CONTRIBUTED PAPER

Nonorthogonal Multiple Access for 5G and Beyond

This article surveys the state-of-the-art research on power-domain nonorthogonal multiple access (NOMA), addressing the theoretical advances, major challenges, and opportunities.

By Yuanwei Liu, *Member IEEE*, Zhijin Qin, *Member IEEE*, Maged Elkashlan, *Member IEEE*, Zhiguo Ding, *Member IEEE*, Arumugam Nallanathan, *Fellow IEEE*, and Lajos Hanzo, *Fellow IEEE*

ABSTRACT | Driven by the rapid escalation of the wireless capacity requirements imposed by advanced multimedia applications (e.g., ultrahigh-definition video, virtual reality, etc.), as well as the dramatically increasing demand for user access required for the Internet of Things (IoT), the fifth-generation (5G) networks face challenges in terms of supporting large-scale heterogeneous data traffic. Nonorthogonal multiple access (NOMA), which has been recently proposed for the third-generation partnership projects long-term evolution advanced (3GPP-LTE-A), constitutes a promising technology of addressing the aforementioned challenges in 5G networks by accommodating several users within the same orthogonal resource block. By doing so, significant bandwidth efficiency enhancement can be attained over conventional orthogonal multiple-access (OMA) techniques. This motivated numerous researchers to dedicate substantial research contributions to this field. In this context, we provide a comprehensive overview of the state of the art in power-domain multiplexing-aided NOMA, with a focus on the theoretical NOMA principles, multiple-antenna-aided NOMA design, on the interplay between NOMA and cooperative transmission, on the resource control of NOMA, on the coexistence of NOMA with other emerging potential 5G techniques and on the comparison with other NOMA variants. We highlight the main advantages of power-domain multiplexing NOMA compared to other existing NOMA techniques. We summarize the challenges of existing research contributions of NOMA and provide potential solutions. Finally, we offer some design guidelines for NOMA systems and identify promising research opportunities for the future.

KEYWORDS | Cooperative communication; fifth generation (5G); multiple-input–multiple-output (MIMO); nonorthogonal multiple access (NOMA); power multiplexing; resource allocation

I. INTRODUCTION

A. Brief History of Wireless Standardization

Following the pioneering contributions of Maxwell and Hertz, Marconi demonstrated the feasibility of wireless communications across the Atlantic at the end of the 19th century. By 1928 this technology became sufficiently mature for the police, the gangsters as well as for the rich and famous to enjoy tetherless communications.

An early European development was the Swedish Mobile Telephone System introduced in 1957, which supported 125 users until 1967. In 1966 the Norwegian system was launched, which operated until 1990. Following these, the 1980s led to the rollout of numerous national mobile phone systems, most of which relied on analog frequency modulation and hence were unable to employ digital error-correction codes. Consequently, their ability to exploit the radical advances in digital signal processing remained limited. Hence, the achievable speech quality was typically poor, especially when the mobile station (MS) roamed farther away from the base station (BS).

Hence, during the 1980s, the member states of the European Union launched a large-scale cooperative research program, which led to the standardization of

Manuscript received March 15, 2017; revised June 29, 2017 and September 20, 2017; accepted October 27, 2017. Date of current version November 20, 2017. This work was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/N029720/1 and Grant EP/N029720/2. The work of L. Hanzo was supported by the ERC Advanced Fellow Grant Beam-me-up.
(Corresponding author: Lajos Hanzo.)

Y. Liu, M. Elkashlan and **A. Nallanathan** are with the Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk; maged.elkashlan@qmul.ac.uk; a.nallanathan@qmul.ac.uk).

Z. Qin and **Z. Ding** are with Lancaster University, Lancaster LA1 4YW, U.K. (e-mail: zhijin.qin@lancaster.ac.uk; z.ding@lancaster.ac.uk).

L. Hanzo is with the University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

A brief video and a set of presentation slides can be downloaded from both IEEE Xplore and from ResearchGate (https://www.researchgate.net/publication/320868303_Non-orthogonal_Multiple_Access_for_5G_and_Beyond_Proceedings_of_the_IEEE).

Digital Object Identifier: 10.1109/JPROC.2017.2768666

the second-generation (2G) system known as the global system of mobile (GSM) communications. GSM was the first digital international mobile system, which rapidly spread across the globe. The success of GSM shows the sheer power and attraction of global standardization, motivating competitors to line up behind a common worldwide solution.

Shortly after the ratification of GSM, a number of other digital standards emerged, such as the Pan-American digital advanced mobile phone system (D-AMPS) and the direct sequence code-division multiple-access (DS-CDMA)-based Pan-American system known as IS-95. IS-95 also had an evolved counterpart, namely the Pan-American cdma2000 system, which had three parallel CDMA carriers, leading to the first standardized multicarrier CDMA (MC-CDMA) system [1].

However, given the consumers' thirst for higher bit rates, during the early 1990s, the research community turned its attention to developing the third-generation (3G) system, which was also based on various CDMA solutions. The detailed discussion and the performance characterization of 3G networks may be found in [2].

Despite the 40-year research history of OFDM [3], multicarrier cellular solutions only emerged during the 2000s as the dominant modulation technique in the context of the 3G partnership project's (3GPP) long-term evolution (LTE) initiative. Clearly, during the 2000s, multicarrier solutions have found their way into all the 802.11 wireless standards designed for wireless local area networks (WLANs), while using different-throughput modem and channel coding modes, depending on the near-instantaneous channel quality.

What is so beautiful about multicarrier solutions is their impressive flexibility, since they have a host of different parameters which allow us to appropriately configure them and program them, whatever the circumstances are, regardless of the propagation environment and regardless of the quality-of-service (QoS) requirements, as facilitated by the employment of adaptive modulation and coding (AMC).

Our hope is dear Colleague that would allow us now to briefly review the evolution of signal processing and communications techniques over the past three decades in an anecdotal style with reference to Fig. 1. At the time of writing we are gradually approaching the "5G Place" on our road map of Fig. 1. We are indeed also approaching the bit-rate limits upper-bounded by the channel capacity of both the classic single-input–single-output (SISO) systems as well as

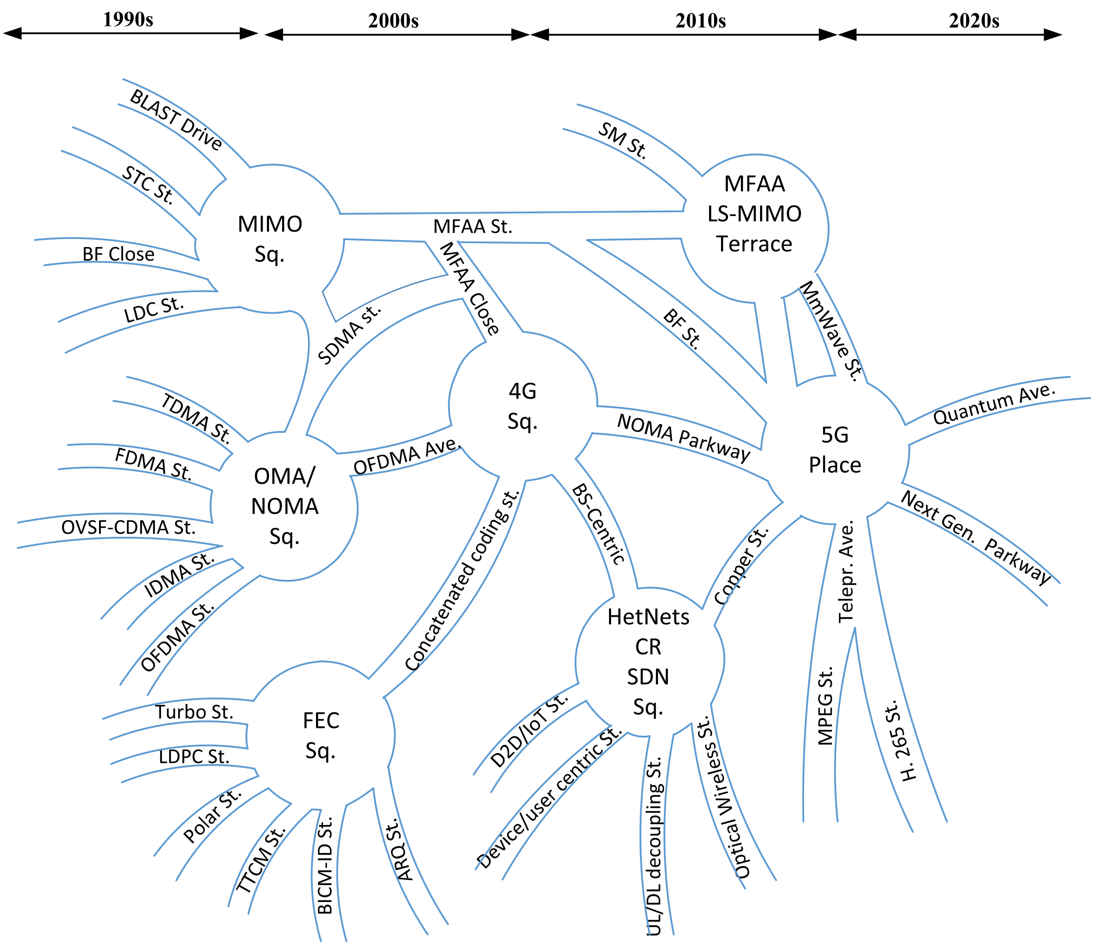


Fig. 1. ***The roadmap for illustrating the brief history of wireless standardization.***

of the MIMO systems. Observe at the top left hand corner of Fig. 1 how the various MIMO solutions, such as bell lab's layered space-time (BLAST) "Drive," space-time coding (STC) "Street," beamforming "Close," and linear dispersion coding (LDC) "Street" merge into multiple-input–multiple-output (MIMO) "Square."

After decades of evolution, the classic orthogonal multiple-access (OMA) schemes, such as time-division multiple-access (TDMA) "Street," frequency-division multiple access (FDMA), orthogonal variable spreading factor-based code-division multiple access (OVSF-CDMA), interleave-division multiple access (IDMA), and orthogonal frequency-division multiple access (OFDMA) "Street" converged to OMA/nonorthogonal multiple-access (NOMA) "Square" of Fig. 1. They have also evolved further along spatial-division multiple-access (SDMA) and multifunctional antenna array "Street"; these solutions have found their way into the fourth-generation (4G) OFDMA systems. As seen at the bottom left corner of Fig. 1, the various advance channel coding schemes have competed for adoption in the 4G standard, which relies of a variety of coding arrangements, including automatic repeat request (ARQ).

At the time of writing the community turned toward the standardization of the 5G systems, with a special emphasis on the NOMA techniques detailed in this treatise, as indicated by the broad NOMA "Parkway," which symbolizes 15 different NOMA proposals. The family of MFAAs also entails the recent spatial modulation (SM) and large-scale (LS) MIMO systems. Since the "road along millimeter wave (mmWave) Street" is rather unexplored and the attenuation is high, the employment of BF is rather crucial, if we want to exploit these rich spectral reserves.

In the bottom right corner of Fig. 1, a number of novel technological advances converge at HetNet "Square," where cognitive radio (CR) and software-defined networks meet device-to-device (D2D) and Internet-of-Things (IoT) networks. A range of sophisticated ideas are also under intensive investigation to resolve the network-centric versus user-centric design options. There is a strong evidence that the latter is more promising, because it is also capable of simultaneous load balancing. There are also strong proposals on decoupling the uplink and downlink teletraffic, with the motivation that mobile-initiated uplink traffic can reach a small-cell BS at a lower transmit power than that of the BS's downlink transmission. Optical wireless based on visible-light communications is also developing quite rapidly, with gigabit copper backhaul networks making promising progress. While no doubt the classic RF systems will continue to evolve toward the next generation, an idea, whose time has come, is quantum communications, as demonstrated by the *Science* article "Satellite-based entanglement distribution over 1200 km" by Yin *et al.* [4].

As the LTE system is reaching maturity and the 4G systems have been commercially deployed, researchers have turned their attention to the 5G cellular network. The latest visual network index (VNI) reports pointed out that by 2020s, the data traffic of mobile devices will become an order of magnitude higher compared to that in 2014 [5]. Apart from meeting the escalating data demands of mobile devices, other challenges of dealing with the deluge of data as well as with the high-rate connectivity required by bandwidth-thirsty applications such as virtual reality (VR), online health care, and the IoT further aggravate the situation. Driven by this, the 5G networks are anticipated with high expectations in terms of making a substantial breakthrough beyond the previous four generations. The often-quoted albeit potentially unrealistic expectations include 1000 times higher system capacity, ten times higher system throughput, and ten times higher energy efficiency per service than those of the 4G networks [6]. Several key directions such as ultra-densification, mmWave communications, massive MIMO arrangements, D2D and machine-to-machine (M2M) communication, full-duplex (FD) solutions, energy harvesting (EH), cloud-based radio access networks (C-RANs), wireless network virtualization (WNV), and software-defined networks (SDNs) have been identified by researchers [7]–[9]. Table 1 lists all acronyms used in this article. Fig. 2 illustrates the whole 5G network structure, including most of the existing/promising techniques.

B. State of the Art of Multiple-Access Techniques

As mentioned before, sophisticated multiple-access (MA) techniques have also been regarded as one of the most fundamental enablers, which have significantly evolved over the consecutive generations in wireless networks [10], [11]. Let us have a deeper looker at the development of MA techniques. As illustrated in Fig. 1, the past three to four decades have witnessed historic developments in wireless communications and standardization in terms of MA techniques. Looking back to the development of the MA formats as we briefly discussed above, in the first generation (1G), FDMA was combined with an analog frequency-modulation-based technology, although digital control channel signaling was used. In the 2G GSM communications, TDMA was used [12]. Then, CDMA, which was originally proposed by Qualcomm [13], became the dominant MA in the 3G networks. In an effort to overcome the inherent limitation of CDMA, namely that the chip rate has to be much higher than the information data rate, OFDMA was adopted for the 4G networks [14]. Based on whether the same time or frequency resource can be occupied by more than one user, the existing MA techniques may be categorized into OMA and NOMA techniques [15]. Among the aforementioned MA techniques, FDMA, TDMA, and OFDMA allow only a single user to be served within the same time/frequency resource block (RB), which belong to the OMA approach. By contrast, CDMA allows multiple users to be supported by the same RB with the aid of applying different unique, user-specific spreading sequences for distinguishing them.

Table 1 List of Acronyms

AF	Amplify-and-Forward
BB	Beamforming Based
BC	Broadcast Channel
BS	Base Station
BF	Beamforming
CB	Cluster-Based
CDMA	Code Division Multiple Access
CF	Compress-and-Forward
CIRs	Channel Impulse Response
CoMP	Coordinated Multipoint
CR	Cognitive Radio
C-RAN	Cloud-based Radio Access Networks
CS	Compressive Sensing
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
D2D	Device-to-Device
DF	Decode-and-Forward
DL	Downlink
DPC	Dirty-Paper Coding
FD	Full-Duplex
FDMA	Frequency Division Multiple Access
IDMA	Leave Division Multiple Access
IMD	Iterative Multi-user Detection
IoT	Internet of Things
LDS	Low-Density Signature
LDPC	Low-Density Parity-Check
LPMA	Lattice Partition Multiple Access
LTE	Long Term Evolution
LMMSE	Linear Minimum Mean Square Error
MA	Multiple Access
MAC	Medium Access Control
M2M	Machine-to-Machine
MNV	Wireless Network Visualization
MPA	Message Passing Algorithms
MUSA	Multi-User Shared Access
MUST	Multi-User Superposition Transmission
NP	Non-deterministic Polynomial-time
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PA	Power Allocation
PDMA	Pattern Division Multiple Access
PF	Proportional Fairness
PLS	Physical Layer Secerity
P2P	Peer-to-Peer
PR	Primary Receiver
PT	Primary Transmitter
PU	Primary User
QoS	Quality of Service
RB	Resource Block
RF	Radio Frequency
RBC	Relaying Broadcast Channel
SA	Signal Alignment
SC	Superposition Coding
SCMA	Sparse Code Multiple Access
SDM	Space Division Multiplexing
SDMA	Space Division Multiple Access
SDN	Software Defined Network
SDR	Software Defined Radio
SD-NOMA	Software Defined NOMA
SIC	Successive Interference Cancelation
SISO	Single-Input Single-Output
SNR	Signal-Noise Ratio
SR	Secondary Receiver
ST	Secondary Transmitter
SU	Secondary User
TCMA	Trellis Coded Multiple Access
TDMA	Time Division Multiple Access
UL	Uplink
V-BLAST	Vertical-Bell Laboratories Layered Space-Time
VLC	Visible Light Communication
VNI	Visual Network Index
WPT	Wireless Power Transfer
ZF	Zero-forcing

Fueled by the unprecedented proliferation of new Internet-enabled smart devices and innovative applications, the emerging sophisticated new services expedite the development of 5G networks requiring new MA techniques. NOMA techniques can be primarily classified into a pair of categories, namely, code-domain NOMA and power-domain NOMA [16].[1]

The most prominent representative code-domain NOMA techniques include trellis-coded multiple access (TCMA) [18], IDMA [19], and low-density signature (LDS) sequence-based CDMA [20]. These solutions are complemented by the more recently proposed multiuser shared access (MUSA) technique [21], pattern-division multiple access (PDMA) [22], and sparse-code multiple access (SCMA).

The power-domain NOMA, which has been recently proposed to 3GPP LTE [23], exhibits a superior capacity region compared to OMA. The key idea of power-domain NOMA is to ensure that multiple users can be served within a given time/frequency RB, with the aid of superposition coding (SC) techniques at the transmitter and successive interference cancellation (SIC) at the receiver, which is fundamentally different from the classic OMA techniques of FDMA/TDMA/OFDMA as well as from the code-domain NOMA techniques. The motivation behind this approach lies in the fact that, again, NOMA is capable of exploiting the available resources more efficiently by opportunistically capitalizing on the users' specific channel conditions [24] and it is capable of serving multiple users at different QoS requirements in the same RB. It has also been pointed out that NOMA has the potential to be integrated with existing MA paradigms, since it exploits the new dimension of the power domain. The milestones of power-domain NOMA are summarized in the timeline of Table 2.

C. Motivation and Contributions

While the above literature review has laid the basic foundation for understanding the development of MA schemes in each generation of cellular networks, the power domain multiplexing-based NOMA philosophy is far from being fully understood. There are some short magazine papers [16], [23], [43], [44] and surveys [45], [46] in the literature that introduce NOMA, but their focus is different from our work. More particularly, Dai *et al.* introduced some concepts of the existing NOMA techniques and identified some challenges and future research opportunities [16] both for power-domain and code-domain NOMA. A magazine paper on power-domain NOMA was presented by Ding *et al.* [23], with

[1]Note that apart from the code and power domain, the spatial domain can also be regarded as another domain for supporting multiple users within the same RB, which is achieved by exploiting the specific "spatial signature" constituted by the channel impulse responses (CIRs) of the users for distinguishing them [3], [10]. A representative MA technique is space-division multiple access (SDMA) [17].

