given in (14.35), are first adjusted and scaled to fit within the dynamic range, denoted as Ψ, of the LED. The values of x_n are redesigned in a manner that is estimated by the designated receiver, allowing the receiver to distinguish between transmitters and omit any data incoming from rogue transmitters. As shown in the transmitter block, Figure 14.21(a), the data bits go through the conventional ACO-OFDM symbols formation steps; QAM, Hermitian symmetry, and IFFT to produce the unipolar output. The key to this original approach is in the blocks highlighted in green. At the transmitter's side, each transmitter has a specific control signal that is sent to the reshaping block. Accordingly, a code word is generated that is tailored for this specific transmitter; hence, each transmitter can be identified by the shape of the received signal. A binary sequence, $b_n = [b_0, b_1, \ldots, b_{N-1}]^T$, is used to represent N_t transmitters, with

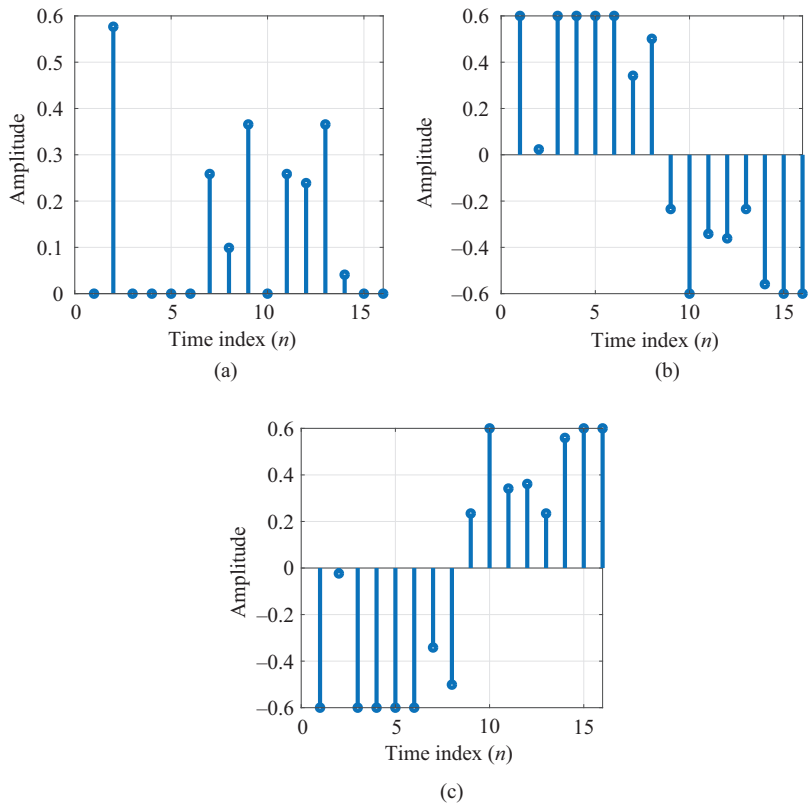


Figure 14.20 *A conventional ACO-OFDM signal (a) and examples of ASM signals representing Tx1 and Tx2 signals for two different transmitters are depicted in (b) and (c), respectively*

$b_i = 0$ or 1 representing an off-state or an on-state, respectively. The sign of the information in the off-state is reversed, while those in the on-state maintain their polarity. In order to differentiate between the transmitters, N_t code words are required, each consisting of N bits that assign the different on-state/off-state to different portions of the transmitted signal. In ASM, the time-domain ACO-OFDM signal is divided into $\log_2 N_t$ groups, with $N/\log_2 N_t$ bits in the code word taking identical values. A general representation of s_n can be given by

$$s_n = p_n \left(\frac{\Psi}{2} - \mu \cdot x_n \right) \tag{14.35}$$

where p_n is related to the transmitter index code word and is given by $p_n = 2b_n - 1$ and μ is the scaling factor applied to the OFDM symbols. At the receiver end, an ML detector is considered as an optimum detector to validate the ASM concept. ML estimation is first invoked on the envelope-like lower rate stream, followed by estimation of the bits encoded on the multilevel samples forming the relatively higher speed stream. After synchronization, the received signal y_n is fed into an ML estimator block, shown in Figure 14.21(b). The received signal y_n can be given by

$$y_n = s_n + I_c + w_n \tag{14.36}$$

with I_c denoting the DC biasing current and w_n is the AWGN at the receiver. According to [66], x_n can be accurately modeled using an independent identically distributed Gaussian random process with the following probability density function (PDF)

$$f_{x,n}(x) = \frac{1}{2}\delta(x) + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2} \right) \tag{14.37}$$

where $\delta(x)$ denotes Dirac delta function. The demodulation process can be modeled as a hypothesis test with nuisance parameters [67]. The N_t hypotheses are denoted by

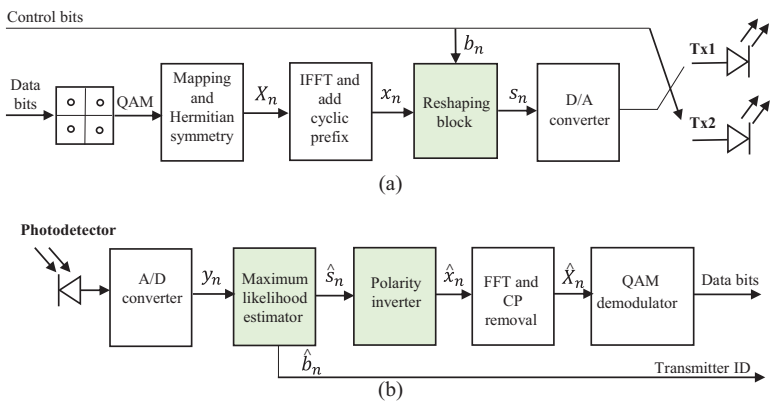


Figure 14.21 ASM system block diagram highlighting ASM's additional block in green depicting (a) transmitter with two possible transmitting elements selected by the control bits and (b) receiver

$\mathscr{H}_i$, $i = 1, \ldots, N_t$, corresponding to the transmitter PHY ID code words conditioned in the form of the vector $p^{\mathscr{H}_i} = \left[p_0^{\mathscr{H}_i}, p_1^{\mathscr{H}_i}, \ldots, p_{N-1}^{\mathscr{H}_i}\right]^T$. The auxiliary vectors, $u^{\mathscr{H}_i}$, given by $u^{\mathscr{H}_i} = y - \frac{\Psi}{2} p^{\mathscr{H}_i}$, where $y = [y_0, \ldots, y_{N-1}]^T$, are defined. Conditioned on the hypotheses and the corresponding estimates, $\hat{s}_{\mathscr{H}_i}$, the likelihood function can be expressed as the joint PDF of $u^{\mathscr{H}_i}$ given by

$$f_{u^{\mathscr{H}_i}}\left(u_0^{\mathscr{H}_i}, \ldots, u_{N-1}^{\mathscr{H}_i} \middle| \hat{s}_{\mathscr{H}_i}, \mathscr{H}_i\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(u_n^{\mathscr{H}_i} + p_n^{\mathscr{H}_i} \hat{s}_{n,\mathscr{H}_i} - I_c\right)^2\right] \tag{14.38}$$

The ML estimator, shown in Figure 14.21, leads to different values for the likelihood functions based on estimates calculated under various hypotheses for the given received signal y. However, it returns only the hypothesis that maximizes the likelihood function as the decoded binary sequence, i.e.:

$$\begin{aligned} \hat{b} &= \arg\max_i \left[f_{u^{\mathscr{H}_i}}\left(u_0^{\mathscr{H}_i}, \ldots, u_{N-1}^{\mathscr{H}_i} \middle| \hat{s}_{\mathscr{H}_i}, \mathscr{H}_i\right)\right] \\ &= \arg\max_i \sum_{n=0}^{N-1} \left(u_n^{\mathscr{H}_i} + p_i^{\mathscr{H}_i} \hat{s}_{n,\mathscr{H}_i} - I_c\right)^2 \end{aligned} \tag{14.39}$$

Once the binary sequence is estimated, the inversion pattern used at the transmitter is known, i.e., the transmitter PHY ID. Then, the signal is reverted back to its original ACO-OFDM unipolar form, by the waveform extractor depicted in Figure 14.21(b), and is demodulated by a conventional ACO-OFDM receiver. Without the ML estimator, the received samples will remain in the pulse-shaped form and the amplitude of the samples are degenerated. Hence, an eavesdropper using a conventional ACO-OFDM receiver, i.e., without the ML block, will receive a distorted signal instead.