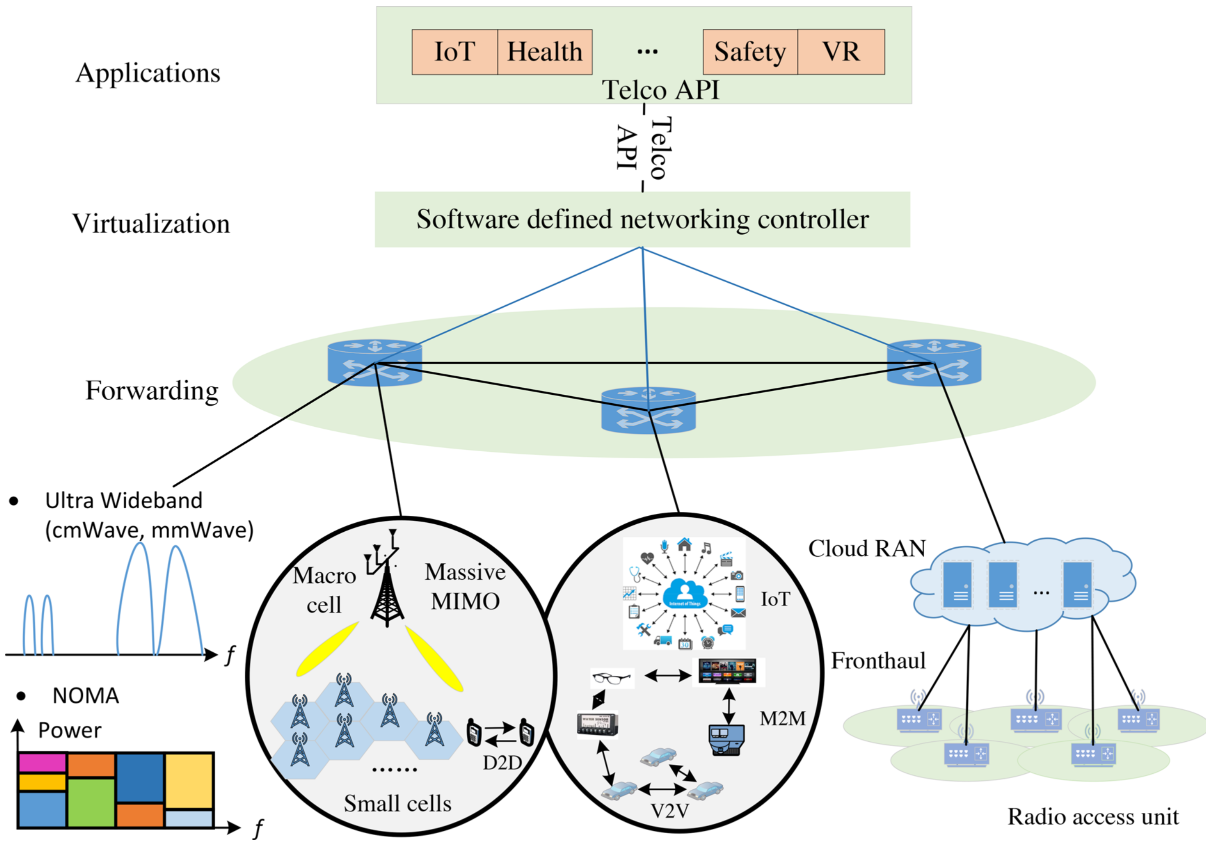


Fig. 2. *Illustration of the future 5G network architecture.*

particular attention devoted to investigating the application of NOMA in LTE and 5G networks. Shin *et al.* [43] discussed the research challenges and opportunities in terms of NOMA in multicell networks, aiming for identifying techniques to manage the multicell interference in NOMA. As a further advance, Ali *et al.* [44] outlined a general framework for multicell downlink NOMA by adopting a coordinated multipoint (CoMP) transmission scheme by considering distributed power allocation (PA) strategy in each cell. Regarding surveys, in [45], Islam *et al.* have surveyed several recent research contributions on power-domain NOMA, while providing performance comparisons to OMA in different wireless communications scenarios. In [46], Tabassum *et al.* investigated the uplink and downlink of NOMA in single-cell cellular networks, identifying the impact of distance of users on the performance attained.

Although the aforementioned research contributions present either general concepts or specific aspects of NOMA, some important NOMA models, the analytical foundations of NOMA, and some of its significant applications in wireless networks have not been covered. Besides, a clear illustration of the historic development of power-domain NOMA milestones is missing. Finally, the comparisons between power-domain NOMA and other practical forms of NOMA have not been discussed. Motivated by all the aforementioned inspirations, we developed this treatise. More explicitly, the goal of this survey is to comprehensively survey the state-of-the-art research contributions that address the major issues, challenges, and opportunities of NOMA, with particular emphasis on both promising new techniques and novel application scenarios. Table 3 illustrates the comparison of this treatise with the existing magazine papers and surveys in the context of NOMA.

To highlight the significance of this contribution, we commence with a survey of NOMA starting with the basic principles, which provides the readers with the basic concepts of NOMA. We continue in the context of multiple-antenna-aided techniques combined with NOMA, followed by cooperative NOMA techniques. We then address another important issue of NOMA, namely its resource and PA problems. Finally, we elaborate on invoking other 5G candidate techniques in the context of NOMA networks. The contributions of this survey are at least fivefold, which are summarized as follows.

1) We present a comprehensive survey on the recent advances and on the state of the art in power-domain multiplexing-aided NOMA techniques. The basic concepts of NOMA are introduced and key advantages are summarized. The research challenges,

Table 2 Timeline of Power-Domain NOMA Milestones

1972	Cover first proposed SC and SIC concepts [25].
1973	Bergmans theoretically demonstrated that SC is capable of approaching the capacity of the Gaussian broadcast channel (BC) [26].
1986	Verdu discovered the optimal maximum-likelihood multi-user receiver for CDMA systems [27].
1994	Patel and Holtzman proposed to apply SIC and PIC in CDMA systems [28].
2001	Li and Goldsmith studied the capacity regions for fading BCs with applying SC and SIC [29], [30].
2003	Mostafa *et al.* demostrated that SAIC can effectively suppress downlink inter-cell interferences in GSM networks [31].
2004	Tse compared the capacity regions of NOMA to OMA both in downlink and uplink [32].
2005	Andrews summarized the development of interference cancellation for cellular systems [33].
2011	Zhang and Hanzo offered a unified treatment for SC aided systems [34].
2012	Vanka *et al.* designed an experimental platform for investigating the implementing performance of SC [35].
2013	Saito *et al.* proposed the concept of two-user downlink NOMA transmission for bandwidth efficiency enhancement [36].
2014	Ding *et al.* developed a multi-user downlink NOMA transmission scheme with randomly deployed users [37].
2015	Xiong *et al.* designed a practical open source SDR-based NOMA prototype for two-user case [38].
2015	Benjebbour *et al.* measured the experimental results on a NOMA test-bed for two-user case [39].
2015	Choi *et al.* proposed a two-user MISO-NONA design for investigating the potential application of multi-antenna techniques in NOMA [40].
2016	Ding *et al.* proposed a cluster-based multi-user MIMO-NOMA structure [41].
2017	Shin *et al.* proposed to apply coordinated beamforming for multi-cell MIMO-NOMA networks to enhance the cell-edge users' throughput [42].

opportunities, and potential solutions are also identified.

2) We investigate the application of multiple-antenna-aided techniques to NOMA. The pair of most dominant solutions, namely cluster-based MIMO-NOMA and beamformer-based MIMO-NOMA, are reviewed and their benefits are examined. Furthermore, we highlight that specific massive MIMO-NOMA solutions are capable of improving the performance of NOMA networks to a large extent. A range of important challenges are elaborated on in the context of multiple-antenna-assisted NOMA and the associated future opportunities are also underlined.
3) By exploiting the specific characteristics of NOMA, we study the interplay between NOMA and cooperative communications. We demonstrate that cooperative NOMA constitutes a promising technique of improving the reliability of the users experiencing poor channel conditions.
4) We identify the potential issues associated both with power and user allocation, which constitute the fundamental problems to be solved for ensuring fairness in the NOMA networks. We point out the significance of designing efficient algorithms for dynamically allocating the resources to the users. Furthermore, we propose the novel concept of software-defined NOMA (SD-NOMA) network architectures, where resource allocation, including the power, is performed on a generic hardware platform by taking into account the global view of the entire network.
5) We identify the major issues and challenges associated with the coexistence of NOMA and the other emerging 5G technologies. The potential solutions based on the current research contributions corresponding to these technologies are also surveyed. We have also discussed the implementation issues and recent standardization progress for NOMA. Finally, power-domain NOMA and other popular forms of NOMA are contrasted. We spotlight that significance to provide a unified framework for NOMA.

D. Organization

The remainder of this paper is structured as follows. Section II presents the basic principles of NOMA, including a brief overview on multiuser detection and interference cancellation (IC), the key techniques adopted, and its main advantages. Section III investigates the recent advances in multiple-antenna-aided NOMA transmission and identifies the open research challenges. In Section IV, the associated research contributions are surveyed in terms of the interplay between NOMA and cooperative transmission. The resource allocation of NOMA, including user association and PA, are studied and summarized in Section V. Finally,

Table 3 Comparison of This Work With Available Magazine Papers and Surveys. Here, "PD" Refers to "Power Domain," "CD" Refers to "Code Domain," "SD" Refers to "Software Defined," "EP" Refers to "Error Propagation," and "CE" Refers to "Channel Estimation"

—	Classifications	Key Contents	Dai *et al.* [16]	Ding *et al.* [23]	Ali *et al.* [44]	Islam *et al.* [45]	Taba. *et al.* [46]	Cai *et al.* [11]	Song *et al.* [47]	Shin *et al.* [43]	This work
PD	Review	Timeline									✓
		MUD/IC									✓
	Basics	Uplink NOMA	✓	✓		✓	✓	✓	✓	✓	✓
		Fairness		✓		✓		✓		✓	✓
	Performance Enhancement	Cooperative		✓		✓	✓	✓	✓		✓
		MIMO-NOMA	✓	✓		✓	✓	✓			✓
		Multi-cell CoMP	✓		✓	✓	✓			✓	✓
	Resource Management	User Scheduling	✓	✓	✓	✓	✓	✓	✓		✓
		Power Allocation			✓	✓	✓	✓	✓	✓	✓
		SD-NOMA	✓								✓
	Practical Issues	EP of SIC				✓	✓			✓	✓
		CE Errors	✓	✓		✓				✓	✓
		Quantization				✓					✓
		Synchronization				✓					✓
		Security								✓	✓
	Compatibility	HetNets-NOMA			✓	✓		✓			✓
		MmWave-NOMA									✓
		CR-NOMA		✓							✓
		D2D-NOMA									✓
		Modulation						✓			
	Progress	Standardization	✓		✓						✓
CD	Single-Carrier	IDMA									✓
		LDS-CDMA	✓					✓			✓
		LPMA						✓			✓
		MUST		✓							
	Multi-Carrier	LDS-OFDM	✓					✓			✓
		SCMA	✓					✓			✓
		PDMA						✓			✓
		MUSA	✓								

Section VI investigates the coexistence of NOMA with other emerging 5G technologies. Section VII points out the implementation challenges as well as the standardization process of NOMA, while Section VIII discusses several other forms of NOMA techniques. Finally, Section IX concludes this treatise. Fig. 3 illustrates the organization of this paper.

II. BASIC PRINCIPLES OF NOMA

The fundamental concept of NOMA facilitates supporting multiple users in the power domain. In contrast to the conventional MA techniques, NOMA[2] uses a new dimension to perform multiplexing within one of the classic time/frequency domains. In other words, NOMA can be regarded as an "add-on" technique, which has the promising potential of facilitating harmonious integration with the existing legacy solutions. In this section, we introduce the basic concept of NOMA by illustrating the associated key technologies and summarize its salient advantages. We also survey the pivotal research contributions both in downlink and uplink NOMA transmissions in single-antenna scenarios.

A. Multiuser Detection and Interference Cancelation

Before diving into the deep description of the key techniques of NOMA, let us first give a brief overview for the multiuser detection (MUD) and IC.

1) Multiuser Detection: The idea of simultaneously receiving the multiple user signals and detecting them is not new; its history goes back to the 1970s (see Cover's work on BCs [25] and Etten's work on multiuser sequence estimator [48]). Then, Verdu discovered the optimal multiuser receiver of CDMA systems in the 1980s [27], which is based on the maximum-likelihood receiver detecting multiple users in parallel. We take CDMA as an example to illustrate MUD. Rather than allocating orthogonal RBs to different users as in FDMA and TDMA, CDMA codes superimpose the signals of multiple users with the aid of unique, user-specific signature referred to as spreading codes. If we use orthogonal K-chip, we can only support K users. But when using nonorthogonal m-sequences in the spirit of NOMA, we can have many more users. Multiple-access interference (MAI) limits the capacity and performance of CDMA systems. While the MAI caused by a single interfering user is generally small, the system becomes interference limited as the number of interferers or their power increases [49], [50]. MUD exploiting the knowledge of both the spreading code and timing (and possibly amplitude and phase) information of multiple users has been regarded as an efficient strategy of improving the system capacity. Various MUD algorithms, such as the optimal maximum-likelihood sequence estimation [27], [51], turbo decoding, matched filter SIC [52], and parallel interference cancellation (PIC) have been designed to reduce the MAI at an affordable complexity cost. Table 1 in [33] illustrates the high-level comparison for different types of multiuser receivers.

Moreover, there are some recently developed joint detection techniques for downlink cellular systems, which are based on single-antenna interference cancellation (SAIC) receivers. More particularly, this technique relies

[2]The NOMA approach addressed in this survey implies the power-domain NOMA, unless stated otherwise.

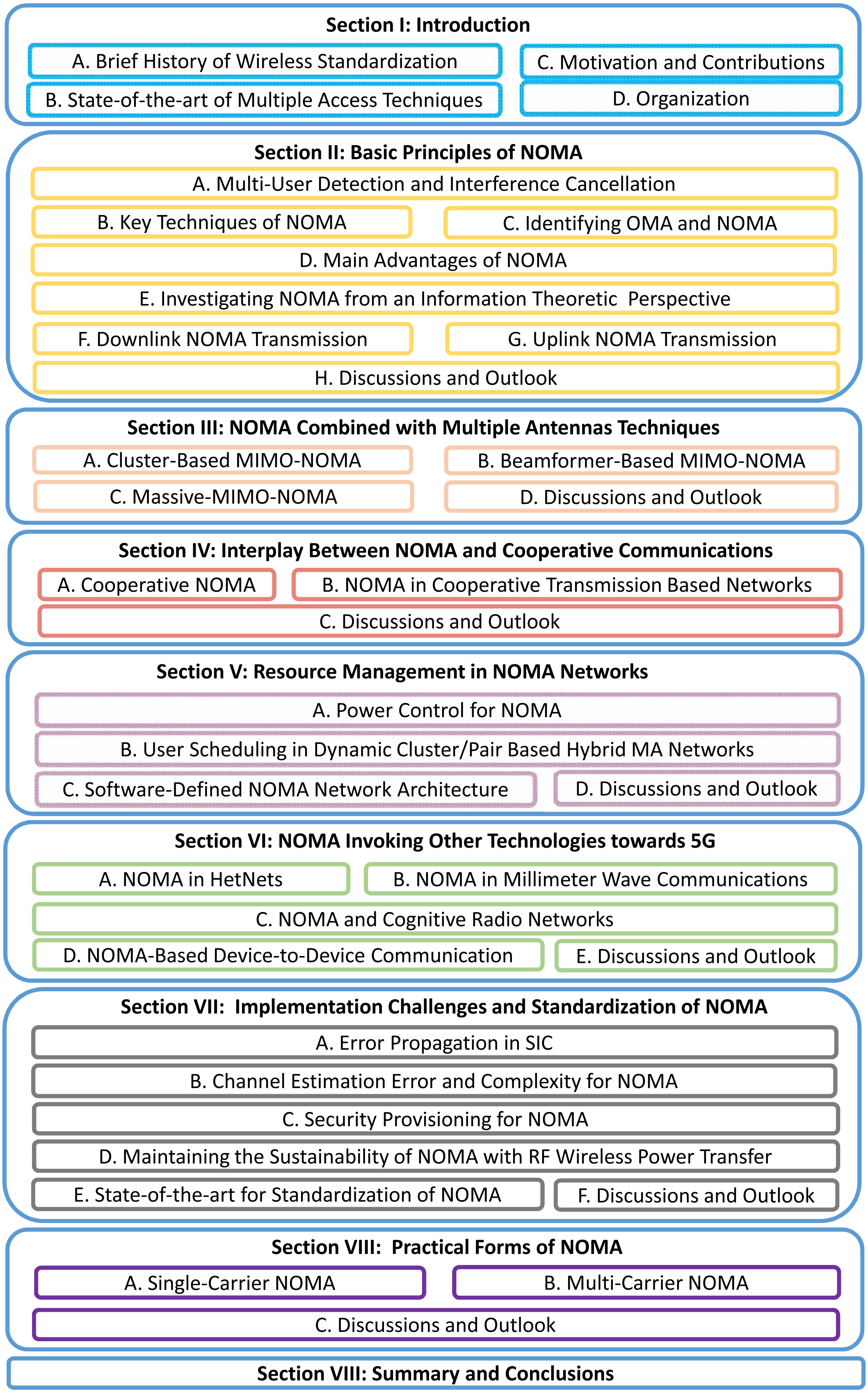


Fig. 3. ***Condensed overview of this survey on NOMA.***

on either maximum-likelihood detection or predetection processing, rather than on IC techniques. This development is attributed to the fact that joint detection has also been developed for asynchronous networks [53]. As a further advance, the proposed SAIC technique has had successful field trial results in the GSM era in terms of suppressing the downlink intercell interference [31].

2) Interference Cancellation: Commonly, the multiuser IC techniques can be divided into two main categories, namely pre-interference cancelation (pre-IC) and post-interference cancellation (post-IC) [54]. More specifically, pre-IC techniques are employed at the transmitter side by suppressing the interference by precoding approaches, such as the famous dirty paper coding (DPC) [55] upon exploiting the knowledge of the channel state information (CSI) at the transmitter. By contrast, the post-IC techniques are usually used at the receiver side for canceling the interference. The post-IC approach can be further divided into two categories, which are parallel and successive [33]. If we carry out accurate power control to ensure that all received signals are similar, PIC outperforms SIC. By contrast, SIC works better, when the received powers are different [28] because the strongest user's signal can be detected first. The detected bit is the remodulate and its interference is deducted from the received signal. Repeating this action in a sequential order gives us the clean weakest signal. It is worth noting that in addition to the performance versus complexity tradeoff, there are also a variety of other tradeoffs between PIC and SIC.

There are also some recent IC techniques. Based on PIC, Sawahashi *et al.* [56] of NTT DoCoMo have developed a coherent three-stage interference canceller for DS-CDMA systems, which is capable of employing accurate pilot symbol-assisted channel estimation and IC. The corresponding performance was experimentally evaluated with the aid of a multipath fading simulator [56]. Regarding the LTE heterogeneous networks (HetNets), some companies such as Qualcomm and Fujitsu have also developed powerful IC techniques. More particularly, at the 3GPP meeting in 2008, Qualcomm proposed a new cell selection scheme, termed cell range extension [57], for allowing more small cell offloading to manage the strong macrocell interference at the handsets [58]. Li *et al.* [59] of Fujitsu proposed a frequency-domain cell-specific reference-signal-aided SIC scheme for 3GPP LTE-Advanced Rel-11 systems. While the aforementioned IC techniques mainly focus on the digital domain, there are also pioneering analog domain solutions. In full-duplex systems, self-interference cancellation (IC) is the key issue [60]. Duarte *et al.* [61] proposed an active analog cancellation-aided full-duplex architecture to cancel the self-interference before the analog-to-digital converter (ADC). The suppression performance of [61] was also characterized with the aid of experimental results.

B. Key Technologies of NOMA

Again, the basic principles of NOMA techniques rely on the employment of superposition coding (SC) at the transmitter and successive interference cancelation (SIC) techniques at the receiver. In fact, neither of these two techniques are new; their roots can be found in the existing literature [25], [26], [28], [33], [35], [62]–[71].

1) Superposition Coding: First proposed by Cover as early as in 1972 [25], the elegant idea of SC is regarded as one of the fundamental building blocks of coding schemes conceived for achieving the capacity of a scalar Gaussian BC [62]. More particularly, it was theoretically demonstrated that SC is capable of approaching the capacities of both the Gaussian BC and of the general BC as defined by Bergmans' paper published as early as 1973 [26] and by Gallager [72]. The fundamental concept of SC is that it is capable of encoding a message for a user associated with poor channel conditions at a lower rate and then superimpose the signal of a user having better channel conditions on it. Inspired by the solid foundations laid down from an information theory perspective, researchers became motivated to apply SC to diverse channels, such as interference channels [63], relay channels [64], MA channels [65], and wiretap channels [66]. While the aforementioned contributions richly motivate the use of SC from a theoretical perspective, further research was required for evolving this technique from theory to practice [35], [73]. Specifically, Vanka *et al.* [35] designed an experimental platform using a software-defined radio (SDR) system to investigate the performance of SC. The set of achievable rate pairs under a specific packet-error constraint was determined.

2) Successive Interference Cancelation: It has been widely exploited that the network capacity can be substantially improved with the aid of efficient interference management, hence SIC is regarded as a promising IC technique in wireless networks. By invoking the following procedure, it enables the user having the strongest signal to be detected first, who has hence the least interference-contaminated signal. Then, the strongest user reencodes and remodulates its signal, which is then subtracted from the composite signal. The same procedure is followed by the second strongest signal, which has in fact become the strongest signal. When all but one of the signals was detected, the weakest user decodes its information without suffering from any interference at all.

Compared to the classic MUD techniques, we summarize the key advantages of SIC as follows.

- In contrast to the MUD techniques of CDMA [74] or MIMO systems [75] where there are multiple observations at receivers [27], [76], [77], power-domain NOMA usually has a single observation at each receiver. In other words, rather than using join-detection-aided MUD considering all users simultaneously, the SIC technique operates in an iterative

manner, which imposes a lower hardware complexity at the receiver than the joint decoding approach [33].[3] The number of multiplications and additions required was summarized in [74].
- It has been demonstrated that SIC is capable of approaching the boundaries of the capacity region of both the BCs and of MA channels [32].
- As discussed before, SIC has a better performance than PIC when the received powers are different [28], which is more suitable for power-domain NOMA supporting the users via different power levels.