The spectral efficiency of previously realized ACO-OFDM system can be expressed as

$$\mathrm{SE}_{\mathrm{ACO-OFDM}} = \frac{1}{4} \log_2(MN_t) \ \text{bps/Hz} \tag{14.40}$$

where N_t denotes the number of transmitters and M is the constellation order. However, the performance of SM is significantly dependent on channel variations; thus, the system's throughput reliability is jeopardized in cases where spatial identification is inaccessible; losing the additional bits extracted from the spatial domain. ASM, on the other hand, eradicates these inconsistencies, allowing spatial identification by design.

14.3.2.2 ASM performance evaluation

To evaluate the efficacy of the proposed system, the pre-discussed configuration is assumed, i.e., a 4×1 MISO configuration. The simulation parameters are 16-QAM and the IFFT length is 64 with the presence of AWGN. Figure 14.22(a) shows the transmitter identification accuracy for various background noise levels vs. the separation distance between the transmitter and the receiver. The accuracy at distance 2.15 m is

highlighted for clarity. Following (14.31) and (14.32), the received signal is inversely proportional to the square of the transmitter–receiver separation, i.e., d^2. Results provided in Figure 14.22(a) show that almost 100% transmitter identification accuracy can be achieved at SNR = 17 dBm, given the transmitted signal power is 17 dBm and the noise power is equal to 0 dBm at $d = 2.15$ m. Intuitively, as the noise power increase, the identification accuracy decreases. However, even at SNR = 12 dBm, above 96% identification accuracy is achieved as depicted by the red curve in Figure 14.22(a). These notable findings demonstrate the competence of ASM for IoT deployments, as this PHY identification capability is dependent solely on the design and does not require high-complexity receivers that require CSI. Moreover, it can be simply scaled up as it is independent of channel uniqueness, providing significant spectral efficiency enhancements. Moreover, since an eavesdropper is oblivious to the reshaping process and the pattern in which it is inserted, PHY security is added to the gains of ASM. Figure 14.22(b) demonstrates this gain by showing the periodogram of the received signal in comparison with that of the transmitted signal for a legitimate receiver and an eavesdropper. It is visually obvious that without the ASM receiver, the eavesdropper receives a signal that does not represent the transmitted signal due to this redesigning process. In summary, in order to utilize MIMO in VLC, techniques need to address the spatial uniqueness problems in VLC. This can be achieved by adjusting and increasing the number of receiving elements, which, in return, will increase the size and complexity of detection. Another alternative would be using digital signal processing approaches, such as in ASM.

14.3.3 Deep learning in VLC

DL techniques are becoming a part of groundbreaking systems in various fields, particularly computer vision and speech recognition. They are applied in various fields to make the systems more efficient. Mostly, they are used in images for applications

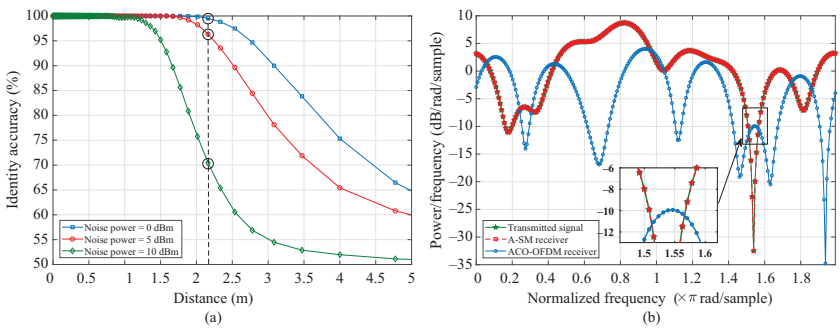


Figure 14.22 ASM performance for 16-QAM ACO-OFDM under AWGN channel model with N = 64: (a) identity accuracy vs. transmitter-receiver separation distance given the transmitted signal power is equal to 17 dBm and (b) signal periodogram comparing the original signal to the received version by the legitimate user and an eavesdropper

such as object detection, handwritten digit classification, and face detection. However, there has been a surge of interest in using DL in communication systems. One of the fundamental communication problems is the reliability in the reconstruction of the transmitted messages from a noisy environment, i.e., communication reliability between the transmitter and the receiver [68]. In classical communication, a profusion of research has been done in order to improve communication reliability. In a classical approach, all the processing blocks in the communication chain are separately optimized in order to achieve performance close to the theoretical. However, such optimization process is considered suboptimal. Alternatively, the idea of DL in communication is based on end-to-end performance enhancement through joint optimization of the whole model to map a set of inputs with certain distribution to a finite set of outputs or targets. Consequently, DL has recently become a potential candidate to achieve reliable communication without any prior mathematical modeling to achieve optimal performance through end-to-end training [69].

Communication techniques are developed based on probability and signal processing theory for channel models. Practical imperfections in these models cannot be replicated. These techniques when tested in scenarios, they have some degree of inaccuracy, resulting in false performance evaluation [70]. DL-based systems can detect faults in the systems without using complex equations. DL can perform a wide range of applications in communication systems like channel modeling and prediction, localization, modulation recognition, and spectrum sensing [69]. The problem of modulation recognition can be solved by studying the use of DL techniques on complex IQ samples. One of the technologies where DL can be applied is VLC. The constraints of the optical signals such as, non-negativity for IM, as well as the average and peak intensity restrictions by the LEDs and the targeted illumination profiles, can be met using deep neural networks (DNNs), which are capable of learning more complicated tasks. Open-source software libraries like Theano, Keras, and Tensorflow have been useful in building network architectures in different domains. In combination with new hardware systems, new applications are designed for DNNs, which have provided a framework for newfangled systems. In the succeeding subsections, two pioneering DNN-based architectures are presented for VLC. The first is the autoencoder OFDM-based VLC system proposed in [71] and the autoencoder architecture presented for optical camera communications (OCCs) in [72].

14.3.3.1 Background

An autoencoder (AE) is a type of a DNN which comprises an encoder and a decoder and serves the purpose of dimension reduction by making use of unsupervised learning. The encoder is a component of AE which compresses the given input into a few numbers of bits. Space represented by these compressed bits is called latent space. The decoder is another component of an AE and it reconstructs the input using the compressed bits in the latent space. AE avoids features that are redundant and include only the pivotal features. It enables capturing the statistical dependencies between different elements of a signal and optimizes a loss function while reconstructing the signal using stochastic gradient descent algorithm during backpropagation. Originally, AE

was used for data compression due to the limitation of the memory resources. Principal component analysis (PCA) is a technique to find a relation in data points in a dataset to predict future models. The main difference between PCA and AE is that AE uses nonlinear transformation, whereas PCA uses linear transformation. DNNs work in two phases:

- Forward pass phase: It generates the output using feed-forward propagation in the network.
- Backward pass phase: It calculates the weighted error for each node in each layer.

First, the entire network performs forward propagation and then the network is trained using back propagation. In back propagation, the flow is from output nodes to input nodes. At the output node, the error is calculated based on the cost function, which can simply be observed as the difference between the actual and desired outputs. Originally, AEs were used for data compression and were used in applications where there was a limitation in memory resources. However, observing the structure of an AE, it is obvious that there are similarities between an AE and a basic communication system. The encoder can be considered as the transmitter, the decoder as the receiver, and the corruption added on the resulted compressed vector can be treated as a communication channel. However, despite of the similarities between an AE and a conventional communication system in terms of structure, they are different in terms of the means of optimization. The optimization in the AE is considered as an end-to-end process, so the reduced dimension and the way of encoding or decoding is learned through the training phase. Figure 14.23 depicts the basic AE block diagram, assuming a single hidden layer. Circles in each layer denote as neurons in the NN. In Layer-1, x_1, x_2, and x_3 are the inputs of the AE. Layer-2 is the hidden layer and Layer-3 is the output layer. Values in the hidden layer are not included in the training dataset. Neurons with value "+1" are the bias units. Then, function f maps the inputs to the hidden layer, similarly another function maps the output of the hidden layer to the final output layer, i.e., $f(x)$ and g maps hidden to output layer which is $g(f(x))$.

14.3.3.2 Autoencoder OFDM-based VLC system

In order to demonstrate the viability of applying DL techniques in VLC, an OFDM-based VLC chain using an AE model is developed. The VLC transmitter is fed with one symbol s out of M possible symbols, determined by the modulation order of the M-QAM under investigation. The encoder maps the input $s \in R^l$ to a reduced dimension $X \in R^{\mathrm{M}}$, where $M < l$. The decoder reconstructs the original input s through decompression of the reduced input dimension X and thus obtains the estimation of the input denoted by $\hat{s}$. l is the system's degrees of freedom (time/frequency domain). The system has a communication rate of $R = k/c$ (bits/channel use), where $k = \log_2(M)$. The whole model is trained such that the values of the weights and biases for both encoder and decoder layers are jointly optimized to minimize the cost function. The categorical cross entropy cost function, L_{cross}, is considered and given by

$$L_{\mathrm{cross}} = H(\hat{s}, s) = H(s) + D_{\mathrm{KL}}(s||\hat{s}) \tag{14.41}$$

where $H(s)$ is the probability mass function (PMF) of the transmitted symbol s and $D_{\mathrm{KL}}(\cdot||\cdot)$ is defined as the Kullback–Leibler divergence between the transmitted and estimated PMF [74]. As shown in Figure 14.24, the inputs to the transmitter are one-hot-encoded symbols s. One-hot encoding is when the element corresponding to the transmitted symbol equals to 1, while the others are 0. For illustration, in the case of 4-QAM, symbol one is presented as (1, 0, 0, 0), symbol two is (0, 1, 0, 0), symbol three is (0, 0, 1, 0), and symbol four is (0, 0, 0, 1). The transmitter consists of multiple dense layers followed by a power normalization layer. The output of the normalization layer is mapped to a complex representation. The complex numbers set by the number of subcarriers N, i.e., length of the IFFT block are mapped and transformed from the frequency-domain to the time-domain. An AWGN channel model is considered, as it is widely used as an accurate channel model for different indoor VLC systems. At the receiver side, the signal is transformed back to the frequency domain using the FFT block then passed to a parallel-to-serial converter. The receiver also consists of multiple dense layers followed by a Softmax activation layer and an argmax layer to decode the highest probability received symbol as $\hat{s}$. Details of the NN architecture are listed in Table 14.3 and the optimized hyperparameters are listed in Table 14.4 with $\mathscr{E}$ as the size of the training dataset. In this implementation, the batch size is equal to the number of subcarriers. It is important to note that a challenging part in the development of the model is to represent and reshape data in a complex form, because NN layers only support real values. Accordingly, custom Keras layers are created for this implementation.

The AE model is tested for different modulation orders, i.e., $M = 4$, 8, 16, and 32. The below graphs highlight the performance of $\pi/2$-shifted 4-QAM. MATLAB is used to verify and compare the results obtained using the AE model with that of a classical simulation model. Figure 14.25(a) shows the learned constellation at the transmitter side after the training phase. It can be noticed that it converges to the ideal case to minimize the cost function. The symbol error rate performance is shown in

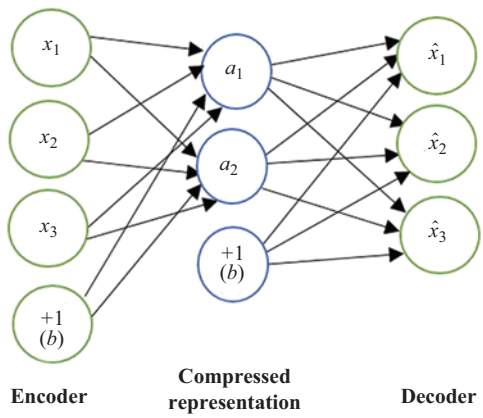


Figure 14.23 Autoencoder architecture [73]

Figure 14.25(b) for different numbers of subcarriers/IFFT lengths. Since, the batch size varies with the number of subcarriers, the case of 8, 16, 32, and 64 subcarriers are implemented. As depicted in Figure 14.25(b), the behavior of the obtained curves from the AE model is consistent with the curves obtained using a classical simulation model in MATLAB. Figure 14.25(c) shows the value of the loss function decreasing as the number of epochs increases. A 100% accuracy can be achieved after less than 15 epochs. In summary, even though this OFDM-based AE model was trained using a single SNR value, its performance was approaching the simulation-based performance that shows the potential of adopting DL in VLC applications.

14.3.3.3 Autoencoder-based optical camera communications

Recently, a VLC technique known as OCC is proposed. In such systems, an imaging detector/camera is used as a receiver to capture the transmitted signals. In both indoor and outdoor applications, OCC systems allow the implementation of various

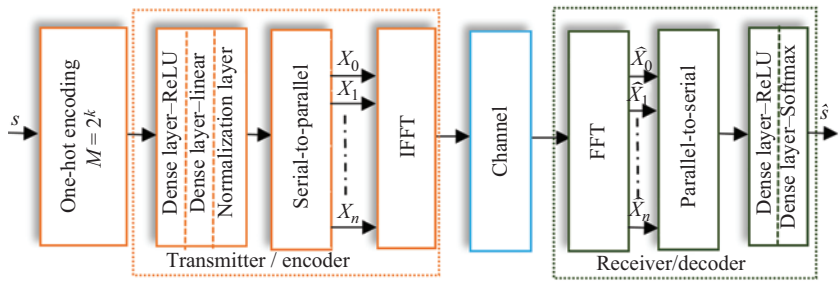


Figure 14.24 OFDM-based AE model

Table 14.3 AE OFDM-based VLC architecture

Encoder network		**Decoder network**	
Layer—activation	**Output dimension**	**Layer—activation**	**Output dimension**
Input	$(\mathscr{E}, M)$	Dense—ReLU	$(\mathscr{E}, M)$
Dense—ReLU	$(\mathscr{E}, M)$	Dense—Softmax	$(\mathscr{E}, c)$
Dense—linear	$(\mathscr{E}, c)$		
Normalize	$(\mathscr{E}, c)$		

Table 14.4 Optimized hyperparameters for AE OFDM-based VLC model

M	**SNR (dB)**	**Epochs**	**Learning rate**	**Training data**	**Testing data**
4	20	50	0.001	8,000	2,000

applications using the cameras integrated in mobile devices, without requiring hardware modifications. In OCC, signals are mainly modulated using single-carrier schemes such as On–Off keying (OOK). This work is primarily focused on proposing a convolutional AE (C-AE) to support OFDM-based OCC through DL, where the OFDM symbol is translated into OOK signaling and transmitted using a 2D-array of LEDs. This will enable an end-to-end performance enhancement by jointly optimizing the LED array and the camera. An OFDM symbol is converted into an $L \times L$ matrix that represents the On and Off states of the pixels of an LED array acting as the transmitting element. Hence, the OFDM symbol is transformed into $L \times L$ OOK