Hence, SIC has been widely studied and diverse versions have been employed in practical systems, such as CDMA [28] as well as SDMA [74] and the vertical-bell laboratories layered space time (V-BLAST) [67]. Furthermore, SIC has also been exploited in several practical scenarios, such as multiuser MIMO networks [68], multihop networks [69], random access systems [70], and in large-scale networks modeled by stochastic geometry [71], just to name a few. A particularly important fact concerning SIC is that it has been implemented in commercialized systems, such as IEEE 802.15.4. Another practical implementation is to use the SIC-aided spatial division multiplexing (SDM) detector in SDM-assisted orthogonal frequency-division multiplexing (OFDM) systems for signal detection by distinguishing the users with the aid of their unique, user-specific impulse responses [3].

C. Identifying OMA and NOMA

We commence with the mathematical definition of NOMA. The general definition of NOMA is a MA technique, which allows multiple users to simultaneously occupy the same time-and-frequency resource. Based on this definition, we may have power-domain NOMA, code-domain NOMA, and spatial-domain NOMA, as mentioned in Section I. In this treatise, we focus on power-domain NOMA. The narrow sense definition of power-domain NOMA is to superimpose the signals in the same time-and-frequency resource at different power levels, and then to adopt SIC techniques at receivers for interference cancelation.[4]

For illustrating the relationship between NOMA and OMA mathematically, below we provide a simple analytical characterization by examining the achievable performance with the aid of signal-to-noise ratio (SNR) expressions. Let us consider two-user downlink NOMA transmission. The channel coefficients of user m and user n are h_m and h_n. Let us denote the transmit SNR at the BS by ρ and assume that we have $|h_m|^2 < |h_n|^2$ without loss of generality.

[3]Actually, if we consider the broad sense of power-domain NOMA, most of aforementioned MUD techniques can be potentially applied, depending on the receiving power of users.

[4]The broad sense definition of power-domain NOMA constitutes all MA techniques which apply SC technique at transmitters for power multiplexing. In this sense, some code-domain NOMA can also regarded as generalized power-domain NOMA techniques, which will be detailed in Section VII.

- **OMA:** According to Shannon's capacity theorem, when power control is used, the throughput of OMA can be expressed for user m and user n as [32]

$$R_m^{\mathrm{OMA}} = \beta \log_2\left(1+\frac{\alpha_m \rho}{\beta}|h_m|^2\right) \tag{1}$$

and

$$R_n^{\mathrm{OMA}} = (1-\beta)\log_2\left(1+\frac{\alpha_n \rho}{1-\beta}|h_n|^2\right) \tag{2}$$

respectively, where α_m and α_n are the PA coefficients and satisfy $\alpha_m + \alpha_n = 1$, and β is the resource allocation coefficient, having units of hertz for frequency or seconds for time. For the case that the power control is not considered at the BS, we set $\alpha_m/\beta = \alpha_n/(1-\beta) = 1$. Then (1) and (2) can be rewritten as

$$R_m^{\mathrm{OMA}} = \beta \log_2\left(1+\rho|h_m|^2\right) \tag{3}$$

and

$$R_n^{\mathrm{OMA}} = (1-\beta)\log_2\left(1+\rho|h_n|^2\right). \tag{4}$$

- **NOMA:** Regarding NOMA, the throughput of user m and user n is given by [32]

$$R_m^{\mathrm{NOMA}} = \log_2\left(1+\frac{\rho\,\alpha_m|h_m|^2}{1+\rho\,\alpha_n|h_m|^2}\right) \tag{5}$$

and

$$R_n^{\mathrm{NOMA}} = \log_2\left(1+\rho\,\alpha_n|h_n|^2\right) \tag{6}$$

respectively.

In order to gain more insights into the spectral efficiency advantage of NOMA over OMA, we investigate the following special case as an example. At high SNRs, assuming that the time/frequency resources are equally allocated to each user, based on (3)–(6), the sum throughput of OMA and NOMA can be expressed as $R_{\mathrm{sum},\infty}^{\mathrm{OMA}} \approx \log_2\left(\rho\sqrt{|h_m|^2|h_n|^2}\right)$ and $R_{\mathrm{sum},\infty}^{\mathrm{NOMA}} \approx \log_2\left(\rho|h_n|^2\right)$, respectively. Then, we can express the sum throughput gain of NOMA over OMA as follows:

$$R_{\mathrm{sum},\infty}^{\mathrm{Gain}} = R_{\mathrm{sum},\infty}^{\mathrm{NOMA}} - R_{\mathrm{sum},\infty}^{\mathrm{OMA}} = \frac{1}{2}\log_2\left(|h_n|^2/|h_m|^2\right). \tag{7}$$

When we have $|h_m|^2 < |h_n|^2$, the sum throughput of NOMA is higher than that of OMA, and this gain is imposed when the channel conditions of the two users become more different. Chen *et al.* [78] provided a rigorous mathematical proof to demonstrate that NOMA always outperforms the conventional OMA scheme.

D. Main Advantages of NOMA

While both SC and SIC continue to mature in terms of their theoretical and practical aspects, NOMA is also maturing. Hence, it has been proposed for next-generation networks. By invoking the SC technique, the BS transmits the superposition coded signals of all users. Then, the channel gains of the users are sorted in the increasing or decreasing order. In the traditional OMA schemes, strongest user benefits from the highest power, which is not always the case for NOMA. The NOMA transmission schemes exhibit the following main advantages.

- **High bandwidth efficiency:** NOMA exhibits a high bandwidth efficiency and hence improves the system's throughput, which is attributed to the fact that NOMA allows each RB (e.g., time/frequency) to be exploited by multiple users [36].
- **Fairness:** A key feature of NOMA is that it allocates more power to the weak users. By doing so, NOMA is capable of guaranteeing an attractive tradeoff between the fairness among users in terms of their throughput. There are sophisticated techniques of maintaining fairness for NOMA, such as the intelligent PA policies of [24] and [79] and the cooperative NOMA scheme of [80], which will be detailed in this paper later.
- **Ultrahigh connectivity:** The future 5G system is expected to support the connection of billions of smart devices in the IoT [81]. The existence of NOMA offers a promising design alternative for efficiently solving this nontrivial task by exploiting its nonorthogonal characteristics. More specifically, in contrast to convectional OMA, which requires the same number of frequency/time RBs as the number of devices, NOMA is capable of serving them by using less RBs.
- **Compatibility:** Again, from a theoretical perspective, NOMA can be invoked as an "add-on" technique for any existing OMA techniques, such as TDMA/FDMA/CDMA/OFDMA, due to the fact that it exploits a new dimension, namely the power domain. Given the mature status of SC and SIC techniques both in theory and practice, NOMA may be amalgamated with the existing MA techniques.
- **Flexibility:** Compared to other existing NOMA techniques, such as MUSA [21], PDMA [22], and SCMA [16], [23], [82], NOMA is conceptually appealing and yields a low-complexity design. In fact, the fundamental principles of the aforementioned MA schemes and NOMA are very similar, relying on allocating multiple users to a single RB. Considering the comparison of NOMA and SCMA as an example, SCMA can be regarded as a variation of NOMA which integrates appropriate coding, modulation, and subcarrier allocation.

E. Investigating NOMA From an Information-Theoretic Perspective

Having considered the potential benefits of NOMA, it is important to investigate its performance gain also from an information-theoretic perspective. In fact, the concept of NOMA may also be interpreted as a special case of SC in the downlink broadcast channel (BC). More particularly, by using SC, the capacity region of a realistic imperfect discrete memoryless BC was established by Cover [25]. As an extension of [25], Bergmans found the Gaussian BC capacity region of single-antenna scenarios [83]. Inspired by [25] and [83], several researchers began to explore the potential performance gain from an information-theoretic perspective [84]–[86]. Xu *et al.* [84] developed a new evaluation criterion for quantifying the performance gain of NOMA over OMA. More specifically, considering a simple two-user single-antenna scenario in conjunction with the Gaussian BC, the comparison of TDMA and NOMA in terms of their capacity region was provided in [84]. The analytical results showed that NOMA is capable of outperforming TDMA both in terms of the individual user rates and the sum rate. In [85], Shieh and Huang focused their attentions on examining the capacity region of downlink NOMA by systematically designing practical schemes and investigated the gains of NOMA over OMA by designing practical encoders and decoders. Furthermore, by proposing to use NOMA for relaying broadcast channels (RBCs) for the same of achieving a performance enhancement, So and Sung [86] examined the achievable capacity region of the RBC upon invoking both decode-and-forward (DF) relaying, as well as compress-and-forward (CF) relaying with/without dirty-paper coding (DPC). Regarding the family of MIMO-NOMA systems, in [87] the achievable capacity region of multiuser MIMO systems is investigated upon invoking iterative linear minimum mean square error (LMMSE) detection.

In contrast to the above research contributions considering NOMA in additive white Gaussian noise (AWGN) channels, Xing *et al.* [88] investigated the performance of a two-user case of downlink NOMA in fading channels to exploit the time-varying nature on multiuser channels. As shown in Fig. 4, NOMA achieves a superior performance

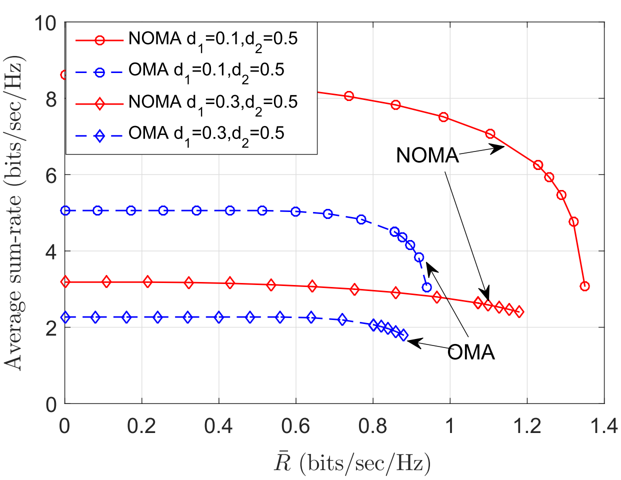


Fig. 4. ***Illustration of the sum rate versus the minimum rate constraints with full CSIT knowledge for different user distances. The full parameter settings can be found in [88].***

over OMA for different distance settings, where the average sum rate was maximized subject to a minimum average individual rate constraint.

F. Downlink NOMA Transmission

Downlink NOMA transmission employs the SC technique at the BS for sending a combination of the signals and the SIC technique may be invoked by the users for interference cancelation, as shown in Fig. 5(b). Numerous valuable contributions have investigated the performance of NOMA in terms of downlink transmission [36], [37], [79], [89], [90]. In [36], a two-user NOMA downlink transmission relying on SIC receivers was proposed. Upon considering a range of further practical conditions in terms of the key link-adaptation functionalities of the LTE, the system-level performance was evaluated in [89] and [90]. A more general NOMA transmission scheme was proposed in [37], which considered a BS communicating with M randomly deployed users. It was demonstrated that NOMA is capable of achieving superior performance compared to OMA in terms of both its outage probability and its ergodic rate. In [79], the fairness issues were addressed by adopting appropriate PA coefficients for the M users in a general NOMA downlink transmission scenario.

Motivated by reducing the signaling overhead required for CSI estimation, some researchers embarked on investigating the performance of downlink NOMA transmissions using partial CSI at the transmitter [91]–[93]. More explicitly, Yang *et al.* [91] studied the outage probability of NOMA by assuming either imperfect CSI or second-order-statistics-based CSI, respectively. In [92], by assuming the knowledge of statistical CSI, Shi *et al.* investigated the outage performance of NOMA by jointly considering both the decoding order selection and the PA of the users. In [93], by assuming that only the average CSI was obtained at the BS, Cui *et al.* studied both the optimal decoding order as well as the optimal PA of the users in downlink NOMA systems. Both the transmit power of the BS and rate fairness of users were optimized. By considering only a single-bit feedback of the CSI from each user to the BS, the outage performance of a downlink transmission scenario was studied by Xu *et al.* [94]. Based on the analytical results derived, the associated dynamic PA optimization problem was solved

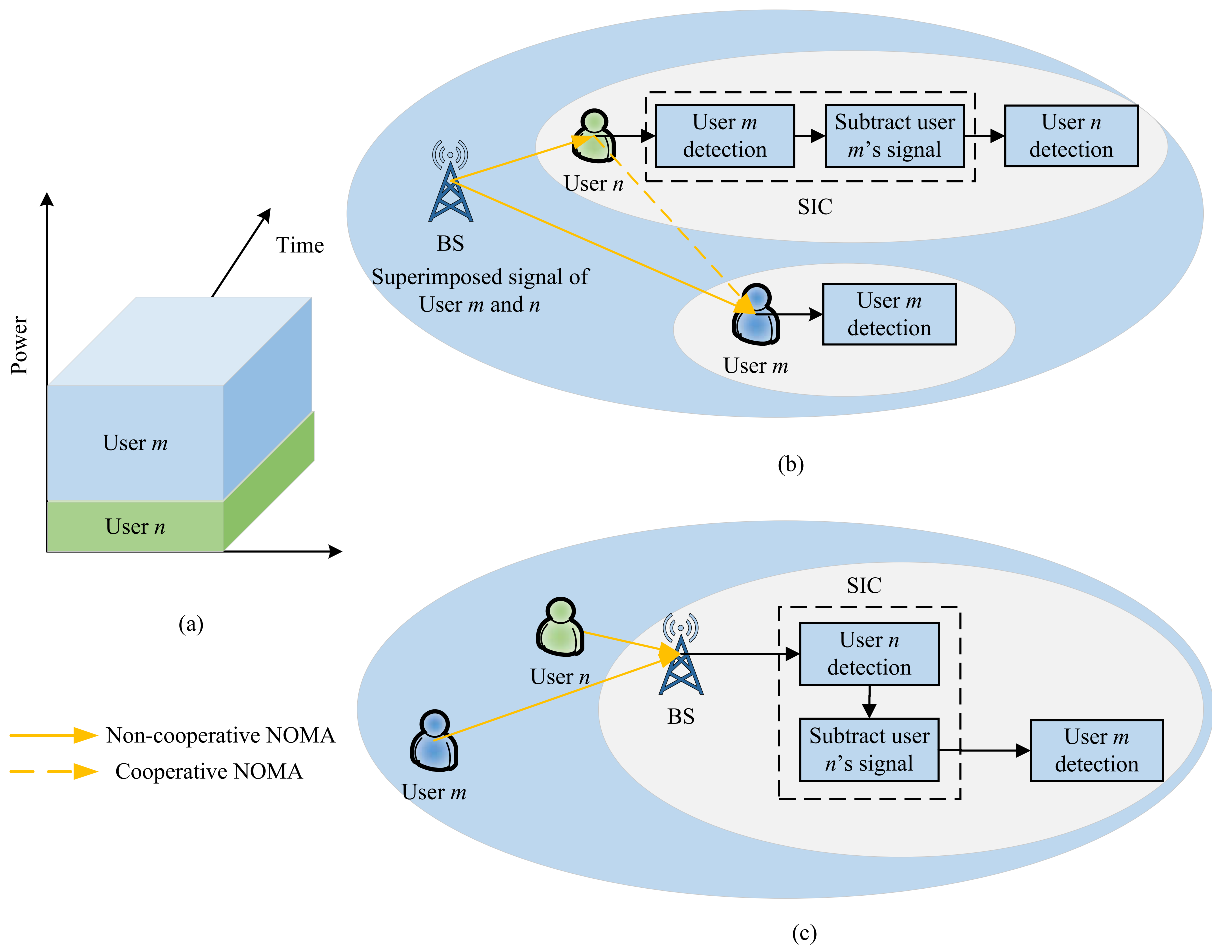


Fig. 5. *Illustration of NOMA transmission. (a) Power multiplexing NOMA. (b) Downlink NOMA transmission. (c) Uplink NOMA transmission.*

by minimizing the outage probability. In [95], Zhang *et al.* investigated an energy-efficiency optimization problem in downlink NOMA systems in conjunction with different data rate requirements of the users. It was shown that NOMA is also capable of outperforming OMA in terms of its energy efficiency.

Apart from wireless communication systems, the potential use of NOMA in other communication scenarios has also attracted interest. A pair of representative communication scenarios are visible light communication (VLC) [96]–[99] and quantum-aided communication [100]. To elaborate, in [97], Marshoud *et al.* applied the NOMA technique in the context of VLC downlink networks. By doing so, the achievable throughput was enhanced. It is worth pointing out that although the potential performance enhancement is brought by invoking NOMA into VLC networks, the hardware implementation complexity is also increased at transmitters and receivers as SC and SIC techniques are adopted. In [101], Botsinis *et al.* considered quantum-assisted multiuser downlink transmissions in NOMA systems, where a pair of bioinspired algorithms were proposed.

G. Uplink NOMA Transmission

In uplink NOMA transmission, multiple users transmit their own uplink signals to the BS in the same RB, as shown in Fig. 5(c). The BS detects all the messages of the users with the aid of SIC. Note that there are several key differences between uplink NOMA and downlink NOMA, which are listed as follows.

- **Transmit power:** In contract to downlink NOMA, the transmit power of the users in uplink NOMA does not have to be different, it depends on the channel conditions of each user. If the users' channel conditions are significantly different, their received SINR can be rather different at the BS, regardless of their transmit power.
- **SIC operations:** The SIC operations and interference experienced by the users in the uplink NOMA and downlink NOMA are also rather different. More specifically, as shown in Fig. 5(b) for downlink NOMA, the signal of user n is decontaminated from the interference imposed by user m, which is achieved by first detecting the stronger signal of user m, remodulating it and then subtracting it from the composite signal. It means that SIC operation is carried out on a strong user in downlink for canceling the weak user's interference. By contrast, in uplink NOMA, SIC is carried out at the BS to detect strong user n first by treating user m as interference, as shown in Fig. 5(c). Then, it remodulates the recovered signal and subtracts the interference imposed by user n to detect user m.
- **Performance gain:** The performance gain of NOMA over OMA is different for downlink and uplink. The capacity region of NOMA and OMA both for downlink and uplink is illustrated in [43, Fig. 1]. The capacity region of NOMA is outside OMA, which means that the use of NOMA in downlink has superior performance in terms of throughput. While in uplink, NOMA mainly has the advantages in terms of fairness, especially compared to the OMA with power control.

Recently, there have been less research contributions on uplink NOMA than on downlink NOMA [102]–[108]. The downlink and uplink NOMA throughput regions were quantified by Higuchi and Benjebbour [102] for both symmetric and asymmetric channels. They demonstrated that the performance benefits of NOMA over OMA become more significant, when the channel conditions of the users are more different. Al-Imari *et al.* [103] designed a novel NOMA scheme for the classic OFDMA uplink. The efficiency gains of the proposed scheme over the conventional OMA scheme were quantified in terms of both fairness and spectral efficiency. As an extension of [103], a novel iterative multiuser detection was designed in [104] for further enhancing the performance of the NOMA uplink. In an effort to reduce the implementation complexity, Chen *et al.* [105] conceived a user-pairing policy for the NOMA uplink. In [106], Zhang *et al.* proposed a novel power control strategy for the NOMA uplink and investigated both the outage probability and the delay-limited sum rate. In [108], a wirelessly powered uplink NOMA transmission scheme was investigated by Diamantoulakis *et al.* by applying a harvest-then-transmit-style EH protocol. A general MIMO-NOMA framework applying stochastic geometry both for downlink and uplink transmission was proposed by Ding *et al.* [107].

H. Discussions and Outlook

Given the increasing number of research contributions on NOMA, its advantages are becoming increasingly clear, especially in terms of its bandwidth efficiency, energy efficiency, and fairness. Stochastic geometry constitutes a powerful mathematical and statistical tool, which is capable of capturing the topological randomness of the networks and hence provides tractable analytical results for the average network behavior [109]. This is particularly essential in large-scale networks supporting a large number of randomly distributed BSs and mobile users. At the time of writing the stochastic-geometry-based modeling of NOMA networks is still in its infancy [37], [107], [110]–[114]. The derivation of closed-form expressions for obtaining deep insights remains an open challenge. Hence, further research efforts are required for analyzing the average performance of multiuser multicell NOMA networks [42], [115] in order to provide inspiration both for the long-term investigations in academia and practical guidelines for industry.

Another important issue to be addressed in the NOMA context is that of the ordering of users as required by SIC. Order statistics [116] can be a helpful tool to assist the associated performance analysis for the sake of obtaining insightful guidelines.