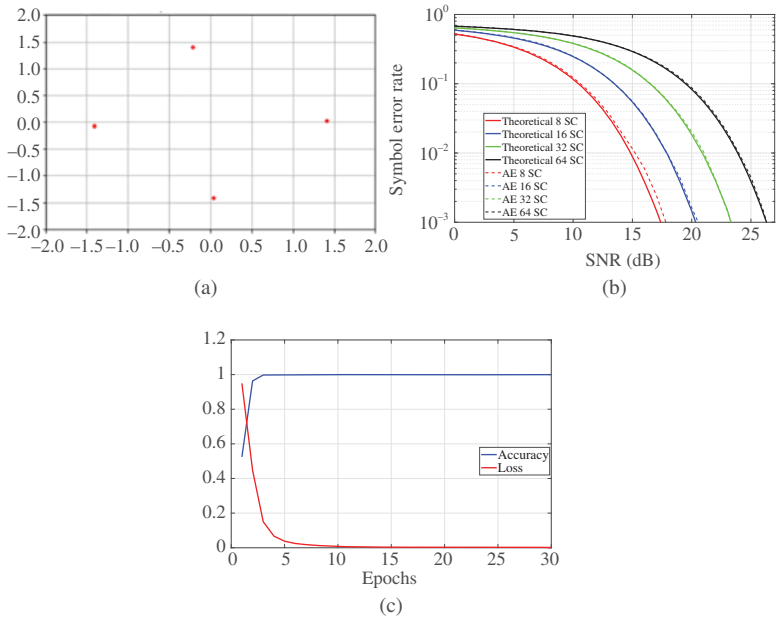


Figure 14.25 *(a) Constellation diagram for QPSK, (b) comparison of SER performance of AE model and classical MATLAB simulation model, and (c) plots of accuracy and loss*

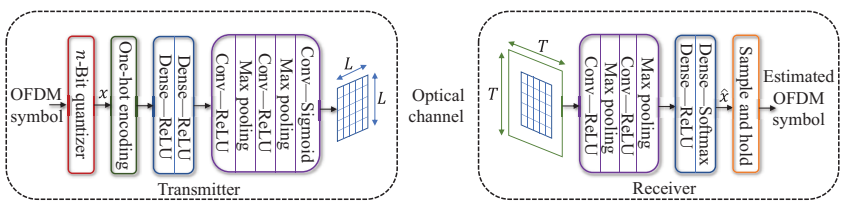


Figure 14.26 *Deep learning framework for OFDM-based OCC equipped with an $L \times L$ LED array and a camera with a captured image of dimension $T \times T$*

signaling. In order to achieve this transformation, the time-domain OFDM symbol is first passed to an n-bit quantizer to create a message $x \in M$ discrete levels, with $M = 2^n$. It is then represented as a one-hot vector and passed to the remaining C-AE network layers, as depicted in Figure 14.26. A rudimentary optical channel is simulated by upscaling the $L \times L$ output from the encoder by 4 with zero-padding of 4 bits. Then, AWGN is applied to the signal with variance $\sigma^2 = (1/R\gamma)$, where R is the communication rate and γ is the SNR. At the receiving end, a camera captures the LED array with dimension $T \times T$ such that $T \geq L$. The captured image is then passed to the decoding stage of the C-AE, in order to produce the estimate $\hat{x}$. The detailed structure of the proposed C-AE is given in Table 14.5. Keras with Tensorflow backend is used for the proposed C-AE implementation using a graphics processing unit backend to perform a multistage training strategy. The C-AE is trained with a single SNR value of 20 dB using the Adam optimizer algorithm and a learning rate of 0.001. In order to achieve binary signaling, i.e., OOK, a parameterizable sigmoid activation function using the Keras lambda layer with the function $x = \frac{1}{1+e^{-\delta x}}$ is implemented. Due to issues with exploding gradients when $\delta > 4$, the network is trained until it converges for $\delta = 1, 2, 3, 4$ using an iterative process. Then, the trained weights are fitted onto a network with $\delta = 1{,}000$, which is feasible for OOK implementation. The loss function is categorical cross-entropy, the optimum training occurs with batch size $= M$, and the C-AE is trained for 100 epochs at each δ. Figure 14.27 shows the normalized confusion matrices using 10^6 time-domain samples of OFDM symbols at $M = 16$ and $M = 64$ for different SNR values during the testing stage. As shown in Figure 14.27(a) and (b), when $M = 16$, samples are reconstructed with almost 100% accuracy even at an SNR as low as 5 dB. On the other hand, at $M = 64$, the accuracy is still maintained at SNR = 10 dB, giving about 99% accuracy. However, at an SNR of 5 dB, accuracy drops to 60%.

In summary, applying DL in VLC is still in its primary stages and hence there is still a great room for improvements. However, the preliminary results show that there is a significant potential for applying DL in a multitude of applications, including traditional VLC and OCC systems.

Table 14.5 Proposed network architecture

Encoder network		**Decoder network**	
Layer—activation	**Output dimension**	**Layer—activation**	**Output dimension**
Dense—ReLU	$M \times 1$	Conv—ReLU	$T \times T$
Dense—ReLU	$16L^2 \times 1$	Max pooling	$T/2 \times T/2$
Conv—ReLU	$4L \times 4L$	Conv—ReLU	$T/2 \times T/2$
Max pooling	$2L \times 2L$	Max pooling	$T/4 \times T/4$
Conv—ReLU	$2L \times 2L$	Dense—ReLU	$M \times 1$
Max pooling	$L \times L$	Dense—Softmax	$M \times 1$
Conv—Sigmoid	$L \times L$		

14.4 Conclusion

The visible light spectrum is 1,000 times larger than the entire RF spectrum of 300 GHz, thus representing a high bandwidth medium. VLC is motivated by several benefits, including license-free operation, minimal health concerns, and energy efficiency. VLC has created a whole range of interesting applications in which dual functionality of simultaneous illumination and data communication is provided. The advancement in LED fabrication has resulted in producing high energy-efficient light sources that can be adopted by VLC technology. As a result, communication can be performed without any entailed energy cost as the cost has already been paid for illumination. Nonetheless, there are a number of practical challenges to be addressed before its widespread acceptance. Due to the potential synergy of lighting and communication, it is unsurprising that VLC has been a subject of increasing interest in academia and industry. In addition, there are standardization efforts in the IEEE 802.15.7 wireless personal area networks and 802.11 WLAN. This chapter discussed, novel modulation techniques proposed in VLC, highlighting MCC and its lightweight version. This chapter also discussed the state-of-the-art topics in VLC literature, such as the coexistence of VLC with RF technology and the entailed challenge in designing technology-independent techniques for hybrid networks. Augmented MIMO was also discussed, and the spatial uniqueness problem in VLC was presented, with ASM

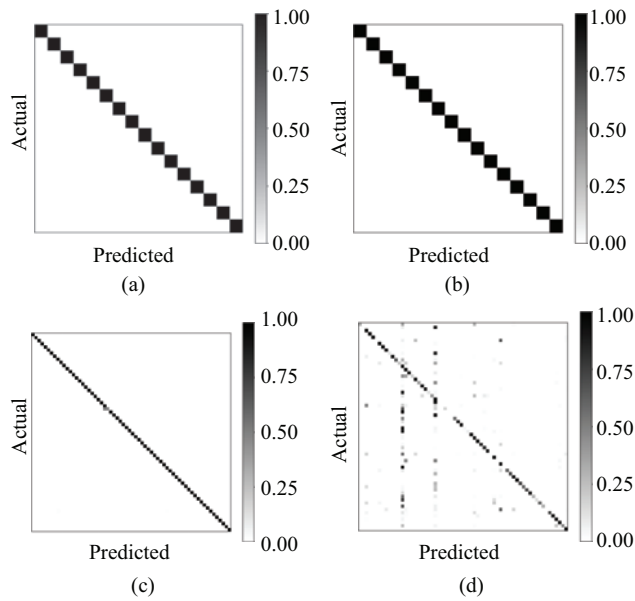


Figure 14.27 Normalized confusion matrices for (a) M = 16, SNR = 10 dB, (b) M = 6, SNR = 5 dB, (c) M = 64, SNR = 10 dB, and (d) M = 64, SNR = 5 dB

introduced as a possible solution. The potential of adopting DL techniques in VLC was explored, the end-to-end autoencoder design as a replacement for conventional transceivers in VLC for applications with photodiode detection and OCCs was discussed. Future research in VLC is anticipated to continue the study of coexistence with other technologies, multiuser access techniques, MIMO transmission, backscatter communication, PHY security, and machine learning.

References

[1] Elgala H, Mesleh R, and Haas H. Indoor Optical Wireless Communication: Potential and State-of-the-Art. IEEE Communications Magazine. 2011;49(9):56–62.

[2] Chowdhury MZ, Hossan MT, Islam A, *et al.* A Comparative Survey of Optical Wireless Technologies: Architectures and Applications. IEEE Access. 2018;6:9819–9840.

[3] Tsonev D, Videv S, and Haas H. Towards a 100 Gb/s Visible Light Wireless Access Network. Optics Express. 2015;23(2):1627–1637.

[4] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021; 2017. https://tinyurl.com/tawwqqj.

[5] Cisco. Cisco Vision: 5G – Thriving Indoor; 2016. https://tinyurl.com/svaejkk.

[6] Khreishah A, Shao S, Gharaibeh A, *et al.* A Hybrid RF-VLC System for Energy Efficient Wireless Access. IEEE Transactions on Green Communications and Networking. 2018;2(4):932–944.

[7] Centre for Energy-Efficient Telecommunications (CEET). The power of wireless cloud; 2013. https://tinyurl.com/wtb5w3u.

[8] Burchardt H, Serafimovski N, Tsonev D, *et al.* VLC: Beyond Point-to-Point Communication. IEEE Communications Magazine. 2014;52(7):98–105.

[9] Damnjanovic A, Montojo J, Wei Y, *et al.* A Survey on 3GPP Heterogeneous Networks. IEEE Wireless Communications. 2011;18(3):10–21.

[10] Lee JS, Su YW, and Shen CC. A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi. In: IEEE 33rd Annual Conference on Industrial Electronics (IECON); 2007. pp. 46–51.

[11] Andrews JG, Claussen H, Dohler M, *et al.* Femtocells: Past, Present, and Future. IEEE Journal on Selected Areas in Communications. 2012;30(3): 497–508.

[12] Pahlavan K and Krishnamurthy P. Principles of Wireless Networks: A Unified Approach. Prentice Hall PTR; 2011.

[13] Islim MS and Haas H. Modulation Techniques for Li-Fi. ZTE Communications. 2016;14(2):29–40.

[14] Armstrong J and Schmidt BJC. Comparison of Asymmetrically Clipped Optical OFDM and DC-biased Optical OFDM in AWGN. IEEE Communications Letters. 2008;12(5):343–345.

[15] Armstrong J and Lowery AJ. Power Efficient Optical OFDM. Electronics Letters. 2006;42(6):370–372.

[16] Armstrong J. OFDM for Optical Communications. Journal of Lightwave Technology. 2009;27(3):189–204.
[17] Elgala H and Little TDC. Reverse Polarity Optical-OFDM (RPO-OFDM): Dimming Compatible OFDM for Gigabit VLC Links. Optics Express. 2013;21(20):24288–24299.
[18] Zafar F, Karunatilaka D, and Parthiban R. Dimming Schemes for Visible Light Communication: The State of Research. IEEE Wireless Communications. 2015;22(2):29–35.
[19] Gacio D, Alonso J, Garcia J, *et al.* PWM Series Dimming for Slow-Dynamics HPF LED Drivers: The High-Frequency Approach. IEEE Transactions on Industrial Electronics. 2012;59(4):1717–1727.
[20] Norsworthy SR, Schreier R, Temes GC, *et al.* Delta-Sigma Data Converters: Theory, Design, and Simulation. vol. 97. IEEE Press, New York; 1997.
[21] Zhang T, Ghassemlooy Z, Rajbhandari S, *et al.* OFDM-PWM Scheme for Visible Light Communications. Optics Communications. 2017;385: 213–218.
[22] Han SH and Lee JH. An Overview of Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission. IEEE Wireless Communications. 2005;12(2):56–65.
[23] Hussein AF, Elgala H, and Little TDC. Visible Light Communications: Toward Multi-Service Waveforms. In: 15th IEEE Annual Consumer Communications Networking Conference (CCNC); 2018 Jan. pp. 1–6.
[24] Hussein AF, Elgala H, and Little TDC. Evolution of Multi-Tier Transmission Towards 5G Li-Fi Networks. In: 2018 IEEE Global Communications Conference (GLOBECOM); 2018. pp. 1–7.
[25] Mesleh R, Elgala H, and Haas H. On the Performance of Different OFDM Based Optical Wireless Communication Systems. IEEE/OSA Journal of Optical Communications and Networking. 2011;3(8):620–628.
[26] Mesleh R, Elgala H, and Haas H. Performance Analysis of Indoor OFDM Optical Wireless Communication Systems. In: IEEE Wireless Communications and Networking Conference (WCNC); 2012.
[27] Hussein AF, Elgala H, Fahs B, *et al.* Experimental Investigation of DCO-OFDM Adaptive Loading Using Si PN-Based Receiver. In: 26th Wireless and Optical Communication Conference (WOCC). IEEE; 2017.
[28] Panta K and Armstrong J. Spectral Analysis of OFDM Signals and its Improvement by Polynomial Cancellation Coding. IEEE Transactions on Consumer Electronics. 2003;49(4):939–943.
[29] Zhao Y. In-band and Out-band Spectrum Analysis of OFDM Communication Systems Using ICI Cancellation Methods. In: WCC 2000 – ICCT 2000 International Conference on Communication Technology Proceedings. vol. 1; 2000. pp. 773–776.
[30] Derickson D and Müller M. Digital Communications Test and Measurement: High-Speed Physical Layer Characterization. Pearson Education; 2007.
[31] International Telecommunication Union. G.975.1: Forward error correction for high bit-rate DWDM submarine systems; 2004. https://tinyurl.com/y4dou234.