Most of the existing contributions are mainly focused on the theoretical performance analysis or optimization. Most of the implementation issues, such as the effect of imperfect CSI on both SISO-NOMA [117] and on MISO-NOMA [118]–[120], on limited channel feedback [121], [122], on efficient receiver designs [123], and its combination with advanced adaptive coding and modulation schemes, etc., are still open areas. Correspondingly, further research is expected in the aforementioned areas to conceive a holistic architecture for NOMA.

III. NOMA COMBINED WITH MULTIPLE-ANTENNA TECHNIQUES

Multiple-antenna techniques are of significant importance, since they offer the extra dimension of the spatial domain, for further performance improvements. The application of multiple-antenna techniques in NOMA has attached substantial interest both from academia [40], [41], [102], [107], [124]–[134] and from industry [36], [102], [135], [136]. The distinct NOMA features such as channel ordering and PA inevitably require special attention in the context of multiple antennas. More specifically, in contrast to the SISO-NOMA scenarios whose channels are all scalars, the channels of MIMO-NOMA scenarios are represented in the form of matrices, which makes the power-based ordering of users rather challenging. As a consequence, conceiving an appropriate beamforming/precoding design is essential for multiple-antenna-aided NOMA systems. NOMA relying on beamforming (BF) constitutes an efficient technique of improving the bandwidth efficiency by exploiting both the power domain and the angular domain. There are two popular MIMO-NOMA designs, namely 1) the cluster-based (CB) MIMO-NOMA design; and 2) the beamformer-based (BB) MIMO-NOMA design, which will be introduced in the following. Table 4 summarizes some of the existing contributions on NOMA applying multiple antennas and illustrates their comparison.

A. Cluster-Based MIMO-NOMA

One of the popular NOMA designs is associated with the cluster-based structure, partitioning users into several different clusters. Explicitly, as shown in Fig. 6, the NOMA users are partitioned into M clusters and each cluster consists of L_m users, where $m \in \{1, 2, \ldots, M\}$. Then, we design appropriate beams for the corresponding clusters. Upon applying effective transmit precoding and detector designs, it becomes possible to guarantee that the beam associated with a particular cluster is orthogonal to the channels of users in other clusters. Hence, the intercluster interference can be efficiently suppressed. When considering each cluster in isolation, there is a difference among the users' channel conditions, hence we are faced again with the conventional NOMA scenarios. Thus, SIC can be readily invoked for mitigating the intracluster interference between users of the same cluster. Recently, many important research contributions investigated beamforming-aided NOMA [40], [41], [102], [107], [127], [138].

Specifically, Choi [40] proposed two-stage multicast zero-forcing (ZF)-based beamforming for downlink intergroup/cluster interference mitigation, where the total transmit power of each group/cluster was minimized during the second stage. Higuchi and Benjebbour [102] invoked receive beamforming at the NOMA users and a transmit beamformer at the BS. Higuchi and Kishiyama [135] then proposed a novel scheme, which combined open-loop random beamforming in conjunction with intrabeam SIC for downlink NOMA transmission. However, random beamforming fails to guarantee a constant QoS at the users' side. To overcome this limitation, in [41], Ding *et al.* proposed a TPC and detection scheme combination for a cluster-based downlink MIMO-NOMA scenario relying on fixed PA. By adopting this design, their MIMO-NOMA system

Table 4 Important Contributions on Multiple-Antenna-Aided NOMA. "DL" and "UL" Represent Downlink and Uplink, Respectively. "BF" and "OP" Represent Beamforming and Outage Probability, Respectively. "SU" and "MU" Represent Two-User and Multiple-User Cases. The "Sum-Rate Gain" Implies the Gain Brought by Invoking NOMA Technique Over Conventional OMA Technique in Terms of Sum Rate

Ref.	Scenarios	Direction	Main Objectives	Techniques/Characteristics
[40]	MU-MISO	DL	Transmit power	Two stage BF
[102]	MU-MISO	DL/UL	Throughput	Less constraint at the BS
[125]	MU-MISO	DL	Sum-rate	Ordered vector constraints
[135]	MU-MIMO	DL	Sum-rate	Random BF
[41]	MU-MIMO	DL	Sum-rate gain	No need for perfect CSIT at the BS
[107]	MU-MIMO	DL/UL	OP	Obtaining larger diversity gain
[127]	SU-MIMO	DL	OP	Creating channel differences artificially
[124]	SU-MIMO	DL	Ergodic capacity	Optimal/sub-optimal power allocation
[126]	SU-MIMO	DL	Sum-rate	Low complexity decoding
[137]	MU-MIMO	DL	OP	Limited feedback requirements
[138]	MU-MIMO	DL	Max-min rate	Addressed max-min fairness issue
[139]	MU-MISO	DL	Secrecy OP	Generating artificial noise

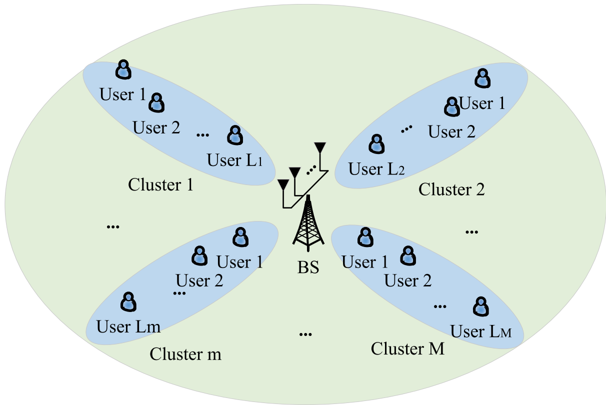


Fig. 6. ***Illustration of cluster-based MIMO-NOMA.***

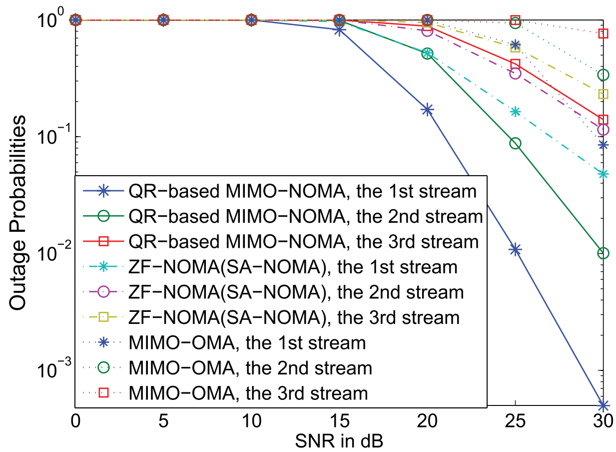


Fig. 7. ***Comparison of MIMO systems with various multiple-access techniques using a two-user case. $M = N = 3$, where M is the number of antennas at transmitters, and N is the number of antennas at receivers. As such, each user is to receive three data streams from the BS. $R_m = 1.2$ bit per channel use (BPCU) and $R_n = 5$ BPCU are the targeted data rates of user m and user n, respectively.***

can be decomposed into several independent SISO NOMA arrangements. Furthermore, in order to establish a more general framework considering both downlink and uplink MIMO-NOMA scenarios, the so-called signal alignment (SA) technique was proposed in [107]. Stochastic-geometry-based tools were invoked to model the impact of the NOMA users' locations [107]. In contrast to the research contributions in [41] and [107], which are intercluster interference free design, an intercluster interference allowance design for CB MIMO-NOMA was proposed in [140]. More particularly, a user who experiences the highest channel condition was selected as a cluster head and was capable of completely canceling all intracluster interference in [140]. Note that the existing NOMA designs have routinely relied on assuming different channel conditions for the different users, which is however a somewhat restrictive assumption. In order to circumvent this restriction, Ding *et al.* [127] designed a new MIMO-NOMA scheme, which distinguishes the users according to their QoS requirements with particular attention to IoT scenarios for the sake of supporting the SIC operation. Furthermore, they compared this new MIMO-NOMA design to two NOMA schemes, which order users according to the prevalent channel conditions. More particularly, the ZF-NOMA scheme of [41] and the SA-NOMA scheme [107] were used as benchmarks in [127]. Fig. 7 illustrates the outage probability defined as the probability of erroneously detecting the message intended for user m in the ith data stream, $i = 1, 2, 3$ at user n, where the QR decomposition is used to augmenting the differences between the users' effective channel conditions according to the associated QoS requirements. As shown in Fig. 7, the QR-based MIMO-NOMA scheme is capable of outperforming both ZF-NOMA and SA-NOMA[5] as well as MIMO-OMA,[6] since it exploits the heterogeneous QoS requirements of different users and applications. In [138], the fairness issues of the MIMO-NOMA scenario were addressed by applying appropriate user allocation algorithms among the clusters and dynamic PA algorithms within each cluster.

B. Beamformer-Based MIMO-NOMA

Another technique of implementing MIMO-NOMA is to assign different beams to different users, as shown in Fig. 8. By doing so, the QoS can be satisfied by calculating the beamformer weights in a predefined order, commencing with the most demanding QoS requirement. By adopting this approach, several contributions have been made in terms of MIMO-NOMA. Considering the illustration of Fig. 8 as an example, user 1 to user N occupy the same RB, similarly to user $(N + 1)$ to user $(N + M)$. Again, within the same RB we may employ SIC at each user, according to the particular ordering of the different users' received signal power. In [125], Hanif *et al.* solved the downlink sum-rate maximization problems, which resulted in obtaining the

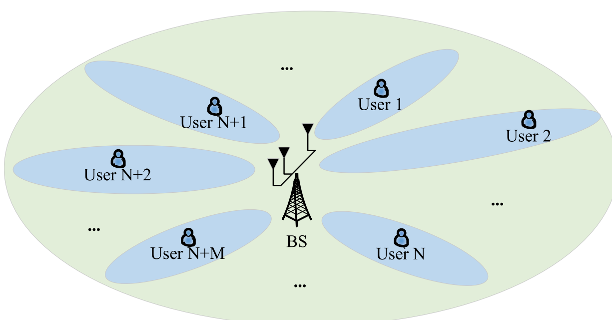


Fig. 8. ***Illustration of beamformer-based MIMO-NOMA.***

[5]Note that when $M = N$, ZF-NOMA achieves the same performance as SA-NOMA [127].

[6]Fig. 7 is focused on the performance of user n, since the QoS requirements have been guaranteed with the aid of appropriate PA [127].

corresponding optimal TPC vectors. Sun *et al.* [124] first investigated the power optimization problem constructed for maximizing the ergodic capacity and then showed that their proposed MIMO-NOMA schemes are capable of achieving significantly better performance than OMA. In an effort to reduce the decoding complexity imposed at the users, a layered-transmission-based MIMO-NOMA scheme was proposed by Choi [126], who also investigated the associated PA problem. It was demonstrated that upon invoking this layered transmission scheme, the achievable sum rate increases linearly with the number of antennas.

C. Massive MIMO-NOMA

Massive MIMO may be considered as one of the key technologies [9] in 5G systems as a benefit of improving both the received SNR and the bandwidth efficiency. It was shown in [141] that massive MIMO is capable of substantially increasing both the capacity as well as the energy efficiency. These compelling benefits of massive MIMO sparked off the interest of researchers also in the context of NOMA. In [137], Ding and Poor conceived a two-stage TPC design for implementing massive MIMO-NOMA. Particularly, a beamformer was adopted for serving to a cluster of angularly similar users and then they decomposed the MIMO-NOMA channels into a number of SISO-NOMA channels within the same cluster. A one-bit CSI feedback scheme was proposed for maintaining a low feedback overhead and a low implementation complexity.

D. Discussions and Outlook

Although numerous research contributions have been made in the field of MIMO-NOMA scenarios, most of the existing treatises assumed that the perfect CSI is known at the transmitter. Under this assumption, the TPC matrix may be readily designed according to the ordered channels between the users and the BS based on their received power. However, in practical wireless communication systems, obtaining the channel state information at the transmitter (CSIT) is not a trivial problem, which requires the classic-pilot-based training process. This is particularly important for massive MIMO-NOMA scenarios, where a large number of antennas is used at the BS, hence potentially imposing an excessive pilot overhead and complexity. Additionally, when the Doppler frequency is doubled, so is the pilot overhead and the computational complexity. Although a low-complexity massive MIMO-NOMA scheme relying on a one-bit feedback flag was proposed in [137], naturally, the performance was degraded compared to the case of perfect CSIT. Motivated by all the aforementioned problems, efficient CSIT estimation techniques are required for supporting massive MIMO-NOMA systems. By exploiting the sparsity of the user's channel matrix, compressive sensing (CS) can be invoked for reducing the pilot overhead [142], as well as for reducing the feedback overhead transmitted from the downlink receivers to the BS. In the CS-based CSI estimation conceived for massive MIMO-NOMA, machine learning may be invoked both for identifying the measurement matrix design of CS and for the sparse domain selection process of channel estimation in massive MIMO-NOMA. Explicitly, this is another valuable research topic.

The channel estimation problem can also be jointly considered with the data detection, where soft extrinsic information is exchanged between the receiver blocks, such as the channel estimator, data detector, and channel decoder. In this context it is worth emphasizing that once some of the soft estimates become sufficiently reliable, then they can be subjected to hand decisions and then used as additional pilots in the context of decision-directed channel estimation [74]. Finally, noncoherent MIMO-NOMA dispensing with channel estimation has to be investigated. Additionally, when considering the practical implementation issues, the coexistence of NOMA and MIMO inevitably necessitates the redesign of receivers because of the distinct characteristics of SIC.

IV. INTERPLAY BETWEEN NOMA AND COOPERATIVE COMMUNICATIONS

Due to the hostile characteristics of wireless channels, the attenuation of the signals may vary dramatically during their transmission. As a countermeasure, cooperative communication is a particularly attractive technique of offering cooperative diversity gains, hence extending the network's coverage [143], [144]. In this section, we summarize the NOMA contributions on cooperative communications, mainly from the perspective of promising cooperative NOMA techniques and the application of NOMA in cooperative networks.

A. Cooperative NOMA

The key idea behind cooperative NOMA is to rely on strong NOMA users acting as DF relays to assist weak NOMA users. Still considering the two-user downlink transmission of Fig. 5(b) as our example, cooperative NOMA requires two time slots for its transmission. The first slot is for the direct transmission phase, which is the same as the noncooperative NOMA of Fig. 5(b) indicated by solid lines. During the second time slot, which is the cooperative phase, user n will forward the decoded message to user m by invoking the DF relaying protocol of Fig. 5(b) indicated by the dashed line. Proposed by Ding *et al.* [80], this novel concept intrigued researchers, since cooperative NOMA fully benefits from SIC and DF decoding. In [80], a general downlink NOMA transmission scenario was considered, where the BS supported M users with the aid of cooperative NOMA protocols relying on M slots. In an effort to seek a more efficient cooperative NOMA protocol, Liu *et al.* [110] proposed a new EH-assisted cooperative NOMA scheme. A sophisticated stochastic-geometry-based model was invoked for evaluating the system's performance and user pairing was adopted for reducing the implementation complexity. Compared to conventional NOMA, the key advantages of cooperative NOMA transmissions can be summarized as follows.

- **Low system redundance:** Again, upon applying SIC techniques in NOMA, the message of the weak user has already been decoded at the strong user, hence it is natural to consider the employment of the DF protocol for weak signal. Explicitly, the weak signal can be remodulated and retransmitted from a position closer to the destination.
- **Better fairness:** A beneficial feature of cooperative NOMA is that the reliability of the weak user is significantly improved. As a consequence, the fairness of NOMA transmission can be improved [79], particularly in the scenarios when the weak user is at the edge of the cell illustrated by the BS.
- **Higher diversity gain:** Cooperative NOMA is capable of achieving an improved diversity gain for the weak NOMA user, which is an effective technique of overcoming multipath fading. It was analytically demonstrated [80] that the diversity gains of the weak users in cooperative NOMA are the same as those of the conventional cooperative networks, even for using EH relays [110].

Fig. 9 illustrates the superior performance of cooperative NOMA over noncooperative NOMA as well as over OMA in terms of its outage probability. It is noted that cooperative NOMA achieves a higher diversity gain than noncooperative NOMA and OMA, which demonstrates the effectiveness of cooperative NOMA, as aforementioned. Note that cooperative NOMA constitutes one of many techniques of improving the transmission reliability of NOMA networks, especially for the weak NOMA users who have poor channel conditions. There are also other techniques for enhancing the performance of NOMA networks, which will be detailed in Section IV-B. It is also worth pointing out that the performance of cooperative NOMA is related to the error propagation issue for SIC, which will be introduced in Section VII.

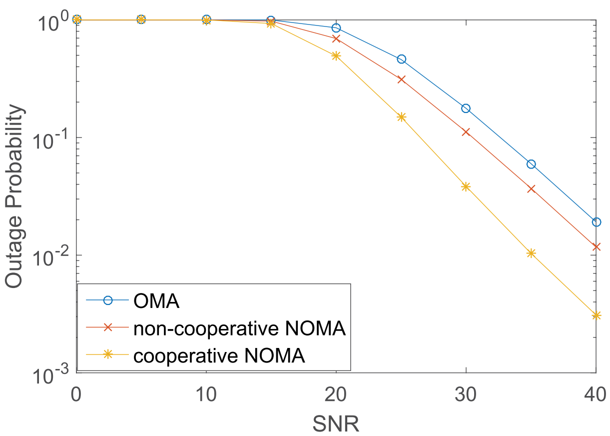


Fig. 9. ***Performance comparison between cooperative NOMA and noncooperative NOMA. The BS is located at (0, 0), user m is located at (4, 0), user n is located at (1, 0), the path loss exponent is 3, and the power allocation coefficient for user m and user n is (0.8, 0.2). The targeted data rate is 1 bits per channel use (BPCU).***

B. NOMA in Cooperative-Transmission-Based Networks

Cooperative communication and NOMA techniques mutually support each others, hence the joint action of both techniques further improves the cooperative network performance [145]–[151]. Relaying and CoMP transmission are capable of improving cooperative networks.

1) Relay-Aided NOMA Transmission: Relaying has recently attracted considerable attention as a benefit of its network coverage extension [143]. In this context, Men and Ge investigated the outage performance of the single-antenna AF relay-aided NOMA downlink [145]. By contrast, in [146], multiple antennas were used. The potential gains of NOMA over OMA were quantified in both scenarios. As a further development, a DF-based two-stage relay selection protocol was proposed in [152], which relied on maximizing the diversity gain and minimizing the outage probability. In [153], Duan *et al.* proposed a novel two-stage PA scheme for a dual-hop relay-aided NOMA system, where the destination jointly decoded the information received both from the source and from the relay by applying the classic MRC technique. In [149], Kim and Lee considered a coordinated direct and relay-aided transmission scheme, where the BS simultaneously transmitted both to a nearby user and to a relay by invoking NOMA techniques during the first phase, while reaching a distant user with the aid of the relay.

2) Multicell NOMA With Coordinated Multipoint Transmission: When considering multicell scenarios, the performance of the cell-edge users is of particular concern. This is particularly important for downlink NOMA, since the SIC operations are usually carried out for cell-center users rather than for cell-edge users. Hence, the cell-edge users may not be well served. CoMP transmissions constitute an effective technique of improving the performance of cell-edge users. The key concept of CoMP in multicell NOMA is to enable multiple BSs to carry out coordinated beamforming or joint signal processing for the cell-edge users [44]. There are several research contributions in the context of handling the intercell interference in NOMA networks [42], [115], [147], [148], as detailed below.

In [115], Han *et al.* extended a single-cell NOMA to multicell network NOMA. A new precoding approach was proposed for mitigating the intercell interference, which led to an enhanced spectral efficiency and energy efficiency. In [147], Choi incorporated NOMA into CoMP for

Table 5 Contributions on the Interplay of NOMA and Cooperative Communications

Ref.	Model Topology	Direction	Techniques	Main Metrics	Characteristics
[80]	Idealized topology	DL	DF	OP+ergodic rate	Maximum diversity gain achievable
[110]	Stochastic geometry	DL	DF	OP	Wireless powered spatial randomly relays
[146]	Idealized topology	DL	AF	OP+ergodic rate	Multiple-antenna aided relays
[152]	Idealized topology	DL	DF	OP	Novel two-stage relay selection policy
[115]	Idealized topology	DL	CoMP	SE + EE	Zero-forcing like distributed precoding
[147]	Idealized topology	DL	CoMP	Sum-rate	Using Alamouti code at coordinated BSs
[42]	Idealized topology	DL	CoMP	Sum-rate	MIMO-NOMA coordinated beamforming
[148]	Idealized topology	DL	CoMP	OP+sum-rate	Joint transmission/Opportunistic NOMA in CoMP
[149]	Idealized topology	DL	CoMP	OP+ergodic rate	Each stream achieves a diversity order of one
[151]	Idealized topology	DL	CRAN	Throughput	All the BSs are controlled by a cloud
[153]	Idealized topology	DL	DF	Sum-rate	Novel two-stage power allocation
[155]	Idealized topology	DL	DF	OP	Invoking full-duplex technique at relays
[154]	Stochastic geometry	DL	CRAN	OP	Cluster point process based model

the sake of attaining a bandwidth efficiency improvement. A new coordinated superposition coding scheme relying on Alamouti's space-time code was proposed. As a further advance, Shin *et al.* [42] investigated the performance of multicell MIMO-NOMA networks, applying coordinated beamforming for dealing with the intercell interference in order to enhance the cell-edge users' throughput. Tian *et al.* [148] conceived an opportunistic NOMA scheme for CoMP systems and compared it to the conventional joint-transmission-based NOMA. In [151], Vien *et al.* proposed a NOMA-based PA policy for downlink cloud radio access network (C-RAN) scenarios, which can also be regarded as a coordinated transmission scenario, where the BSs jointly form a cloud. In [154], Martin *et al.* proposed a novel NOMA-aided C-RAN network, associated with using stochastic geometry. It revealed that the proposed framework is capable of substantially enhancing the performance of cell-edge users.