[32] Ghassemlooy Z, Popoola W, and Rajbhandari S. Optical Wireless Communications; System and Channel Modelling with MATLAB. CRC Press; 2013.
[33] Feng L, Hu RQ, Wang J, *et al.* Applying VLC in 5G Networks: Architectures and Key Technologies. IEEE Network. 2016;30(6):77–83.
[34] Figueiredo M, Alves LN, and Ribeiro C. Lighting the Wireless World: The Promise and Challenges of Visible Light Communication. IEEE Consumer Electronics Magazine. 2017;6(4):28–37.
[35] Matheus LEM, Vieira AB, Vieira LFM, *et al.* Visible Light Communication: Concepts, Applications and Challenges. IEEE Communications Surveys Tutorials. 2019 Fourth quarter;21(4):3204–3237.
[36] Kashef M, Ismail M, Abdallah M, *et al.* Energy Efficient Resource Allocation for Mixed RF/VLC Heterogeneous Wireless Networks. IEEE Journal on Selected Areas in Communications. 2016;34(4):883–893.
[37] Shao S, Khreishah A, Ayyash M, *et al.* Design and Analysis of a Visible-Light-Communication Enhanced WiFi System. IEEE/OSA Journal of Optical Communications and Networking. 2015;7(10):960–973.
[38] Zou Y, Zhu J, Wang X, *et al.* A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends. Proceedings of the IEEE. 2016;104(9):1727–1765.
[39] Zhang J, Duong TQ, Woods R, *et al.* Securing Wireless Communications of the Internet of Things from the Physical Layer, An Overview. Entropy. 2017;19(8):1–16.
[40] Classen J, Steinmetzer D, and Hollick M. Opportunities and Pitfalls in Securing Visible Light Communication on the Physical Layer. In: Proceedings of the 3rd Workshop on Visible Light Communication Systems. VLCS '16. New York, NY, USA: Association for Computing Machinery; 2016. pp. 19–24.
[41] Zhang J, Marshall A, Woods R, *et al.* Design of an OFDM Physical Layer Encryption Scheme. IEEE Transactions on Vehicular Technology. 2017;66(3):2114–2127.
[42] Hamamreh JM, Furqan HM, and Arslan H. Classifications and Applications of Physical Layer Security Techniques for Confidentiality: A Comprehensive Survey. IEEE Communications Surveys Tutorials. 2018:1–57.
[43] Mostafa A and Lampe L. Physical-Layer Security for Indoor Visible Light Communications. In: 2014 IEEE International Conference on Communications (ICC); 2014. pp. 3342–3347.
[44] Mostafa A and Lampe L. Physical-Layer Security for MISO Visible Light Communication Channels. IEEE Journal on Selected Areas in Communications. 2015;33(9):1806–1818.
[45] Zaid H, Rezki Z, Chaaban A, *et al.* Improved Achievable Secrecy Rate of Visible Light Communication with Cooperative Jamming. In: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2015. pp. 1165–1169.
[46] Liu X, Wei X, Guo L, *et al.* A New Eavesdropping-Resilient Framework for Indoor Visible Light Communication. In: 2016 IEEE Global Communications Conference (GLOBECOM); 2016. pp. 1–6.

[47] Mukherjee A. Secret-Key Agreement for Security in Multi-Emitter Visible Light Communication Systems. IEEE Communications Letters. 2016;20(7): 1361–1364.

[48] Khadr MH, Elgala H, Ayyash M, *et al.* Technology Independent Security Aware OFDM (SA-OFDM). In: IEEE International Symposium on Personal, Indoor and Mobile Radio Communications; 2019. pp. 1–6.

[49] Khalighi MA and Uysal M. Survey on Free Space Optical Communication: A Communication Theory Perspective. IEEE Communications Surveys Tutorials. 2014 Fourth quarter;16(4):2231–2258.

[50] Rusek F, Persson D, Lau BK, *et al.* Scaling Up MIMO: Opportunities and Challenges With Very Large Arrays. IEEE Signal Processing Magazine. 2013;30(1):40–60.

[51] Fath T and Haas H. Performance Comparison of MIMO Techniques for Optical Wireless Communications in Indoor Environments. IEEE Transactions on Communications. 2013;61(2):733–742.

[52] Werfli K, Chvojka P, Ghassemlooy Z, *et al.* Experimental Demonstration of High-Speed 4×4 Imaging Multi-CAP MIMO Visible Light Communications. Journal of Lightwave Technology. 2018;36(10):1944–1951.

[53] Wei L, Zhang H, and Song J. Experimental Demonstration of a Cubic-Receiver-Based MIMO Visible Light Communication System. IEEE Photonics Journal. 2017;9(1):1–7.

[54] Lin B, Ghassemlooy Z, Tang X, *et al.* Experimental Demonstration of Optical MIMO NOMA-VLC with Single Carrier Transmission. Optics Communications. 2017;402:52–55.

[55] Wang TQ, He C, and Armstrong J. Performance Analysis of Aperture-Based Receivers for MIMO IM/DD Visible Light Communications. Journal of Lightwave Technology. 2017;35(9):1513–1523.

[56] He C, Cincotta S, Mohammed MMA, *et al.* Angular Diversity Aperture (ADA) Receivers for Indoor Multiple-Input Multiple-Output (MIMO) Visible Light Communications (VLC). IEEE Access. 2019;7:145282–145301.

[57] Park K, Oubei HM, Alheadary WG, *et al.* A Novel Mirror-Aided Non-Imaging Receiver for Indoor 2×2 MIMO-Visible Light Communication Systems. IEEE Transactions on Wireless Communications. 2017;16(9):5630–5643.

[58] Wang Q, Wang Z, Qian C, *et al.* Multi-User MIMO-OFDM for Indoor Visible Light Communication systems. In: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2015. pp. 1170–1174.

[59] Feng Z, Guo C, Ghassemlooy Z, *et al.* The Spatial Dimming Scheme for the MU-MIMO-OFDM VLC System. IEEE Photonics Journal. 2018;10(5): 1–13.

[60] Al-Ahmadi S, Maraqa O, Uysal M, *et al.* Multi-User Visible Light Communications: State-of-the-Art and Future Directions. IEEE Access. 2018;6:70555–70571.

[61] Chen C, Zhong W, and Wu D. On the Coverage of Multiple-Input Multiple-Output Visible Light Communications [Invited]. IEEE/OSA Journal of Optical Communications and Networking. 2017;9(9):D31–D41.

[62] Lu L, Li GY, Swindlehurst AL, *et al.* An Overview of Massive MIMO: Benefits and Challenges. IEEE Journal of Selected Topics in Signal Processing. 2014;8(5):742–758.

[63] Dawoud DW, Héliot F, Imran MA, *et al.* Spatial Quadrature Modulation for Visible Light Communication in Indoor Environment. In: 2017 IEEE International Conference on Communications (ICC); 2017. pp. 1–6.

[64] Li E, Zhang W, Sun J, *et al.* Energy-Spectral Efficiency Tradeoff of Visible Light Communication Systems. In: 2016 IEEE/CIC International Conference on Communications in China (ICCC); 2016. pp. 1–5.

[65] Khadr MH and Elgala H. Augmented Spatial Modulation (ASM): Secure and Efficient IoT Compatible MIMO Based VLC Links. In: IEEE International Symposium on Personal, Indoor and Mobile Radio Communications; 2019. pp. 1–6.

[66] Chen L, Krongold B, and Evans J. Performance Analysis for Optical OFDM Transmission in Short-Range IM/DD Systems. Journal of Lightwave Technology. 2012;30(7):974–983.

[67] Wang TQ and Huang X. Fractional Reverse Polarity Optical OFDM for High Speed Dimmable Visible Light Communications. IEEE Transactions on Communications. 2018;66(4):1565–1578.

[68] Dörner S, Cammerer S, Hoydis J, *et al.* Deep Learning Based Communication Over the Air. IEEE Journal of Selected Topics in Signal Processing. 2018;12(1):132–143.

[69] O'Shea T and Hoydis J. An Introduction to Deep Learning for the Physical Layer. IEEE Transactions on Cognitive Communications and Networking. 2017;3(4):563–575.

[70] O'Shea TJ, Karra K, and Clancy TC. Learning to Communicate: Channel Auto-encoders, Domain Specific Regularizers, and Attention; 2016.

[71] Pachpande PG, Khadr MH, Hussien H, *et al.* Autoencoder Model for OFDM-based Optical Wireless Communication. In: OSA Advanced Photonics Congress (AP) 2019 (IPR, Networks, NOMA, SPPCom, PVLED). Optical Society of America; 2019. p. SpT2E.3.