C. Discussions and Outlook

Table 5 summarizes all the aforementioned research contributions relying on the interplay between NOMA and cooperative communications. In Table 5, "DL" and "OP" represent the downlink and outage probability, respectively. The terminology of "idealized topology" in the table indicates that the users may not be randomly positioned and that the path loss is assumed to be constant. The aforementioned literature has investigated the potential performance enhancements of NOMA with the aid of cooperative techniques. Nonetheless, there are several open research opportunities. For example, although employing relays is capable of improving the reception reliability, while extending the coverage area, it requires an extra time slot for relaying. Developing full-duplex relays for NOMA networks constitutes a promising research direction for eliminating the requirement of an extra slot [155]–[157]. However, invoking full-duplex relays will require the elimination of both self-interference and interuser interference.

Additionally, the existing research contributions on C-RAN NOMA are still in their infancy. It is worth pointing out that C-RAN techniques are capable of efficient interference management and large-scale data control. This is particularly important for large-scale C-RAN NOMA networks, where sophisticated interference management (e.g., distributed beamforming design), dynamic user association, and efficient PA can be jointly considered.

V. RESOURCE MANAGEMENT IN NOMA NETWORKS

One of the main challenges in NOMA networks is to strike an attractive compromise between the bandwidth efficiency and the energy efficiency of the networks by intelligently controlling the PA of the superimposed signals [47], dynamically scheduling the users for the subchannels or by forming spatially correlated clusters. Motivated by this, in this section, we provide a comprehensive review of the existing resource allocation contributions, mainly with a special emphasis on power control and user/resource allocation.

A. Power Control for NOMA

Power control/allocation has been a pivotal research direction throughout each generation of networks, since inappropriately allocating power among users will inevitably increase the overall energy consumption while inflicting extra interference and hence degrading the overall performance of the networks. Again, in NOMA, more power is required for users with poor channel conditions and less power for users with better channel conditions, which guarantees fairness among NOMA users.

Numerous valuable contributions have been published on the PA problem in NOMA, which may be divided into two main categories, optimal PA from the perspective of all users and cognitive PA from the perspective of the weak users, which will be detailed in Sections V-A1 and V-A2.

1) Optimal Power Allocation: The contributions on NOMA which investigate optimal PA can be also classified

Table 6 Strategies Considering the Fairness Issues of NOMA

Fairness Strategy	Characteristics	Bandwidth efficiency	Fairness	Complexity	Ref.
Ordered power allocation	Allocates more power for weak users	Moderate	Moderate	Low	[36], [37], [89], [110]
Max-min rate fairness	Absolutely fairness for users	Low	High	High	[79], [93], [138]
Proportional fairness	Maximizes users' geometric mean	Moderate	Moderate	Low	[158], [159], [164], [165]
α utility function	Imposes fairness	High	Moderate	Moderate	—
Weighted sum rate	Achieves fairness with the aid of MAC layer	High	Low	Low	[161]
Jain's fairness index comparison	Only a fairness performance metric	—	—	—	[103], [108], [163]

into two main categories: a) single-channel/carrier PA; and b) multiple-channel/carrier/cluster PA. In this section, we mainly focus our attention on single-channel/carrier PA, while multiple-channel/carrier/cluster power allocation will be introduced in Section V-B.

The aim of PA in single-channel/carrier scenarios is to optimize the individual/sum rate while giving cognizance to the fairness issues. Compared to OMA, optimal PA in NOMA imposes more constraints associated with channel-quality and power-based ordering in an effort to guarantee fairness by allocating reasonable data rates to weak users as well. The recent research contributions on the fairness issues of NOMA can be summarized as follows. Table 6 summarizes most of the strategies considering the fairness issues of NOMA.

- **Ordered power allocation:** A simple yet effective approach to guarantee the fairness of NOMA systems is to allocate more power to users with poor channel conditions. By doing so, weak users can still achieve adequate rates. It is worth pointing out that adding these transmit-power-based ordering constraints will impose additional complexity on the optimization problem formulated, especially for multiple-antenna-aided NOMA scenarios [40], [125].
- **Max-min rate-fairness:** The max-min PA problem maximizes the rate of the weakest of all NOMA users [79], [93], [138]. In [79], the PA strategy was investigated under the scenarios of knowing either the instantaneous CSI or the average CSI. Note that by adopting the max-min rate as our objective function we can guarantee a certain grade of rate fairness, but at the price of sacrificing the system's sum rate.
- **Proportional fairness:** Proportional fairness (PF) is capable of maximizing the geometric mean of user rates. In NOMA scenarios, a feasible proportional fairness policy is to schedule users based on the instantaneous user rates, while additionally guaranteeing a certain long-term-average target rate [158], [159].
- ***α* utility function:** The α utility function constitutes a generalization of proportional fairness and max-min fairness. The definition of the α utility function is as follows [160]:

$$f_\alpha(x) = \begin{cases} \log x, & \text{if } \alpha = 1 \\ x^{1-\alpha}/(1-\alpha), & \text{otherwise}. \end{cases} \tag{8}$$

Note that for $\alpha = 1$, (8) reduces to that of proportional fairness, while for $\alpha \to \infty$, it results in the max-min fairness optimization problem. Hence, the α utility function is capable of treating the fairness issues of NOMA in a generic way.

- **Weighted sum rate:** The key idea behind the weighted sum rate is to consider an additional positive weighing factor for each user's achievable rate, which reflects the priority of each user in the context of resource allocation [161]. By doing so, a certain grade of fairness can be achieved from the specific perspective of the media-access control (MAC) layer.
- **Jain's fairness index comparison:** Jain's fairness index [162] is widely used for striking a tradeoff between the sum rate and fairness of a communication system. Motivated by this, many NOMA contributions characterized their PA in terms of Jain's fairness index [103], [108], [163]. However, Jain's fairness index is only a metric of evaluating a system's fairness, which cannot prevent the weak users from having low rates. More intelligent algorithms are required for fair PA among the users.

2) Cognitive-Radio-Inspired Power Control: The objective of CR-inspired power control relying on NOMA is to guarantee the QoS of weak users by constraining the power allocated to the strong user. Inspired by the CR concept [166], NOMA can be regarded as a special case of CR networks [24], [167]. More specifically, still considering a downlink scenario supporting two users, Fig. 10 compares conventional CR and CR-inspired NOMA. The BS can be viewed as the combination of a primary transmitter (PT) and a secondary transmitter (ST), which transmits the superimposed signals. The strong user (user n) and the weak user

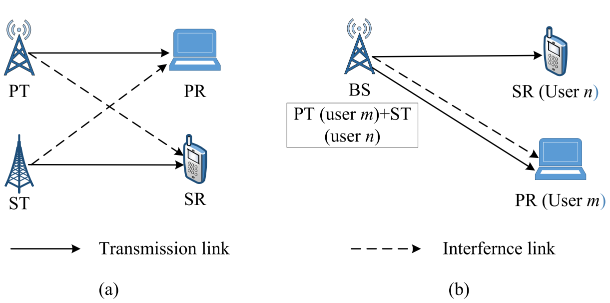


Fig. 10. *Comparison of convectional CR and CR-inspired NOMA. (a) Conventional CR. (b) CR inspired NOMA.*

(user m) can be regarded as a secondary receiver (SR) and a primary receiver (PR), respectively. By doing so, the strong user n becomes capable of accessing the spectrum occupied by the weak user m under predetermined interference constraints, which is the key feature of the classic underlay CR. The concept of CR-inspired PA in NOMA was proposed by Ding *et al.* [24], who investigated the PA of user-pairing-based NOMA systems.

The key advantages of cognitive PA are summarized as follows.

- **Guaranteed QoS:** By applying cognitive PA, the QoS requirements of the weak user are guaranteed, which is especially vital in real-time safety-critical applications.
- **Fairness/throughput tradeoff:** Cognitive PA is capable of striking a beneficial tradeoff between the overall system throughput and the individual user fairness, where the targeted data rate of the weak user has to be satisfied by appropriate PA.
- **High flexibility:** Cognitive PA offers a high degree of freedom for the BS to explore the opportunistic support of the strong user.
- **Low complexity:** Compared to the optimal PA approach, cognitive PA imposes a lower complexity during PA. This becomes particularly useful when the channel ordering and PA constraints are not convex and hence finding an appropriate PA scheme becomes a challenge, especially in multiple-antenna-aided NOMA scenarios.

Motivated by its advantages mentioned above, the cognitive PA policy was invoked for characterizing MIMO-NOMA systems in [107] and [127]. Particularly, in addition to investigating the convectional downlink cognitive PA conceived for MIMO-NOMA scenarios, Ding *et al.* [107] also designed a more sophisticated CR NOMA PA scheme for uplink MIMO-NOMA scenarios. In [127], in an effort to find a PA strategy suitable for SU-MIMO IoT scenarios, a cognitive PA policy was designed for ensuring that SIC may indeed be carried out at the strong user.

B. User Scheduling in Dynamic Cluster/Pair-Based Hybrid MA Networks

Due to the low-complexity design of NOMA and the new degree of freedom by the power dimension, it is widely accepted both by the academia and industry that NOMA constitutes a convenient "add-on" technique for establishing spectral efficient hybrid MA networks. In an effort to design hybrid MA networks, the users have to be assigned to different pairs/clusters. By doing so, how to pair/cluster the users becomes an interesting problem, which is valuable to examine. Several excellent contributions have researched the potential issues aiming for obtaining efficient user pairing/clustering strategies with the objective of improving the performance of both a single pair and of the entire networks, as detailed below.

1) A Single-User Pair's Performance: User pairing is capable of beneficially reducing the complexity of the NOMA design, when establishing advanced hybrid MA networks. By focusing on fully exploiting the potential gains of NOMA over conventional OMA from the perspective of a single-user pair, Ding *et al.* [24] investigated both the sum rate and the individual rates of the two users. It was demonstrated that a higher sum-rate gain can be attained by NOMA over OMA, when the channel conditions of the pair of NOMA users are rather different, which is in line with our expectation in the context of SIC-aided detection. In [110], the impact of spatial location on a pair of NOMA users was examined by employing a stochastic geometry model, in conjunction with three different user selection schemes. In [111], Qin *et al.* examined the secrecy performance of a specifically selected user pair in large-scale networks and showed that the secrecy diversity order of the user pair was determined by the user having the weaker channel. User pairing was invoked for MIMO-NOMA scenarios in [107] and [127]. We note that the user paring techniques of [24], [107], [110], [111], and [127] can be further improved by PA, which is hence a promising research direction.

2) User Clustering/Pairing From the Perspective of the Overall System's Performance: Note that allocating NOMA users to different orthogonal RBs (e.g., subchannels/subcarries/clusters) generally turns out to be a nondeterministic polynomial-time (NP)-hard problem. It is computationally prohibitive to obtain optimal results by performing an exhaustive search, when the number of users and RBs becomes high, especially when the number of user/RBs is dynamically fluctuating. Hence, investigating efficient low-complexity user allocation algorithms, which are capable of achieving attractive performance-complexity tradeoffs is necessary. Some of the research contributions in the field of resource allocation can be found in [159], [164], and [168]–[177]. We list some of the promising approaches as follows. The characteristics of those approaches are also summarized in Table 7.

Table 7 Summary of Resource Control Solutions for NOMA

UA approaches	Advantages	Disadvantages	Ref.
Combinatorial relaxation	Convert to convex problem	Existence of duality gap	—
Monotonic optimization	Optimal solution	Relatively high complexity	[161]
Matching theory	Achieve near-optimal performance	Requires predefined preference list	[169], [172], [173], [183]
Heuristic algorithms	Flexible complexity-performance tradeoff	Unstable performance	[164], [170], [171]

- **Combinatorial relaxation:** A popular method of tackling this kind of issue is to relax the constraint of the combinatorial problem by using continuous variable of $\eta \in [0,1]$ instead of a binary variable of $\eta \in \{0,1\}$, where η indicates whether a user is allocated to a RB block. By doing so, the original NP-hard problem is transferred to a convex problem and can be solved by invoking the classic Lagrangian dual method [178]. However, this kind of relaxation will result in a nontrivial duality performance gap between the original problem and the simplified one.
- **Monotonic optimization:** Due to the nonconvex nature of many sophisticated resource allocation problems, finding the optimal solution using convex optimization theory is rather challenging. In addition to convexity, the monotonicity is another important property, which can be exploited for efficiently solving nonconvex optimization problems [179], [180]. Note that the resource allocation problem of NOMA scenarios is usually nonconvex due to its interuser interference in the same RB. In this context, Sun *et al.* [161] invoked the classic monotonic optimization approach for developing an optimal solution for their joint power and subcarrier allocation problem. A low complexity suboptimal approach was also proposed for striking a performance-versus-complexity tradeoff.
- **Matching theory:** Matching theory, as a powerful mathematical modeling tool conceived for solving the combinatorial user allocation problems, is capable of overcoming some of the shortcomings imposed by the widely used game theory, such as the distributed-limited implementation and unilateral equilibrium deviation [181]. As such, matching theory has recently emerged as a promising technique of tackling the resource allocation problem in wireless networks, also in NOMA scenarios [169], [182]–[184]. Particularly, Di *et al.* [169] invoked many-to-many two-sided matching theory for resource allocation in the downlink of NOMA systems. A potential shortcoming of matching theory is that both the users and resources should have a predefined preference list. However, these preference lists may vary as the channel conditions fluctuate, which hence requires further research.
- **Heuristic algorithms:** Heuristic algorithms are commonly used for solving computationally complex problems. The core idea of heuristic algorithms is to obtain approximate solutions for the original optimization problem at an acceptable computational complexity. Inspired by this, many heuristic algorithms such as metaheuristics [171], greedy techniques [170], and others [164] have been adopted by researches in NOMA scenarios for allocating resources. Given the diverse nature of heuristic algorithms [185], they may lead to significantly different performance-versus-complexity tradeoffs.

It is worth pointing out that apart from applying intelligent user scheduling strategies, optimizing PA for each RB/user is capable of further improving the attainable network performance. However, the joint optimization of user scheduling and PA is a nontrivial problem. Except for [161], which jointly considers the PA as well as user association and obtains the optimal solution, the commonly adopted approaches rely on decoupling this correlated problem into a pair of subproblems [169]–[171]. As such, the aforementioned PA and user scheduling algorithms can be invoked for finding a suboptimal solution. However, the existing literature of resource allocation is predominantly focused on maximizing the performance attained, while there is a distinct lack of a systematic performance versus computational complexity analysis. Motivated by filling this gap, Lei *et al.* [163] characterized the tractability of NOMA resource allocation problems under several practical constraints. A combined Lagrangian duality and dynamic programming algorithm was also proposed in [163] for solving the joint optimization of the PA and user scheduling problem of NOMA networks.

C. Software-Defined NOMA Network Architecture

Inspired by the concepts of the emerging SDN paradigm [186], we propose the novel SD-NOMA concept, which provides the desired degree of flexibility for controlling resources. By decomposing the resource allocation and control problem into complex tractable problems, the SD-NOMA concept makes it much easier to create new abstractions in terms of power optimization, interference management, user association, and dynamic user clustering/pairing. Before introducing the proposed SD-NOMA strategy, we first review the compelling concept of SDR. The essential idea behind SDR is to implement the baseband signal processing algorithms in software instead of hardware, which hence exhibit a high grade of flexibility and reconfigurability, when designing agile instantaneously adaptive communication systems. In this spirit, a practical open source SDR-based NOMA prototype has been designed for a typical two-user downlink scenario by Xiong *et al.* [38], which bridges the theoretical and practical aspects of NOMA. Both the hardware and software architectures of the NOMA prototype systems designed were elaborated on. The link-level simulation results showed that the performance of the SIC procedure was significantly influenced both by the PA as well as by the modulation scheme employed.

In contrast to the SDR-based NOMA, the key idea of our proposed SD-NOMA network architecture is that the SDN controller has a global view of both the available resources and of the teletraffic across the whole network,

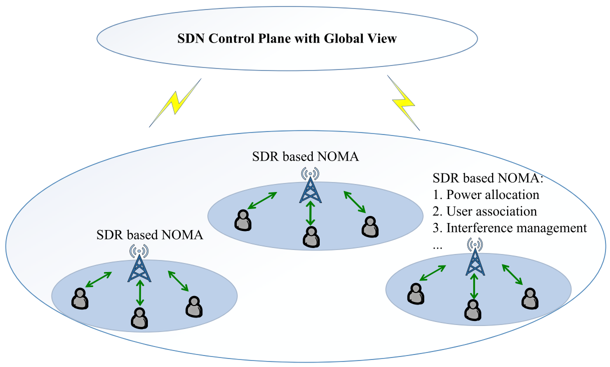


Fig. 11. ***Illustration of software-defined NOMA network architecture.***

as shown in Fig. 11. However, each component may still rely on the SDR-based NOMA concept. In other words, the SD-NOMA network architecture can globally control the SDR-based NOMA components via a central SDN controller. Considering the user association and PA in the context of a cluster-based MIMO-NOMA scheme [41] as an example, the SDN controller becomes capable of associating the different users with a potential cluster and allocating the most appropriate power levels to different users, while at the same time considering the interference arriving not only from the users in the same cluster but also from other clusters. This approach formulates the PA of NOMA as a global optimization problem, since the SDN controller is aware of the entire network's state. Therefore, SD-NOMA constitutes an attractive 5G network structure.

D. Discussions and Outlook

The distinctive cochannel interference characteristics of NOMA usually lead to a nonconvex optimization problem. The aforementioned research contributions have provided numerous beneficial techniques of dealing with the resource control problem. However, for complexity reasons, most of these techniques opted for decoupling the user scheduling and PA into consecutive subproblems, which usually result in suboptimal solutions. The problem of finding optimal solutions by jointly considering both the user scheduling and PA is still far from being well understood. Hence, advanced optimization techniques requiring low-complexity algorithms are required for solving this problem.

In summary, the design of attractive NOMA solutions hinges on finding the Pareto-optimal solutions, as discussed, for example, in an *ad hoc* networking context in [187], where none of the system characteristics can be improved without degrading at least one of the others. In the interactive seven-mode networking demo, a three-component objective function constituted by the BER, the power, and the delay was animated.

VI. COMPATIBILITY OF NOMA WITH OTHER TECHNOLOGIES TOWARD 5G AND BEYOND

As one promising technique in future 5G networks, one of the main challenges of NOMA is how well it is compatible with other emerging techniques for meeting the requirements of 5G. In this section, we survey the existing research contributions considering the coexistence of NOMA with other 5G proposals.

A. NOMA in HetNets

Dense HetNets, as one of the hot 5G technologies [9], are capable of significantly improving the network capacity. The core idea of HetNets is essentially to move the low-power BSs closer to the served users in order to form small cells under the oversailing macrocells. However, due to the cochannel nature of macrocells and small cells, the users surfer both from interlayer interference as well as from intralayer interference. To intelligently cope with the interference arriving from the other cochannel layers, in [150] Xu *et al.* proposed a cooperative NOMA scheme for HetNets and minimized the interuser interference with the aid of DPC precoding. The effect of distinctive power disparity between the macro BSs and pico BSs was investigated.