[72] Walter I, Khadr MH, and Elgala H. Deep Learning Based Optical Camera Communications. In: OSA Advanced Photonics Congress (AP) 2019 (IPR, Networks, NOMA, SPPCom, PVLED). Optical Society of America; 2019. p. SpTh3E.2.

[73] Pachpande PG. Communications Using Deep Learning Techniques. Master's Thesis. University at Albany; 2019. Available from: https://tinyurl.com/rty5pnw.

[74] Cover TM and Thomas JA. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing, 2nd Edition). Wiley-Interscience; 2006.

Chapter 15

Conclusions and future developments

Himal A. Suraweera[1], Jing Yang[2], Alessio Zappone[3] and John S. Thompson[4]

This book has provided a thorough discussion of research trends in the field of energy-efficient wireless systems and networks. Our treatment of this topic has been presented in three main parts. The first part described new mathematical tools and concepts that can be used to analyze communication systems in order that they can be configured to operate in a very energy-efficient manner. The second part of this book moved on to discuss how wireless devices and networks can be powered using either renewable energy sources, such as wind or solar power, or even by energy harvesting, where a radio receiver can exploit the energy available from the radio spectrum directly. The final part of this book explores a number of promising new technologies for energy-efficient operation, including concepts such as massive multiple input multiple output (MIMO), interference cancellation approaches, and innovative visible light communications.

Ever since the 3GPP standards body defined the first new radio release 15 in mid-2018, the evolution of fifth-generation (5G) wireless technologies has progressed with new opportunities to improve the energy efficiency [1]. In this concluding chapter, we discuss some of them and take a look into several promising green solutions for the realization of 5G phase II and beyond. These points are organized into two main sections, dealing with "Flattening the Energy Curve to Support 5G Evolution" and "Potential Solutions for a Green Future."

15.1 Flattening the energy curve to support 5G evolution

5G has brought with it an increase in energy consumption due to the need to densify the network, meet the rising traffic demands of end-users, and in the use of new

[1]Department of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya, Sri Lanka
[2]School of Electrical Engineering and Computer Science, The Pennsylvania State University, State College, PA, USA
[3]Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy
[4]Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, UK

high-frequency millimeter wave (mmWave) bands. As a result, operators are faced with the challenge of improving the energy efficiency of 5G network deployment. In order to meet energy efficiency targets, network should adopt the latest technology solutions and operate the infrastructure intelligently.

A holistic approach for energy reduction considering the entire network will yield significant energy savings. Traditionally, operators have focused on providing capacity enhancements for hot spots, city centers, and populated urban areas. Traffic volume carried in most real networks fluctuates from site to site, and it has been observed that a small percentage of medium-to-high-capacity sites carry the significant network traffic [2]. On the other hand, although a large percentage of sites carry low traffic, they must remain active most of the time and thus waste large amounts of energy. Clearly, there will be opportunities to improve the overall energy efficiency, when such sites are also considered for energy optimization.

Softwarization and virtualization of resources in 5G will help to flatten the energy curve. To this end, 5G network slicing technology allows operators to provide diverse services by creating multiple logical networks with increased energy efficiency. Further as 5G evolves, data analytics together with energy saving software will play a key role in reducing the network energy consumption. For example, data analytics is helpful to identify base station inactive periods. It is interesting to note that even when there is no output power, most hardware components of a base station remain operational to transmit synchronization and reference signals [3]. Therefore, when there is no traffic, deep sleep modes are useful to lower the energy consumption. Further, as temporal and spatial traffic variations occur during the day, software can be used to switch off active hardware to promote energy sustainability. There is also scope to reduce the signaling overhead so that base stations can go into prolong periods of sleep. Data analytics is also useful as a real-time assessment tool of radio transceiver hardware performance. Hence, hardware anomalies that increase the power consumption can be diagnosed or even predicted well in advance. Today, power consumption in data centers is a major problem. A study published in [4] estimates that in 2018, data center energy usage was 205 TWh of electricity or roughly 1% of the world's demand for electricity in that year. For a greener future, 5G operators need to look for clean energy sources to run their data centers. As 5G evolves, it is expected that more artificial intelligence (AI)-powered edge computing devices will be deployed at the fringes of the network. However, training of AI models consumes large amounts of power. In order to address this problem, it is worthwhile to investigate on smart targeted data, new learning architectures, and low power custom hardware.

Power amplifiers determine the power efficiency and battery lifetime of base stations and mobile handsets. High-frequency mmWave power amplifiers required for 5G signal transmission are inefficient and pose a serious challenge for green operation. The market of handset-dedicated power amplifiers at present are dominated by III–V processes (such as Gallium Nitride or Gallium Arsenide) and these can also be used in 5G femtocell applications. Moreover, power amplifier integration on CMOS could be beneficial for implementation in 5G macro cells [5]. In all cases, use of predistortion can improve the power amplifier linearity.

Newer 5G new radio releases are expected to address new verticals such as massive industrial wireless sensor networks, intelligent transportation systems, and vehicle-to-anything communications. The sensors in such applications will be embedded everywhere and are further required to operate for years without battery replacement. Such sensors can be powered by ambient energy sources such as solar, wind, vibrations, or incoming radio signals similar to the passive RFID tags and ambient backscattering systems discussed in Chapter 7 of this book. Another solution to power these zero-energy devices is to adopt dedicated radio frequency beacons to harvest sufficient amounts of energy. In addition, advancements in printed electronics will allow miniaturization and embedding of simple tags almost into any surface or packaging [6]. Finally, as the 5G ecosystem continues to expand and cover regions such as oceans, sparsely populated remote areas of the earth and space. Networks deployed in these environments would feature heterogeneous topologies, technologies, and devices coexisting together to provide connectivity and extend coverage. It is important to consider green metrics and design principles for energy-efficient implementation. On a parallel development, recently drone-based aerial communication for 3D coverage has received significant research interest [7]. 5G use cases would benefit from drone deployment, for example, drones show promise in disaster recovery situations, as aerial base stations or to handle increased traffic in hotspot areas. However, limited available on-board power of drones reduces their usefulness in many cases. As such, successful integration of 5G systems and drones will depend on investigating solutions such as dynamic repositioning or design of energy aware network optimization with drone base stations.

15.2 Potential solutions for a greener future

While 5G evolution continues to take place, research has already shifted toward the initial design and development of the next generation of wireless communications. As we move ahead into 6G [8], service providers are expected to further increase the network capacity, extend coverage, and introduce different types of advanced use cases. The following solutions are currently being studied for increasing the energy efficiency of future wireless systems and networks.

AI-based energy-efficient design: AI is emerging as a promising technology to enable the management of complex wireless networks in real time with limited computational complexity [9]. The use of deep learning enables most of the processing complexity to shift offline for training the neural network. The trained network can then be used online, with a much lower complexity, to predict the best policy to follow. AI-based techniques hold the promise to learn and dynamically adjust important system and channel parameters for wireless transmission with minimal complexity. This saves considerable energy that normally is used for overhead signaling and computations in the digital signal processor. On the other hand, training deep neural networks is an energy-intensive process that requires a large amount of training data, and which should be repeated when the wireless environment has significantly changed. Thus, more research effort is needed to understand how deep learning methods perform

in rapidly time-varying environments or to develop more efficient training methods capable of operating with limited training data and low energy consumption. This is also important to enable AI algorithms to run in low-cost devices with limited processing capabilities, such as mobile terminal or Internet of Things sensors, which is a key aspect to support the rise of truly self-organizing wireless networks.

Cell-free massive MIMO: Most centralized MIMO base stations are mounted in towers and rooftops. Such installations are vulnerable to path loss blockage effects and require transmissions at high power. An alternative solution would be to spread antennas in the environment in order to create a distributed antenna system such as cell-free massive MIMO setup [10]. In the cell-free massive MIMO architecture, a particular user will be surrounded by several antennas and hence low power signals can be used. However, antennas need to be connected with many meters of cables and the complexity of controlling antennas in the network becomes high. One technology to overcome this problem is Ericsson's radio stripe that is a large-scale distributed, serial, and integrated antenna system.