Additionally, as a benefit of aiming to take full advantage of both massive MIMO and HetNets techniques including high array gains, reliable BS-user links, etc., massive MIMO-aided HetNets solutions are promising for the emerging 5G systems [188]. The basic philosophy of massive MIMO-assisted HetNets is to install hundreds/thousands of antennas at the macro BSs for offering an unprecedented level of spatial degrees of freedom, while using a single antenna at the densely positioned small-cell BSs. In an effort to further enhance the bandwidth efficiency of small cells, a promising NOMA-aided and massive MIMO-based framework was proposed in [189], where a massive MIMO system was adopted by the macrocells to simultaneously serve N users and then user-pairing-based NOMA transmissions were adopted by the small cells, as shown in Fig. 12. More specifically, stochastic geometry was invoked for modeling the K-tier HetNets considered and to analyze the efficiency

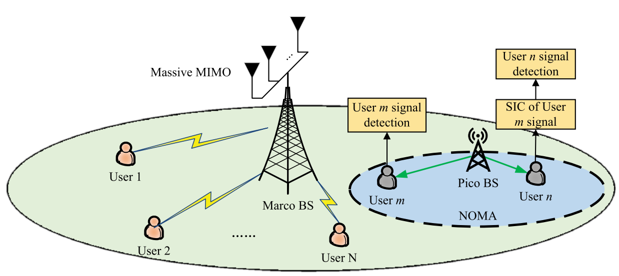


Fig. 12. ***Illustration of NOMA and massive MIMO-based hybrid HetNets.***

attained. The key feature of this framework is that it integrates the potential advantages of both NOMA (e.g., high bandwidth efficiency, faireness/throughput tradeoff, etc.) and of HetNets (low power consumption, spatial spectrum reuse, etc.) while simultaneously relying on a sophisticated-cluster-based MIMO-NOMA precoder design [40], [41], [107], [125], [127], [135].

B. NOMA in Millimeter-Wave Communications

MmWave communications have been recognized as a promising technique in 5G networks and beyond due to their large bandwidths in the high-frequency spectrum [190]. The severe propagation path loss of mmWave channels and their low penetration capabilities [191] necessitate the redesign of MA techniques, especially when aiming to support massive connectivity in dense networks, where hundreds/thousands of users have to be served within a small area. NOMA can be regarded as a potent MA technique capable of coexisting with mmWave networks due to the following reasons.

- The highly directional beams used in mmWave communications lead to correlated channels, which may degrade the performance of convectional OMA systems, but they are amenable to NOMA.
- The sharp beams of mmWave networks effectively suppress the interbeam interference among users, which is suitable for supporting NOMA in each beam.
- The application of NOMA in mmWave is capable of enhancing the bandwidth efficiency and supporting massive connectivity.

Inspired by this, some initial research contributions examining NOMA in mmWave networks have been disseminated in [113] and [192]. In [113], Ding *et al.* investigated the coexistence of NOMA and mmWave solutions relying on random beamforming. Due to the potential line-of-sight (LOS) blockages of mmWave systems, a thinning-process-aided stochastic geometry model was used to evaluate the performance. As a further advance, Cui *et al.* [192] investigated the performance of NOMA-mmWave networks relying on partial CSI feedback. More particularly, the user scheduling and PA were jointly considered with the aid of matching theory and branch-and-bound approaches. Fig. 13 illustrates the sum rate of NOMA-mmWave versus the SNR at different frequencies. It is demonstrated that the proposed NOMA-mmWave system is capable of outperforming conventional OMA-mmWave systems.

C. NOMA and Cognitive Radio Networks

The 2010s have witnessed the rapidly increasing penetration of mobile devices (e.g., smartphones, tablets, and laptops) all over the world, which gave rise to increasing demand for spectral resources. As reported by the U.S. Federal Communications Commission (FCC), there are

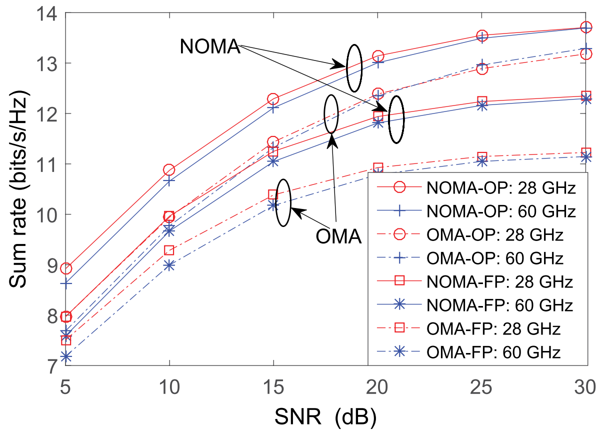


Fig. 13. ***Illustration of the superiority of NOMA-mmWave over OMA-mmWave networks both at 28 and 60 GHz. "OP" refers to optimized power allocation. "FP" refers to fixed power allocation. The detailed parameter settings are found in [192].***

significant temporal and spatial variations in the exploitation of the allocated spectrum. Given this fact, the CR concept—a terminology coined by Mitola [193]—inspired the community to mitigate the spectrum scarcity problem. The basic concept of CR is that at a certain time of the day or in a geographic region, the unlicensed secondary users (SUs) are allowed to opportunistically access the licensed spectrum of primary users (PUs). These CR techniques may be categorized into the interweave, overlay, and underlay paradigms [166].

- **Interweave:** The interweave CR can be regarded as an interference avoidance paradigm, where the SUs are required to sense the temporary slivers of the space-frequency domain of PUs before they access the channels [194]–[197]. The concurrent transmission of SUs and PUs is not allowed under the interweave paradigm.
- **Overlay:** The overlay paradigm essentially constitutes an interference mitigation technique. With the aid of the classic dirty paper encoding technique, overlay CR ensures that a cognitive user becomes capable of transmitting simultaneously with a noncognitive PU [166]. Additionally, SUs are capable of forwarding the information of PUs' to the PU receivers, while superimposing their own signals as a reward for their relaying services.
- **Underlay:** The underlay CR operates like an intelligent interference control paradigm, where the SUs are permitted to access the spectrum allocated to PUs as long as the interference power constraint at the PUs is satisfied [198].

Table 8 provides a summary of the interweave, overlay, and underlay paradigms and briefly illustrates the differences among them.

One of the core challenges in both CR and NOMA networks is the interference management, while improving the bandwidth efficiency. Hence, it is natural to link them for

Table 8 Summary and Comparison of the Three Main CR Paradigms

Paradigm	Concurrent Transmission	Constraints at SU	Interference Management	Complexity	NOMA-CR
Underlay	Not allowed	Limited by the interference power constraint	Interference avoiding	Low	[24], [112]
Overlay	Allowed	No power constraints	Interference mitigating	High	—
Interweave	Allowed	Limited by the range of sensing hole	Interference controlling	Medium	—

achieving an improved bandwidth efficiency. Liu *et al.* investigated the application of NOMA in large-scale underlay CR networks by relying on the stochastic geometry model [112]. The diversity order of the NOMA users was characterized analytically in two scenarios. The classic OMA-based underlay CR was also used as a benchmark to show the benefits of the proposed CR-NOMA scheme. As mentioned in Section VI, Ding *et al.* [24] proposed a novel PA policy for NOMA, namely the CR-inspired NOMA PA, which constitutes a beneficial amalgam of NOMA and underlay CR.

To the best of our knowledge, CR-NOMA studies only exist in the context of the underlay CR paradigm. Hence, both the interweave and overlay CR paradigms have to be investigated in NOMA networks. It is worth pointing out that a significant research challenge of NOMA is to dynamically cluster/pair the NOMA users first, followed by dynamically allocating the clusters/pairs to different orthogonal subchannels. In the context of the interweave paradigm, intelligent sensing has to be applied first, followed by user clusering/pairing of NOMA users, depending on the specific channel conditions sensed.

D. NOMA-Based Device-to-Device Communication

Due to the recent rapid increase in the demand for local area services under the umbrella of cellular networks, an emerging technique, namely device-to-device (D2D) communication, may be invoked for supporting direct communications amongst devices without the assistance of cellular BSs [199]. The main advantages of integrating D2D communications into cellular networks are: 1) low-power support of proximity services for improving the energy efficiency; 2) reusing the frequency of the oversailing cellular networks in an effort to increase the bandwidth efficiency; and 3) the potential to facilitate new types of peer-to-peer (P2P) services [200].

Note that one of the common features of both D2D and NOMA is that of enhancing the bandwidth efficiency by managing the interference among users within each RB. Motivated by this, it is desirable to invoke intelligent joint interference management approaches for fully exploiting the potential benefits of both D2D and NOMA. In [183], Zhao *et al.* designed a novel NOMA-based D2D communication scheme, where several D2D groups were permitted to share the same RB with the cellular users. In contrast to the conventional D2D pair's transmission, the novel "D2D group" concept was introduced, where a D2D transmitter was able to simultaneously communicate with multiple D2D receivers with the aid of NOMA. It was demonstrated that the proposed NOMA-based D2D scheme is capable of delivering higher throughput than conventional D2D communication.

E. Discussions and Outlook

The aforementioned techniques are capable of improving the bandwidth efficiency by exploiting NOMA, but they will also pose some challenges. For example, adopting NOMA in HetNets, CR networks, and D2D communication scenarios will impose increased cochannel interference on the existing networks. Hence, intelligent interference management is desired.

Apart from the aforementioned solutions, there are other emerging techniques, such as C-RAN, full-duplex, and other solutions. As the extension of [161], Sun *et al.* [201] studied the multiple carrier NOMA resource allocation problem for the scenario, where a full-duplex BS supported several half-duplex users. While the current research contributions have laid a solid foundation, numerous open questions have to be resolved.

VII. IMPLEMENTATION CHALLENGES AND STANDARDIZATION OF NOMA

Although NOMA has been recognized as a promising candidate for 5G and beyond, there are still several implementation challenges to be tackled. In this section, we identify some of the implementation issues and point out some potential approaches to solve them. Moreover, the standardization progress of NOMA will also be discussed in this section to show how NOMA paves the way to 5G and beyond.

A. Error Propagation in SIC

SIC is the key technique of user detection in NOMA systems. Nevertheless, a main drawback of implementing SIC is the interuser error propagation issue, which propagates from one user to another, because a decision error results in subtracting the wrong remodulated signal from the composite multiuser signal, hence resulting in residual interference [202], [203]. More specifically, if the messages of previous users are not correctly decoded, the reconstruction of those signals will result in flawed decoding of the remaining users' messages, which leads the accumulation of the decoding errors. Most existing research contributions in the context NOMA are based on the assumption that the SIC receivers are capable of perfectly canceling the interference. Actually this assumption cannot be readily satisfied in practice due to the inaccurate PA and imperfect channel decoding. Several researchers have recognized the opacity of error propagation issues and investigated the effect of imperfect SIC on uplink NOMA [114] systems, on full-duplex NOMA networks [204], and on C-RAN NOMA networks [205].

Some recent research contributions have identified several possible techniques for solving the error propagation issues. The first one is to use multistage channel estimation to improve the SIC [206] at the cost of increasing the complexity. The second possible approach is to apply iterative SIC-aided receiver to overcome the error propagation effects, as proposed by Zhang and Hanzo [34]. Another strategy is to take the channel estimation error into considerations when designing the power control algorithms, which was proposed by Buehrer [207]. More particularly, more power is allocated to the users who carry out SIC later for compensating the residual interferences. Nevertheless, invoking this approach imposes a capacity payoff [33].

B. Channel Estimation Error and Complexity for NOMA

Channel estimation plays a more significant role in NOMA than in OMA systems, since the channel estimation errors will result in ambiguous user ordering as well as inaccurate power control, which will in turn affect the SIC decoding accuracy. Naturally, perfect CSI cannot be obtained in practice. High-performance near-optimal conventional channel estimation algorithms impose an unacceptable system overhead and computational complexity on NOMA systems, especially for MIMO-NOMA scenarios. In order to solve those problems, researchers dedicated their efforts mainly to the following three aspects.

The first approach is to propose more effective channel estimation designs for striking a good complexity/performance tradeoff. As mentioned in Section III, CS can be applied as an effective technique of reducing the complexity [142]. The second approach is to rely on partial CSI. This is motivated by the fact that the large-scale fading fluctuates less rapidly than the small-scale fading. As mentioned in Section II, the use of partial CSI in downlink NOMA was studied in [91]–[93]. The third approach is to use limited feedback for reducing the overhead [94]. Finally, iterative joint decision-directed channel estimation and data detection can be invoked as detailed in [75].

C. Security Provisioning for NOMA

Security issues are of great significance in each generation of networks. The broadcast nature of the wireless medium makes it vulnerable to eavesdropping. Physical layer security (PLS), which was proposed by Wyner in as early as 1975 [208], has become an appealing technique of improving the confidentiality of wireless communications. In contrast to the traditional approaches, which design cryptographic protocols in the upper layers, PLS aims to exploit the specific characteristics of wireless channels in the physical layer for transmitting confidential messages. The key idea of achieving perfect secrecy in wiretap channels is to ensure that the capacity of the desired channel is higher than that of the eavesdropper's channel. Triggered by the rapid development of wireless networks, PLS has been considered in diverse scenarios [209]–[213].

Motivated by the security concerns of wireless communications, PLS measures have also been proposed for NOMA networks in order to combat eavesdropping [111], [139], [139], [214]–[216]. In [111], Qin *et al.* examined the PLS of single-antenna NOMA in large-scale networks by invoking stochastic geometry, where the BS communicates with randomly distributed NOMA users. A protected zone may be adopted around the BS, where the intended users benefit from a high capacity in order to establish an eavesdropper-exclusion area to enhance the PLS with the aid of careful channel ordering of the NOMA users. As a further development, Liu *et al.* [139] investigated the PLS of multiple-antenna NOMA scenarios, where artificial noise was generated at the BS for degrading the channels of eavesdroppers. Zhang *et al.* [214] investigated the PLS of SISO-NOMA networks, where the secrecy sum rate was maximized and the optimal PA was characterized in closed-form expressions.

D. Maintaining the Sustainability of NOMA With RF Wireless Power Transfer

One of the key objectives of future 5G networks is to maximize the energy efficiency and to support energy-constrained wireless devices. Energy consumption is a critical factor in maintaining the sustainability of wireless networks, especially for the devices, which invoke a high cost of replacing the batteries. EH is an effective technique of extending the battery recharge period. Hence, it has recently received remarkable attention [217]–[222]. However, the traditional EH techniques relying on solar-wind-vibration and thermoelectric effects, which depend on the location, environment, time of the day, etc. In contrast to the conventional EH techniques, radio-frequency (RF) wireless power transfer (WPT) provides a more flexible approach for powering energy-constrained devices. Another motivation behind this approach lies in the fact that most devices are surrounded by ubiquitous RF signals. As a consequence, even interfering signals can be regarded as the potential EH sources.

To elaborate, NOMA relying on RF WPT techniques in wireless networks has been studied in [108], [110], [223], and [224]. Particularly, in [223], the application of WPT to NOMA networks was investigated, where the users are randomly located. As a further development, in [110], a new cooperative simultaneous wireless information and power transfer (SWIPT)-based NOMA protocol was proposed. In order to elaborate on the impact of user association, three user selection schemes based on the user distances from the base station were proposed in [110]. The analytical results of [110] confirmed that invoking SWIPT techniques did not degrade the diversity gain compared to that of conventional NOMA. In [108], the uplink of NOMA transmission was considered in the context of energy constrained users, who can harvest energy from the BS by adopting the "harvest-then-transmit" protocol. The results demonstrated that NOMA is also capable of providing considerable throughput, fairness, and energy-efficiency improvements.

E. State of the Art for Standardization of NOMA

NOMA has recently been included into LTE-A, terms MUST [225]. More specifically, at the 3GPP meeting in May 2015, it was decided to include MUST into LTE Advanced. Afterwards, at the 3GPP meeting in August 2015, 15 different forms of MUST have been proposed by Huawei, Qualcomm, NTT DOCOMO, Nokia, Intel, LG Electronics, Samsung, ZTE, Alcatel Lucent, etc. Finally, at the 3GPP meeting in December 2015, NOMA was included into LTE Release 13 [226]. It is worth noting that the MUST technique may be made compatible with the existing LTE structure. In other words, NOMA allows two users to be served at the same OFDM subcarrier without changing the current structure. Various nonorthogonal transmission schemes have been proposed for the MUST items [225], [226], which can be generally classified into three categories [227], namely: 1) superposition transmission with an adaptive power ratio on each component constellation and non-Gray-mapped composite constellation; 2) superposition transmission with an adaptive power ratio on component constellations and Gray-mapped composite constellation; and 3) superposition transmission with a label-bit assignment on composite constellation and Gray-mapped composite constellation. The examples of transmitter processing candidate can be found in [23, Fig. 6]. The NOMA-like MUST architecture works as follows. The coded bits of near users and far users are first separately input in a bit converter and then modulated with the aid of quadrature amplitude modulation (QAM) or quadrature phase shift keying mapper (QPSK). The modulated symbols are superposed with allocating appropriate powers to transmit. Actually, the first category can be regarded as a special case of the second category, since coded bits in the first category are modulated directly to mappers without inputting into the bit converter. Both of these two categories are capable of supporting flexible power partition among users [228]. For the third category, it is a bit-partition-based scheme which is in contrast to the power partition scheme adopted in the first and second categories. Regarding the detailed differences among the mentioned three schemes, interested readers may refer to [23] and [228] to identify the transmit structures of the three schemes. Regarding receivers, interference cancelation is required to be carried out. The interference scenarios are different between network-assisted interference cancellation system (NAICS) and MUST since the receivers of NAICS and MUST are employed by different groups of users. More particularly, for a two-user case which consists of a cell-edge user and a cell-center user, NAICS receivers are typically used for cell-edge users while MUST is typically used for cell-center users. By doing so, the network performance can be enhanced.

In addition to LTE-A, NOMA had also been included in the forthcoming digital TV standard, by the Advanced Television Systems Committee (ATSC) 3.0 [229], termed as layered division multiplexing (LDM) for providing significant improvements in terms of service reliability, system flexibility, and spectrum efficiency. This standard will generate significant impact on digital TV industry. Moreover, in the white paper of NTT DOCOMO, NOMA has been identified as a key technique for 5G. The system-level performance of NOMA has also been demonstrated by NTT DOCOMO [90]. It is worth pointing out that the key challenges of implementing NOMA in industry is the fact that the decoding complexity increases at receivers as the number of users increases. Another potential challenge is that the security and privacy of far users should be protected at near user side due to the characteristic of SIC, which may depend on key generations from the upper layer. Although NOMA is reviewed as a promising candidate for 5G and beyond, there are various forms of NOMA (which will be detailed in Section VIII). The standardization process for NOMA is still ongoing.

F. Discussions and Outlook

In addition to the aforementioned implementation issues, there are also other imperfections to be dealt with for NOMA. Since NOMA relies on MA at different power levels, the strength of the received signals is deliberately different, which bring about new challenges for accurate analog-to-digital (A/D) conversion. On the one hand, for strong signals, a large voltage range is needed. On the other hand, for weak signals, high-resolution ADCs are required for supporting accurate quantization at small levels. In practice, considering the cost and system complexity, it is impossible to apply ADCs, which have both large voltage range as well as high resolution. The quantization errors are unavoidable. It is important to seek a good performance/complexity tradeoff.

Accurate synchronization is another significant issue in NOMA, which is to a degree neglected in the existing literature. Actually, perfect synchronous transmissions cannot be achieved in practice due to the dynamic mobile environment of users, especially for uplink NOMA transmission. In order to solve the synchronization issues, two possible approaches can be adopted. The first one is to propose accurate pilot design for reducing the time synchronization errors. The second approach is to investigate novel asynchronous communication schemes. Haci *et al.* [230] proposed a new IC scheme for asynchronous NOMA-aided OFDM systems.

It was demonstrated that the system performance largely depends on the relative time offset among the interfering users. Other imperfections such as the impact of filtering distortion between transmitters and receivers on the performance degradation of NOMA systems are still unknown, which is a promising future direction to consider.

VIII. PRACTICAL FORMS OF NOMA

In the previous sections, power-domain NOMA has been investigated from diverse perspectives, including multiple-antenna techniques, CR NOMA, the resource management problems of NOMA and the coexistence of NOMA as well as other 5G solutions. In practice, NOMA can be implemented in various forms, such as code-domain and power-domain NOMA, as summarized in this section. We may classify the practical forms of NOMA also into single-carrier and multiple-carrier NOMA. Their advantages and disadvantages will be discussed in this section.

A. Single-Carrier NOMA

We commence our discussions from the single-carrier NOMA design, which will lay the foundations for the multiple-carrier NOMA design in Section VIII-B.

1) IDMA: The key idea of the IDMA scheme relies on a unique user-specific chip interleaver for distinguishing the signals of different users [231]. Hence, IDMA may be viewed as chip-interleaved CDMA, which has the benefit of a high diversity gain, because if one or two chips are corrupted, the corresponding spreading sequence could still be recovered with the aid of low-complexity chip-by-chip iterative multiuser detection (IMD) strategy. A comprehensive comparison of IDMA and CDMA was presented in [232] in terms of its performance versus complexity.

2) LDS-CDMA: LDS-based CDMA constitutes an enhanced version of CDMA [233], which is inspired by the investigations of the low-density parity-check (LDPC) codes [234]. For conventional CDMA, one possible solution is to assign orthogonal spreading sequences for each user, and hence low-complexity receiver can be invoked at receivers for eliminating interference. Nonetheless, this orthogonal spreading code design is only capable of supporting an equal number of users and chips, which motivates the development of the nonorthogonal spreading code design. So-called sparse spreading sequences are employed instead of the dense spreading sequences of conventional CDMA, such as orthogonal variable spreading factor (OVSF) codes. However, such design requires sophisticated multiuser decoding at receivers. For LDS-CDMA, a message passing algorithm (MPA) based on the multiuser detection technique can be applied at the receiver, which is capable of achieving the near-maximum-likelihood (ML) detection performance.

3) LPMA: Lattice-partition-based multiple access (LPMA) is a new downlink nonorthogonal multiuser superposition transmission scheme, which was proposed by Fang *et al.* [235]. LPMA achieves a beneficial multiplexing gain both in the power domain and in the code domain. More explicitly, power-domain multiplexing has the potential of increasing the throughput by superimposing different-power streams. By contrast, code-domain multiplexing superimposes several streams by exploiting that the linear combination of lattice codes also results in a lattice code. More specifically, LPMA encodes the information of users by applying lattice coding at the transmitters and invokes SIC at the receivers for detection. It is worth pointing out that a specific advantage of LPMA is that it has the potential of circumventing a specific impediment of power-domain NOMA, namely that the performance gain attained relies on the channel quality difference of users. However, LPMA imposes a higher encoding and decoding complexity than power-domain NOMA.

B. Multiple-Carrier NOMA

Given the remarkable advantages of OFDM, this mature wave form design is likely to be incorporated in 5G networks. Explicitly, one of the main benefits of multiple-carrier techniques is that instead of the short dispersion-sensitive symbols of serial modems, many parallel long-duration dispersion-resistant symbols are transmitted over dispersive channels. Hence, these long-duration symbols are only mildly affected by the same CIR and can be readily equalized by a single-tap frequency-domain equalizer. The NOMA techniques are expected to coexist with OFDM. In this sections, several forms of multiple-carrier NOMA designs are discussed.

1) LDS-OFDM: LDS-OFDM [236] is essentially a combination of LDS-CDMA and OFDM, which can be regarded as an advanced variant of LDS-CDMA in a multiple-carrier form. Therefore, the same sparse spreading sequences are applied at the transmitters and the same MPA-based detection is adopted at the receivers. The data-stream mapping process consists of two steps: a) each bit of the data streams is first spread by the low density spreading sequences; and b) the data streams are then transmitted over different subcarriers by applying an OFDM modulator. There are several techniques of mapping the spreading sequences to OFDM, each having different pros and cons. For example, each chip may be mapped to a different subcarrier for achieving frequency-domain diversity. Alternatively, each subcarrier may convey a spread low-rate stream. However, this design results in a higher complexity due to the application of the MPA detection compared to a conventional OFDMA design.

2) SCMA: SCMA [237] also relies on sparse spreading codes for ensuring that each RB can support more than one user with the aid of low density spreading, hence

Table 9 Comparison of the Existing NOMA Solutions

Paradigm	Carrier	Spreading	Modulation	Flexibility	Complexity	Decoding	Key Advantages	Main Challenges
PD-NOMA	SC	Not used	Separated	High	Low	SIC	High Spectral Efficiency	Error Propagation
IDMA	SC	Low-rate	Separated	Medium	Low	IMD	High Diversity Gain	Interleaver Design
LDS-CDMA	SC	Sparse	Separated	High	Medium	MPA	CSI is not required	Coding Redundancy
LPMA	SC	Lattice	Separated	Medium	Medium	SIC	Gain benefits PD and CD	Specific Channel Coding
LDS-OFDM	MC	Sparse	Integrated	Low	Medium	MPA	Wideband Signals Suitable	Coding Redundancy
SCMA	MC	Sparse	Integrated	Low	High	MPA	Enhanced Diversity than LDS	Codebook Design
PDMA	MC	Sparse	Separated	Medium	High	MPA-SIC	Multi-dimension Diversity	Patterns Design

resulting in a scheme similar to LDS-CDMA and LDS-OFDM. Considering an example of six users and four subcarriers, a typical signature matrix can be illustrated as follows:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{9}$$

It is worth noting that for $\mathbf{S}$ in (9), each column has only two nonzero entries, which implies that each user is only allowed to occupy two subcarriers. This is one of the key characteristics of SCMA. In addition to these sparse spreading sequences, SCMA also relies on multidimensional constellations for generating its codebooks in order to achieve a so-called constellation shaping gain. More particularly, SCMA allows the employment of fewer constellation points to be used at a given throughput. For example, the four-level quadrature amplitude modulation (4-QAM) constellation can be reduced to a smaller constellation, because some bits can be conveyed in the sparse-code domain. However, this requires sophisticated codebook design [238]–[240]. SCMA relies on the joint encoding of the signals of multiple subcarriers, hence joint decoding is required at the receivers. For this large joint alphabet MPA-based detection is capable of achieving near-ML performance at a fraction of the ML complexity.

3) PDMA: PDMA is another form of sparse-signature-matrix-based multiple-carrier NOMA [22]. Again, we use the previous example of six users and four subcarriers relying on the pattern matrix of

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{10}$$

At the transmitter side, in contrast to the matrix design of SCMA in (9), where the number of users in each RB has to be the same, PDMA allows a variable number of users to be mapped to each RB. For example, the first user in (10) is seen to activate all the subcarriers for transmission, while the sixth user only activates a single subcarrier. Another main difference between PDMA and SCMA is that PDMA does not aim for achieving a constellation shaping gain, which beneficially avoids the complex multidimensional constellation design of SCMA. At the receiver side, the MPA may also be adopted for joint decoding, which is similar to LDS-CDMA and SCMA. Moreover, other MPA-SIC detection strategies or the turbo detection design of [241] can also be used for improving the decoding performance.

C. Discussions and Outlook

While we have provided a detailed introduction to the most popular power-domain NOMA variants, there are also other NOMA schemes, such as multiuser superposition transmission (MUST) scheme of [225], the MUSA arrangement of [21], etc. Table 9 compares their key characteristics at a glance, where "SC" refers to single-carrier schemes and "MC" refers to multiple-carrier schemes. Note that each form has its own advantages and disadvantages. For example, SCMA is an open-loop grant-free access scheme, which is suitable for uplink transmission, and it relies both on sophisticated codebook design as well as on joint MPA detection. MUST is capable of enhancing the bandwidth efficiency but it can only be used for downlink transmission. As a consequence, a unified NOMA framework is sought, which is capable of supporting the access of numerous users in general scenarios. This is a promising future problem for researchers to contribute to. Additionally, more research contributions are also needed on the following aspects: 1) optimal codebook/coding designs for SCMA and PDMA; 2) hybrid MPA-SIC designs are sought for low-complexity decoding; and 3) the joint designs of new modulation and MA schemes.

IX. SUMMARY AND CONCLUSION

In this paper, the recent literature of power-domain multiplexing-aided NOMA proposed for 5G systems has been surveyed with an emphasis on the following aspects: the basic principles of NOMA, the amalgams of multiple antenna techniques and NOMA, the interplay of NOMA and cooperative communications, the resource control of NOMA, its coexistence with other key 5G techniques, and the implementation challenges and standardization. Apart from surveying the existing NOMA contributions, we have highlighted the key advantages of NOMA itself as well as the inherent features of other techniques in Section II. Investigating the inherent integration of multiple-antenna-aided and cooperative techniques with NOMA is particularly important, since they are capable of providing extra spatial diversity, either with the aid of centralized or distributed beamforming designs. It is worth pointing out that when designing centralized beamformers, the most influential factor is the appropriate ordering of the matrix-based channels. Design guidelines were provided both for multiple-antenna-aided and cooperative NOMA in Sections III and IV, respectively.

Bearing in mind that NOMA multiplexes users associated with different power levels with the aid of superposition coding techniques, the power sharing among the users should be carefully optimized for each scenario. For example, in practical scenarios, a laptop displaying high-definition online video may share an RB with low-rate wireless sensors. Clearly, these applications have very different target rates, and processing capabilities. More particularly, high-capability laptops and smartphones can readily carry out SIC, but this is quite the contrary for sensors. As such, intelligent resource control including PA and user scheduling should be investigated by taking into account the recommendations listed in Section V. What is also worth emphasizing is that the proposed software-defined NOMA strategy is capable of dealing with all the aforementioned issues from a global perspective for the sake of holistically optimizing the performance of the entire network.

The benefits of invoking NOMA in combination with the emerging large-scale MIMO, cooperative transmission, wireless power transfer, HetNets, etc., were discussed. However, given the fact that NOMA imposes extra intrauser interference on the system, it brings about a range of open challenges, especially those related to sophisticated interference management. This problem would become more challenging in CR, D2D, and HetNets scenarios. Indeed, the related research of NOMA relying on sophisticated interference coordination is still in its infancy and hence has to be more deeply investigated, as recommended in Section VI.

The implementation issues of NOMA such as error propagation, channel estimation errors, security issues, etc., were discussed in Section VIII. Moreover, the standardization progress of NOMA was also discussed. It was noted that there are still several gaps between research and implementation to fill in, which require more effort in this field, as described in Section VIII.

Apart from power-domain NOMA, other NOMA schemes have also been advocated both by the industry and academia, as described in Section VIII. All NOMA schemes discussed in this treatise share the same spirit of nonorthogonal transmissions to enhance the attainable bandwidth efficiency and to provide connectivity for numerous users within the limited number of RBs. Naturally, the problem of designing a unified NOMA framework would be beneficial for supporting 5G scenarios. ■

REFERENCES

[1] L. Hanzo, L.-L. Yang, E.-L. Kuan, and K. Yen, *Single and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Networking and Standards*. Hoboken, NJ, USA: Wiley, 2003.

[2] L. Hanzo, J. S. Blogh, and S. Ni, *3G, HSPA and FDD Versus TDD Networking: Smart Antennas and Adaptive Modulation*. Hoboken, NJ, USA: Wiley, 2008.

[3] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. Hoboken, NJ, USA: Wiley, 2005.

[4] J. Yin, *et al.*, "Satellite-based entanglement distribution over 1200 kilometers," *Science*, vol. 356, no. 6343, pp. 1140–1144, 2017.

[5] "*Cisco visual networking index: Global mobile data traffic forecast update 2014-2019*," Cisco, San Jose, CA, USA, white paper, Dec. 2015.

[6] C.-X. Wang, *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[7] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[8] E. Hossain and M. Hasan, "5G cellular: Key enabling technologies and research challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, Jun. 2015.

[9] J. G. Andrews, *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[10] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the MIMO era: Evolution of standards, waveform design, and future perspectives," *Proc. IEEE*, vol. 99, no. 8, pp. 1343–1385, Aug. 2011.

[11] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys Tuts.*, to be published.

[12] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems*, 2nd. Hoboken, NJ, USA: Wiley, 1999.

[13] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley, III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, May 1991.

[14] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[15] P. Wang, J. Xiao, and L. P, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 4–11, Sep. 2006.

[16] L. Dai, B. Wang, Y. Yuan, S. Han, C. L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[17] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Turbo multi-user detection for OFDM/SDMA systems relying on differential evolution aided iterative channel estimation," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1621–1633, Jun. 2012.

[18] F. Brannstrom, T. M. Aulin, and L. K. Rasmussen, "Iterative detectors for trellis-code multiple-access," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1478–1485, Sep. 2002.

[19] L. Liu, J. Tong, and L. Ping, "Analysis and optimization of CDMA systems with chip-level interleavers," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 1, pp. 141–150, Jan. 2006.

[20] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[21] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for Internet of Things," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2016, pp. 1–5.

[22] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—A novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, Apr. 2016.

[23] Z. Ding, *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[24] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[25] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[26] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 197–207, Mar. 1973.

[27] S. Verdú, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Commun.*, vol. IT-34, no. 9, pp. 890–897, Sep. 1986.

[28] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Sel.*

Areas Commun., vol. 12, no. 5, pp. 796–807, Jun. 1994.

[29] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels.I. Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[30] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels.II. Outage capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1103–1127, Mar. 2001.

[31] A. Mostafa, R. Kobylinski, I. Kostanic, and M. Austin, "Single antenna interference cancellation (SAIC) for GSM networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 2. Oct. 2003, pp. 1089–1093.

[32] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[33] J. G. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 19–29, Apr. 2005.

[34] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 503–520, 3rd Quart., 2011.

[35] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.

[36] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. 77th IEEE VTC-Spring*, Dresden, Germany, Jun. 2013, pp. 1–5.

[37] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[38] X. Xiong, W. Xiang, K. Zheng, H. Shen, and X. Wei, "An open source SDR-based NOMA system for 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 24–32, Dec. 2015.

[39] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (NOMA): Concept, performance evaluation and experimental trials," in *Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, Oct. 2015, pp. 1–6.

[40] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.

[41] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[42] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.

[43] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor (Nov. 2016). "*Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges*." [Online]. Available: https://arxiv.org/abs/1611.01607

[44] M. S. Ali, E. Hossain, and D. I. Kim (Mar. 2017). "*Coordinated multi-point (CoMP) transmission in downlink multi-cell NOMA systems: Models and spectral efficiency performance*." [Online]. Available: https://arxiv.org/abs/1703.09255

[45] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[46] H. Tabassum, Md. S. Ali, E. Hossain, Md. J. Hossain, and D. I. Kim (2016). "*Non-orthogonal multiple access (NOMA) in cellular uplink and downlink: Challenges and enabling techniques*." [Online]. Available: https://arxiv.org/abs/1608.05783

[47] L. Song, Y. Li, Z. Ding, and H. V. Poor (2016). "*Resource management in non-orthogonal multiple access networks for 5G and beyond*." [Online]. Available: https://arxiv.org/abs/1610.09465

[48] W. van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. COM-24, no. 2, pp. 276–283, Feb. 1976.

[49] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Pers. Commun.*, vol. 2, no. 2, pp. 46–58, Apr. 1995.

[50] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 124–136, Oct. 1996.

[51] W.-H. Sheen and G. L. Stuber, "MLSE equalization and decoding for multipath-fading channels," *IEEE Trans. Commun.*, vol. 39, no. 10, pp. 1455–1464, Oct. 1991.

[52] A. J. Viterbi, "Very low rate convolution codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 4, pp. 641–649, May 1990.

[53] H. Schoeneich and P. A. Hoeher, "Single antenna interference cancellation: Iterative semiblind algorithm and performance bound for joint maximum-likelihood interference cancellation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 3. Dec. 2003, pp. 1716–1720.

[54] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 312–335, 1st Quart., 2013.

[55] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 439–441, May 1983.

[56] M. Sawahashi, K. Higuchi, H. Andoh, and F. Adachi, "Experiments on pilot symbol-assisted coherent multistage interference canceller for DS-CDMA mobile radio," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 2, pp. 433–449, Feb. 2002.

[57] *Range Expansion for Efficient Support of Heterogeneous Networks*, document TSG-RAN WG1 Meeting 54bis R1-083813, 3rd Generation Partnership Project (3GPP), Qualcomm Europe, Prague, Czech Republic, Sep. 2008.

[58] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless. Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[59] B. Li, J. Ji, and X. Wang, "Interference cancellation for HetNet deployment in 3GPP LTE-advanced rel-11," in *Proc. IEEE Proc. Veh. Technol. Conf. (VTC)*, Jun. 2013, pp. 1–5.

[60] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[61] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1558–1562.

[62] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[63] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.

[64] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[65] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, Mar. 2001.

[66] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[67] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals Syst. Electron.*, Oct. 1998, pp. 295–300.

[68] E. Gelal, *et al.*, "Topology control for effective interference cancellation in multiuser MIMO networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 455–468, Apr. 2013.

[69] C. Jiang, Y. Shi, Y. T. Hou, W. Lou, S. Kompella, and S. F. Midkiff, "Squeezing the most out of interference: An optimization framework for joint interference exploitation and avoidance," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 424–432.

[70] C. Xu, L. Ping, P. Wang, S. Chan, and X. Lin, "Decentralized power control for random access with successive interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2387–2396, Nov. 2013.

[71] J. Lee, J. G. Andrews, and D. Hong, "Spectrum-sharing transmission capacity with interference cancellation," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 76–86, Jan. 2013.

[72] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problems Peredachi Inf.*, vol. 10, no. 3, pp. 3–14, Jul./Sep. 1974.

[73] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 503–520, 3rd Quart., 2011.

[74] L. L. Hanzo and T. Keller, *OFDM and MC-CDMA: A Primer*. Hoboken, NJ, USA: Wiley, 2007.

[75] L. Hanzo, Y. Akhtman, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent Versus Non-Coherent and Cooperative Turbo Transceivers*. Hoboken, NJ, USA: Wiley, 2010.

[76] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Pers. Commun.*, vol. 2, no. 2, pp. 46–58, Apr. 1995.

[77] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 124–136, Oct. 1996.

[78] Z. Chen, Z. Ding, X. Dai, and R. Zhang (Dec. 2016). "*A mathematical proof of the superiority of NOMA compared to conventional OMA*." [Online]. Available: https://arxiv.org/abs/1612.01069

[79] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[80] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[81] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, 2017.

[82] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. PIMRC*, Sep. 2013, pp. 332–336.

[83] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.

[84] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks," *IEEE Access*, vol. 3, pp. 1633–1639, Aug. 2015.

[85] S. L. Shieh and Y. C. Huang, "A simple scheme for realizing the promised gains of downlink non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1624–1635, Apr. 2016.

[86] J. So and Y. Sung, "Improving non-orthogonal multiple access by forming relaying broadcast channels," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1816–1819, Sep. 2016.

[87] L. Liu, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving iterative LMMSE detection for MIMO-NOMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[88] H. Xing, Y. Liu, A. Nallanathan, and Z. Ding, "Sum-rate maximization guaranteeing user fairness for NOMA in fading channels," Singapore, Tech. Rep., Dec. 2017.

[89] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annu. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 611–615.

[90] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops*, Dec. 2013, pp. 66–70.

[91] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[92] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink non-orthogonal multiple access with statistical channel state information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4718–4731, Jul. 2016.

[93] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[94] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with one-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[95] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[96] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.

[97] H. Marshoud, V. M. Kapinas, G. K. Karagiannidis, and S. Muhaidat, "Non-orthogonal multiple access for visible light communications," *IEEE Photon. Technol. Lett.*, vol. 28, no. 1, pp. 51–54, Jan. 1, 2016.

[98] L. Yin, W. O. Popoola, X. Wu, and H. Haas, "Performance evaluation of non-orthogonal multiple access in visible light communication," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5162–5175, Dec. 2016.

[99] X. Zhang, Q. Gao, C. Gong, and Z. Xu (Oct. 2016). "*Interference management and power allocation for NOMA visible light communications network*." [Online]. Available: https://arxiv.org/abs/1610.07327

[100] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless," *Proc. IEEE*, vol. 100, pp. 1853–1888, May 2012.

[101] P. Botsinis, *et al.*, "Quantum-aided multi-user transmission in non-orthogonal multiple access systems," *IEEE Access*, vol. 4, pp. 7402–7424, 2016.

[102] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, Mar. 2015.

[103] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 781–785.

[104] M. Al-Imari, P. Xiao, and M. A. Imran, "Receiver and resource allocation optimization for uplink NOMA in 5G wireless networks," in *Proc. ISWCS*, Brussels, Belgium, Aug. 2016, pp. 151–155.

[105] S. Chen, K. Peng, and H. Jin, "A suboptimal scheme for uplink NOMA in 5G systems," in *Proc. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2015, pp. 1429–1434.

[106] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, Mar. 2016.

[107] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[108] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, "Wireless powered communications with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8422–8436, Dec. 2016.

[109] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 2nd ed. Hoboken, NJ, USA: Wiley, 1996.

[110] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[111] Z. Qin, Y. Liu, Z. Ding, Y. Gao, and M. Elkashlan, "Physical layer security for 5G non-orthogonal multiple access in large-scale networks," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[112] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, "Non-orthogonal multiple access in large-scale underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10152–10157, Dec. 2016.

[113] Z. Ding, P. Fan, and H. V. Poor (Jul. 2016). "*Random beamforming in millimeter-wave NOMA networks*." [Online]. Available: https://arxiv.org/abs/1607.06302

[114] H. Tabassum, E. Hossain, and Md.J. Hossain (Oct. 2016). "*Modeling and analysis of uplink non-orthogonal multiple access (NOMA) in large-scale cellular networks using poisson cluster processes*." [Online]. Available: https://arxiv.org/abs/1610.06995

[115] S. Han, I. Chih-Lin, Z. Xu, and Q. Sun, "Energy efficiency and spectrum efficiency co-design: From NOMA to network NOMA," E-LETTER, Tech. Rep., 2014.