Reconfigurable intelligent surfaces (RISs): RISs are planar structures made of several individually tunable elements, called meta-atoms or passive scatterers that can be programmed and appropriately reconfigured to control the phase of the incoming electromagnetic signal, by reflecting or refracting it toward specified locations. Recently, RISs have emerged as a promising technique for future wireless networks [11]. Together with the use of AI, they enable the concept of smart radio environments, wherein RISs are used to coat environmental objects that are present in the propagation environments and are intelligently reconfigured to adapt to the temporal evolution of the propagation channels. Being composed of elementary reflecting elements spaced at sub-wavelength distances over a sheet of special material known as meta-material, RISs are not bound by conventional reflection and diffraction laws but instead can modify the phase and direction of the radio wave impinging on them in a fully customizable way. RISs can be deployed on the walls of buildings or can be used to coat the environmental objects between the communicating devices, which effectively makes the wireless channel a new variable to be optimized, besides the design of the transmitters and receivers. Compared to traditional antenna arrays, RISs have the advantage of granting a large amount of degrees of freedom to exploit, in a more energy-efficient and cost-efficient way, being composed of cheap and nearly passive reflecting elements. Preliminary results have shown that a RIS can be more energy-efficient than traditional relaying schemes [12]. Further research in this direction is required to shed light on the energy efficiency of RIS-based communications.

Combined communication, sensing, and localization: Beyond 5G systems are expected to use new frequencies, including the sub-THz region of the radio spectrum. Such frequencies are suitable for sensing of the wireless environment and obstacle recognition [13]. For example, sensing can be performed so that the local propagation environment of the users could be learnt and profiled in real time. Furthermore, due to the high-rate 6G communication links, location information will be able to rapidly distribute among sensing devices. In this way, there is a possibility to create custom beams toward individual users to deliver a better performance with very high energy efficiency. Sensing also allows localization through which energy-efficient

solutions can be introduced. In cases where precise localization techniques are unable to deploy, alternative solutions such as channel charting [14] will enable to obtain pseudo-locations within the cell, thus ensuring beam space energy aware approaches to be designed. 6G will present system designers with new opportunities to make interpretations of the "context" and coupled with location information, such context-aware applications can be made energy efficient through tailored algorithms and intelligent transmission.

Seamless integration of satellite and terrestrial communications: Recently, broadband satellite communications have gained renewed interest due to technological advances and cost-effective manufacturing leading to new networks enabled by private venture companies such as Starlink and OneWeb. The integration of satellite and terrestrial communications can potentially transform the paradigm of next-generation communications, leading to more energy and spectrum efficient solutions. For example, the satellite can serve as the backhaul link for future terrestrial communication systems operating in remote areas. With wide area coverage, the satellite backhaul can broadcast or multicast by content groups of base stations, reducing the energy consumption in multi-hop unicast terrestrial networks. Besides, satellites can also be used for task off-loading from mobile users, which may significantly reduce latency and improve the energy efficiency compared to the stand-alone 5G terrestrial network.

References

[1] I. Chih-Lin, S. Han and S. Bian, "Energy-efficient 5G for a greener future," Nat. Electron., vol. 3, pp. 182–184, 2020.

[2] M. P. Scharp and O. Persson, "Why we need a new approach to network energy efficiency," Ericsson Blog, 2020. [Online]. Available from: https://www.ericsson.com/en/blog/2020/3/5g-network-energy-efficiency [Accessed July 9, 2020].

[3] P. Frenger, P. Moberg, J. Malmodin, Y. Jading and I. Godor, "Reducing energy consumption in LTE with cell DTX," in Proc. IEEE 73rd Vehicular Technology Conference (VTC Spring), Budapest, Hungary, May 2012, pp. 1–5.

[4] E. Masanet, A. Shehabi, N. Lei, S. Smith and J. Koomey, "Recalibrating global data center energy-use estimates," Science, vol. 267, no. 6481, pp. 984–986, 2020.

[5] D. Y. C. Lie, J. C. Mayeda, Y. Li and J. Lopez, "A review of 5G power amplifier design at cm-wave and mm-wave frequencies," Wireless Commun. Mob. Comput., vol. 2018, 16 pages, 2018, Article ID 6793814.

[6] G. A. Torres Sevilla and M. M. Hussain, "Printed organic and inorganic electronics: Devices to systems," IEEE Trans. Emerg. Sel. Top. Circuits Syst., vol. 7, pp. 147–160, 2017.

[7] Y. Zeng, Q. Wu and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," Proc. IEEE, vol. 107, no. 12, pp. 2327–2375, 2019.

[8] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan and M. Zorzi, "Toward 6G networks: Use cases and technologies," IEEE Commun. Mag., vol. 58, no. 3, pp. 55–61, 2020.

[9] A. Zappone, M. Di Renzo and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?," IEEE Trans. Commun., vol. 67, no. 10, pp. 7331–7376, 2019.

[10] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," IEEE Trans. Green Commun. Networking, vol. 2, no. 1, pp. 25–39, 2018.

[11] M. Di Renzo, A. Zappone, M. Debbah, *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," IEEE J. Sel. Areas Commun., in press. [Online]. Available from: https://arxiv.org/pdf/2004.09352.pdf.

[12] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," IEEE Trans. Wireless Commun., vol. 18, no. 8, pp. 4157–4170, 2019.

[13] A. Bourdoux, A. N. Barreto, B. van Liempd, *et al.*, 6G White Paper on Localization and Sensing [White paper], June 2020. [Online]. Available from: https://arxiv.org/abs/2006.01779 [Assessed July 9, 2020].

[14] C. Studer, S. Medjkouh, E. Gönültas, T. Goldstein and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," IEEE Access, vol. 6, pp. 47682–47698, 2018.

Index

Green Communications for Energy-Efficient Wireless Systems and Networks

The ICT industry is a major consumer of global energy. The energy crisis, global warming problems, dramatic growth in data traffic and the increased complexity of emerging networks are pushing academic and industry research towards the development of energy-saving and energy-efficient architectures, technologies and networks in order to reduce the carbon footprint while ensuring efficient and reliable communication networks, and environmental sustainability. Attractive solutions for the design and implementation of energy efficient wireless networks and 5G technologies include massive MIMO, non-orthogonal multiple access, and energy harvesting communications. Tools from areas such as machine and deep learning are being investigated to establish optimal approaches and understand fundamental limits. Moreover, new promising heterogeneous and decentralized network architectures and the Internet-of-Things (IoT) will have an impact on the successful implementation of future and next generation green wireless communications.

The aim of this edited book is to present state-of-the art research from theory to practice, and all aspects of green communication methods and technologies for the design of next generation green wireless communication systems and networks. This advanced research title will be of interest to an audience of researchers, engineers, scientists and developers from academia and the industry working in the fields of ICTs, signal processing, networking, power and energy systems, environmental and sustainable engineering, sensing and electronics. It will also be a very useful text for lecturers, postdocs, PhD and masters students researching the design of the next generation wireless communication systems and networks.

About the Editors

Himal A. Suraweera is a senior lecturer in the Department of Electrical & Electronic Engineering at The University of Peradeniya, Sri Lanka.

Jing Yang is an assistant professor in the Department of Electrical Engineering at The Pennsylvania State University, USA.

Alessio Zappone is a tenured professor in The University of Cassino and Southern Lazio, Cassino, Italy.

John Thompson is a professor in the School of Engineering at The University of Edinburgh, UK.

ISBN 978-1-83953-067-8

9 781839 530678 >

The Institution of Engineering and Technology
theiet.org
978-1-83953-067-8