[116] H. A. David and N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.

[117] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[118] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10231–10236, Dec. 2016.

[119] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Robust beamforming techniques for non-orthogonal multiple access systems with bounded channel uncertainties," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2033–2036, Sep. 2017.

[120] Y. Tan, J. Zhou, and J. Qin, "Novel channel estimation for non-orthogonal multiple access systems," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1781–1785, Dec. 2016.

[121] X. Liu and H. Jafarkhani (Jan. 2017). "*Downlink non-orthogonal multiple access with limited feedback*." [Online]. Available: https://arxiv.org/abs/1701.05247

[122] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with 1-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[123] B. Tomasi, F. Gabry, V. Bioglio, I. Land, and J.-C. Belfiore (Jan. 2017). "*Low-complexity receiver for multi-level polar coded modulation in non-orthogonal multiple*

access." [Online]. Available: https://arxiv.org/abs/1701.05814

[124] Q. Sun, S. Han, C. L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[125] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[126] J. Choi, "On the power allocation for MIMO-NOMA systems with layered transmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3226–3237, May 2016.

[127] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, Apr. 2016.

[128] B. Kim, *et al.*, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278–1283.

[129] S. Qureshi, H. Kim, and S. A. Hassan, "MIMO uplink NOMA with successive bandwidth division," in *Proc. IEEE Wireless Commun. Netw. Conf. (MILCOM)*, Doha, Qatar, Apr. 2016, pp. 1–6.

[130] L. Liu, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving iterative LMMSE detection for MIMO-NOMA systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[131] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.

[132] Z. Chen, Z. Ding, and X. Dai, "Beamforming for combating inter-cluster and intra-cluster interference in hybrid NOMA systems," *IEEE Access*, vol. 4, pp. 4452–4463, 2016.

[133] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic (Dec. 2016). "*Antenna selection for MIMO non-orthogonal multiple access systems*." [Online]. Available: https://arxiv.org/abs/1612.04943

[134] M. S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[135] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–5.

[136] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. IEEE Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2013, pp. 770–774.

[137] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.

[138] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

[139] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

[140] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[141] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[142] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[143] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[144] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[145] J. Men and J. Ge, "Performance analysis of non-orthogonal multiple access in downlink cooperative network," *IET Commun.*, vol. 9, no. 18, pp. 2267–2273, Dec. 2015.

[146] J. Men and J. Ge, "Non-orthogonal multiple access for multiple-antenna relaying networks," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1686–1689, Oct. 2015.

[147] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[148] Y. Tian, A. Nix, and M. Beach, "On the performance of opportunistic NOMA in downlink CoMP networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 998–1001, May 2016.

[149] J. B. Kim and I. H. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2037–2040, Nov. 2015.

[150] Y. Xu, H. Sun, R. Q. Hu, and Y. Qian, "Cooperative non-orthogonal multiple access in heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[151] Q.-T. Vien, N. Ogbonna, H. X. Nguyen, R. Trestian, and P. Shah, "Non-orthogonal multiple access for wireless downlink in cloud radio access networks," in *Proc. Eur. Wireless Conf.*, May 2015, pp. 1–6.

[152] Z. Ding, H. Dai, and H. V. Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Aug. 2016.

[153] W. Duan, M. Wen, Y. Yan, Z. Xiong, and M. H. Lee (Apr. 2016). "*Use of non-orthogonal multiple access in dual-hop relaying.*" [Online]. Available: https://arxiv.org/abs/1604.01151

[154] F. Martin-Vega, Y. Liu, G. Gomez, M. C. Aguayo-Torres, and M. Elkashlan, "Modeling and analysis of NOMA enabled CRAN with cluster point process," Singapore, Tech. Rep., Dec. 2017.

[155] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Outage performance of full/half-duplex user relaying in NOMA systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Paris, France, May 2017, pp. 1–6.

[156] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device aided cooperative non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4467–4471, May 2017.

[157] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2478–2481, Dec. 2016.

[158] F. Liu, P. Mahonen, and M. Petrova, "Proportional fairness-based power allocation and user set selection for downlink NOMA systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[159] J. Mei, L. Yao, H. Long, and K. Zheng, "Joint user pairing and power allocation for downlink non-orthogonal multiple access systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[160] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[161] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober (Mar. 2016). "*Optimal joint power and subcarrier allocation for MC-NOMA systems*." [Online]. Available: http://arxiv.org/abs/1603.08132

[162] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Eastern Res. Lab., Digit. Equip. Corp., Hudson, MA, USA, Tech. Rep., 1984, vol. 38.

[163] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[164] F. Liu, P. Mahonen, and M. Petrova, "Proportional fairness-based user pairing and power allocation for non-orthogonal multiple access," in *Proc. IEEE Pers. Indoor Mobile Radio Commun. (PIMRC)*, Aug. 2015, pp. 1127–1131.

[165] N. Otao, Y. Kishiyama, and K. Higuchi, "Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation," in *Proc. IEEE ISWCS*, Paris, France, Aug. 2012, pp. 476–480.

[166] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, Apr. 2009.

[167] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[168] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[169] B. Di, S. Bayat, L. Song, and Y. Li, "Radio resource allocation for downlink non-orthogonal multiple access (NOMA) networks using matching theory," in *Proc.*

IEEE Global Commun. Conf. (GLOBECOM), Dec. 2015, pp. 1–6.

[170] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops*, Dec. 2014, pp. 1026–1031.

[171] M. J. Yang and H. Y. Hsieh, "Moving towards non-orthogonal multiple access in next-generation wireless access networks," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Jun. 2015, pp. 5633–5638.

[172] S. Zhang, B. Di, L. Song, and Y. Li, "Radio resource allocation for non-orthogonal multiple access (NOMA) relay network using matching game," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[173] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy efficiency of resource scheduling for non-orthogonal multiple access (NOMA) wireless network," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–5.

[174] M. Elsaadany and W. Hamouda, "Energy efficient design for non-orthogonal AF relaying in underlay spectrum sharing networks," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[175] Y. Fu, Y. Chen, and C. S. Sung, "Distributed downlink power control for the non-orthogonal multiple access system with two interfering cells," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[176] H. Zhang, D.-K. Zhang, W.-X. Meng, and C. Li, "User pairing algorithm with SIC in non-orthogonal multiple access system," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[177] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[178] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[179] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2159–2168, Jul. 2010.

[180] L. P. Qian and Y. Jun, "Monotonic optimization for non-concave power control in multiuser multicarrier network systems," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2009, pp. 172–180.

[181] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: Fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[182] S. Zhang, B. Di, L. Song, and Y. Li (Nov. 2016). "*Sub-channel and power allocation for non-orthogonal multiple access relay networks with amplify-and-forward protocol*." [Online]. Available: https://arxiv.org/abs/1611.04235

[183] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, M. Elkashlan, and J. Alonso-Zarate, "NOMA-based D2D communications towards 5G," in *Proc. IEEE Proc. Global Commun. Conf. (GLOBECOM)*, Washington DC, USA, Dec. 2016, pp. 1–6.

[184] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, A. Nallanathan, and Z. Han, "Resource allocation for non-orthogonal multiple access in heterogeneous networks," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Paris, France, 2017, pp. 1–6.

[185] N. Kokash, "An introduction to heuristic algorithms," Ph.D. dissertation, Dept. Informat. Telecommun., 2005, pp. 1–8.

[186] D. Kreutz, F. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015.

[187] D. Alanis, P. Botsinis, Z. Babar, S. X. Ng, and L. Hanzo, "Non-dominated quantum iterative routing optimization for wireless multihop networks," *IEEE Access*, vol. 3, pp. 1704–1728, 2015.

[188] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, Jun. 2015.

[189] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and A. Nallanathan, "Non-orthogonal multiple access in massive MIMO aided heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[190] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[191] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[192] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan (2017). "*Optimal user scheduling and power allocation for millimeter wave NOMA systems*." [Online]. Available: https://arxiv.org/abs/1705.03064

[193] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, 2000.

[194] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[195] Y. C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[196] Z. Qin, Y. Gao, and C. G. Parini, "Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1174–1185, Feb. 2016.

[197] Z. Qin, Y. Gao, M. Plumbley, and C. Parini, "Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3106–3117, Jun. 2016.

[198] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[199] G. Fodor, *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.

[200] C. Ma, W. Wu, Y. Cui, and X. Wang, "On the performance of successive interference cancellation in D2D-enabled cellular networks," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Kowloon, Hong Kong, Apr. 2015, pp. 37–45.

[201] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[202] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[203] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Pers. Commun.*, vol. 2, no. 2, pp. 46–58, Apr. 1995.

[204] X. Yue, Y. Liu, L. Rongke, A. Nallanathan, and Z. Ding, "Full/half-duplex relay selection for cooperative NOMA networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017.

[205] J. Zhao, Y. Liu, Z. Han, K. K. Chai, Y. Chen, and A. Nallanathan, "Resource allocation in cache-enabled CRAN with non-orthogonal multiple access," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017.

[206] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1450–1460, Aug. 2001.

[207] R. M. Buehrer, "Equal BER performance in linear successive interference cancellation for CDMA systems," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1250–1258, Jul. 2001.

[208] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[209] A. Mukherjee and A. Swindlehurst, "Robust beamforming for security in MIMO wiretap channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 351–361, Jan. 2011.

[210] Z. Ding, K. K. Leung, D. L. Goeckel, and D. Towsley, "On the application of cooperative transmission to secrecy communications," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 359–368, Feb. 2012.

[211] Y. Liu, L. Wang, T. T. Duy, M. Elkashlan, and T. Q. Duong, "Relay selection for security enhancement in cognitive relay networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 1, pp. 46–49, Feb. 2015.

[212] Y. Zou, X. Wang, W. Shen, and L. Hanzo, "Security versus reliability analysis of opportunistic relaying," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2653–2661, Jul. 2014.

[213] Y. Liu, L. Wang, S. A. R. Zaidi, M. Elkashlan, and T. Q. Duong, "Secure D2D communication in large-scale cognitive cellular networks with wireless power transfer," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Jun. 2015, pp. 4309–4314.

[214] Y. Zhang, H. M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 930–933, May 2016.

[215] B. He, A. Liu, N. Yang, and V. K. N. Lau (Dec. 2016). "*On the design of secure non-orthogonal multiple access systems*." [Online]. Available: https://arxiv.org/abs/1612.06961

[216] Z. Ding, Z. Zhao, M. Peng, and H. V. Poor (2016). "*On the spectral efficiency and*

security enhancements of noma assisted multicast-unicast streaming." [Online]. Available: https://arxiv.org/abs/1611.05413

[217] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[218] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3622–3636, Jul. 2013.

[219] Y. Liu, L. Wang, M. Elkashlan, T. Q. Duong, and A. Nallanathan, "Two-way relaying networks with wireless power transfer: Policies design and throughput analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4030–4035.

[220] Z. Qin, Y. Liu, Y. Gao, M. Elkashlan, and A. Nallanathan, "Throughput analysis for compressive spectrum sensing with wireless power transfer," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[221] Y. Liu, L. Wang, S. A. R. Zaidi, M. Elkashlan, and T. Q. Duong, "Secure D2D communication in large-scale cognitive cellular networks: A wireless power transfer model," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 329–342, Jan. 2016.

[222] Y. Liu, S. A. Mousavifar, Y. Deng, C. Leung, and M. Elkashlan, "Wireless energy harvesting in a cognitive relay network," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2498–2508, Apr. 2016.

[223] Y. Liu, Z. Ding, M. Eikashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems with SWIPT," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 1999–2003.

[224] R. Sun, Y. Wang, X. Wang, and Y. Zhang, "Transceiver design for cooperative non-orthogonal multiple access systems with wireless energy transfer," *IET Commun.*, vol. 10, no. 15, pp. 1947–1955, 2016.

[225] H. Lee, S. Kim, and J. H. Lim, "Multiuser superposition transmission (MUST) for LTE-A systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, May 2016, pp. 1–6.

[226] *Study on Downlink Multiuser Superposition Transmission for LTE*, document, 3rd Generation Partnership Project (3GPP), Mar. 2015.

[227] *TP for Classification of MUST Schemes*, document R1-154999, 3rd Generation Partnership Project (3GPP), Mar. 2015.

[228] Y. Yuan, *et al.*, "Non-orthogonal transmission technology in LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 68–74, Jul. 2016.

[229] L. Zhang, "Layered-division-multiplexing: Theory and practice," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 216–232, Mar. 2016.

[230] H. Haci, H. Zhu, and J. Wang, "Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1319–1335, Mar. 2017.

[231] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[232] K. Kusume, G. Bauch, and W. Utschick, "IDMA vs. CDMA: Analysis and comparison of two multiple access schemes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 78–87, Jan. 2012.

[233] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[234] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[235] D. Fang, Y.-C. Huang, Z. Ding, G. Geraci, S.-L. Shieh, and H. Claussen, "Lattice partition multiple access: A new method of downlink non-orthogonal multiuser transmissions," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[236] R. Hoshyar, R. Razavi, and M. Al-Imari, "LDS-OFDM an efficient multiple access technique," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2010, pp. 1–5.

[237] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.

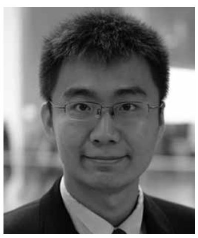
[238] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Sep. 2014, pp. 1–5.

[239] J. Bao, Z. Ma, Z. Ding, G. K. Karagiannidis, and Z. Zhu, "On the design of multiuser codebooks for uplink SCMA systems," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1920–1923, Oct. 2016.

[240] J. Peng, W. Chen, B. Bai, X. Guo, and C. Sun, "Joint optimization of constellation with mapping matrix for SCMA codebook design," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 264–268, Mar. 2017.

[241] B. Ren, *et al.*, "Advanced IDD receiver for PDMA uplink system," in *Proc. IEEE Int. Commun. Conf. China (ICCC)*, Jul. 2016, pp. 1–6.

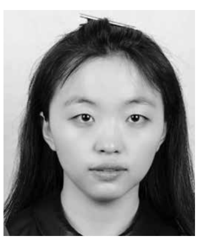
ABOUT THE AUTHORS



Yuanwei Liu (Member, IEEE) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, London, U.K., in 2016.

He has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, since 2017. He was with the Department of Informatics, King's College London, from 2016 to 2017, where he was a Postdoctoral Research Fellow. His research interests include 5G wireless networks, Internet of Things, stochastic geometry, and matching theory.

Dr. Liu received the Exemplary Reviewer Certificate of the IEEE Wireless Communication Letters in 2015 and the IEEE Transactions on Communications in 2017. He has served as a TPC Member for many IEEE conferences, such as GLOBECOM and ICC. He currently serves as an Editor of the IEEE Communications Letters and IEEE Access.



Zhijin Qin (Member, IEEE) received the B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2012 and the Ph.D. degree from Queen Mary University of London, London, U.K., in 2016.

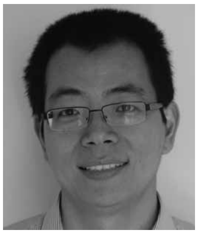
She was with Imperial College London, London, U.K., as a Postdoctoral Research Associate from May 2016 to July 2017. Since August 2017, she has been with the School of Computing and Communications, Lancaster University, Lancaster, U.K., as a Lecturer (Assistant Professor). Her research interests include low-power wide-area network for Internet of Things, compressive sensing and machine learning in wireless communications, and nonorthogonal multiple access for 5G networks.

She has served as the Cochair of the IEEE WoWMoM CORAL Workshop 2017, and a TPC member for many IEEE conferences such as GLOBECOM, ICC, and VTC. She won the best paper award at the Wireless Technology Symposium 2012.

Maged Elkashlan (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 2006.

From 2006 to 2007, he was with the Laboratory for Advanced Networking, University of British Columbia. From 2007 to 2011, he was with the Wireless and Networking Technologies Laboratory, Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. During this time, he held an adjunct appointment at the University of Technology Sydney, Australia. In 2011, he joined the School of Electronic Engineering and Computer Science at Queen Mary University of London, London, U.K., as an Assistant Professor. He also holds visiting faculty appointments at the University of New South Wales, Australia, and Beijing University of Posts and Telecommunications, Beijing, China. His research interests fall into the broad areas of communication theory, wireless communications, and statistical signal processing for distributed data processing and heterogeneous networks.

Dr. Elkashlan currently serves as an Editor of the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, and IEEE Communications Letters. He also serves as Lead Guest Editor for the special issue on "Green Media: The Future of Wireless Multimedia Networks" of the IEEE Wireless Communications Magazine, Lead Guest Editor for the Special Issue on "Millimeter Wave Communications for 5G" of the IEEE Communications Magazine, Guest Editor for the Special Issue on "Energy Harvesting Communications" of the IEEE Communications Magazine, and Guest Editor for the Special Issue on "Location Awareness for Radios and Networks" of the IEEE Journal on Selected Areas in Communications. He received the Best Paper Award at the IEEE International Conference on Communications (ICC) in 2014, the International Conference on Communications and Networking in China (CHINACOM) in 2014, and the IEEE Vehicular Technology Conference (VTC-Spring) in 2013. He received the Exemplary Reviewer Certificate of the IEEE Communications Letters in 2012.



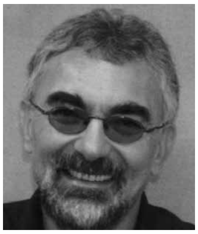
Zhiguo Ding (Member, IEEE) received the B.Eng. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2000 and the Ph.D. degree in electrical engineering from Imperial College London, London, U.K., in 2005.

From July 2005 to August 2014, he was with Queen's University Belfast, Imperial College London, and Newcastle University. Since September 2014, he has been with Lancaster University, Lancaster, U.K., as a Chair Professor. From October 2012 to September 2018, he serves as an academic visitor at Princeton University, Princeton, NJ, USA. His research interests are 5G networks, game theory, cooperative and energy harvesting networks, and statistical signal processing.

Dr. Ding serves as an Editor for the IEEE Transactions on Communications, the IEEE Transactions on Vehicular Technology, and the *Journal of Wireless Communications and Mobile Computing*, and was an Editor for IEEE Wireless Communication Letters and IEEE Communication Letters from 2013 to 2016. He received the best paper award at the IET Communications Conference on Wireless, Mobile and Computing 2009, the IEEE Communication Letters Exemplary Reviewer 2012 Award, and the EU Marie Curie Fellowship 2012–2014.

Arumugam Nallanathan (Fellow, IEEE) has been a Professor of Wireless Communications and Head of the Communication Systems Research (CSR) group in the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K., since September 2017. He was with the Department of Informatics at King's College London from December 2007 to August 2017, where he was a Professor of Wireless Communications from April 2013 to August 2017. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from August 2000 to December 2007. His research interests include 5G wireless networks, Internet of Things (IoT), and molecular communications. He published more than 350 technical papers in scientific journals and international conferences.

Dr. Nallanathan is a corecipient of the Best Paper Award presented at the IEEE International Conference on Communications 2016 (ICC2016) and the IEEE International Conference on Ultra-Wideband 2007 (ICUWB 2007). He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science (ISI) Highly Cited Researcher in 2016. He is an Editor for the IEEE Transactions on Communications. He was an Editor for the IEEE Transactions on Wireless Communications (2006–2011), the IEEE Transactions on Vehicular Technology (2006–2017), IEEE Wireless Communications Letters, and IEEE Signal Processing Letters. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of the IEEE Communications Society and Technical Program Chair and member of Technical Program Committees at numerous IEEE conferences. He received the IEEE Communications Society SPCE outstanding service award 2012 and the IEEE Communications Society RCC outstanding service award 2014.



Lajos Hanzo (Fellow, IEEE) received the M.S. degree in electronics in 1976 and the Ph.D. degree from the Technical University of Budapest, Budapest, Hungary, in 1983. He received the prestigious Doctor of Sciences Research degree in wireless communications from the University of Southampton, Southampton, U.K., in 2004.

In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the United Kingdom. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the Chair in Telecommunications. He has successfully supervised 112 Ph.D. students, coauthored 18 Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1696 research contributions on IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses.

Dr. Hanzo is a Governor of the IEEE Vehicular Technology Society (VTS). During 2008–2012, he was the Editor-in-Chief of IEEE Press and a Chaired Professor at Tsinghua University, Beijing, China. He has 30 000+ citations.