Multiple Access Techniques for Intelligent and Multi-Functional 6G: Tutorial, Survey, and Outlook

Bruno Clerckx, *Fellow, IEEE,* Yijie Mao, *Member, IEEE,* Zhaohui Yang, *Member, IEEE,* Mingzhe Chen, *Member, IEEE,* Ahmed Alkhateeb, *Member, IEEE,* Liang Liu, *Senior Member, IEEE,* Min Qiu, *Member, IEEE,* Jinhong Yuan, *Fellow, IEEE,* Vincent W.S. Wong, *Fellow, IEEE,* and Juan Montojo, *Senior Member, IEEE*

(invited paper)

***Abstract*—Multiple access (MA) is a crucial part of any wireless system and refers to techniques that make use of the resource dimensions (e.g., time, frequency, power, antenna, code, message, etc) to serve multiple users/devices/machines/services, ideally in the most efficient way. Given the increasing needs of multi-functional wireless networks for integrated communications, sensing, localization, computing, coupled with the surge of machine learning / artificial intelligence (AI) in wireless networks, MA techniques are expected to experience a paradigm shift in 6G and beyond. In this paper, we provide a tutorial, survey and outlook of past, emerging and future MA techniques and pay a particular attention to how wireless network intelligence and multi-functionality will lead to a re-thinking of those techniques. The paper starts with an overview of orthogonal, physical layer multicasting, space domain, power domain, rate-splitting, code domain MAs, MAs in other domains, and random access, and highlight the importance of conducting research in universal multiple access (UMA) to shrink instead of grow the knowledge tree of MA schemes by providing a unified understanding of MA schemes across all resource dimensions. It then jumps into rethinking MA schemes in the era of wireless network intelligence, covering AI for MA such as AI-empowered resource allocation, optimization, channel estimation, receiver designs, for different MA schemes, and MA for AI such as federated learning/edge intelligence and over-the-air computation. We then discuss MA for network multi-functionality and the interplay between MA and integrated sensing, localization, and communications, covering MA for joint sensing and communications, multimodal sensing-aided communications, multimodal sensing and digital twin-assisted communications, and communication-aided sensing/localization systems. We finish with studying MA for emerging intelligent applications such as semantic communications, virtual reality, and smart radio and reconfigurable intelligent surfaces, before presenting a roadmap toward 6G standardization. Throughout the text, we also point out numerous directions that are promising for future research.**

***Index Terms*—Multiple Access, Orthogonal Multiple Access, Non-Orthogonal Multiple Access, Space Division Multiple Access, Code Domain Multiple Access, Rate-Splitting Multiple Access, Universal Multiple Access, Artificial Intelligence, Machine Learning, Integrated Sensing and Communications, Semantic Communications, Reconfigurable Intelligent Surfaces, Augmented Reality, Internet-of-Things, 6G.**

Bruno Clerckx is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K (email: b.clerckx@imperial.ac.uk).

Yijie Mao is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: maoyj@shanghaitech.edu.cn).

Zhaohui Yang is with College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, (e-mail: yang_zhaohui@zju.edu.cn).

Mingzhe Chen is with the Department of Electrical and Computer Engineering and Frost Institute for Data Science and Computing, University of Miami, Coral Gables, FL, 33146 USA (Email: mingzhe.chen@miami.edu).

Ahmed Alkhateeb is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: alkhateeb@asu.edu).

Liang Liu is with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China (e-mail: liang-eie.liu@polyu.edu.hk).

Min Qiu and Jinhong Yuan are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: min.qiu@unsw.edu.au; j.yuan@unsw.edu.au).

Vincent W.S. Wong is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vincentw@ece.ubc.ca).

Juan Montojo is with Corporate R&D, Qualcomm Inc., San Diego, CA (e-mail: juanm@qti.qualcomm.com).

The work of Bruno Clerckx was supported in part by UKRI Impact Acceleration Account (IAA) grant EP/X52556X/1 and in part by by UKRI grant EP/X040569/1, EP/Y037197/1, EP/X04047X/1, EP/Y037243/1. The work of Yijie Mao was supported in part by the National Nature Science Foundation of China under Grant 62201347; in part by Shanghai Sailing Program under Grant 22YF1428400; in part by Young Elite Scientists Sponsorship Program by CAST 2023QNRC001, Zhejiang Key R&D Program under Grant 2023C01021, the Fundamental Research Funds for the Central Universities, K2023QA0AL02. The work of Mingzhe Chen was supported in part by the U.S. National Science Foundation under Grants CNS-2312139 and CNS-2332834. The work of Ahmed Alkhateeb was supported in part by the National Science Foundation under Grant No. 2048021. The work of Liang Liu was supported in part by the Research Grants Council, Hong Kong SAR, China, with Grant No. 25215020 and Grant No. 15203222. The work of Min Qiu and Jinhong Yuan was supported in part by the Australian Research Council (ARC) Discovery Project under Grant DP220103596, and in part by the ARC Linkage Project under Grant LP200301482. The work of Vincent Wong was supported in part by the Natural Sciences and Engineering Council of Canada (NSERC) and the Digital Research Alliance of Canada (alliancecan.ca).

I. Introduction

A. From Communication-centric 5G to Intelligent and Multi-functional 6G

NExt generation wireless networks, such as 6G and beyond, will face challenges such as higher data throughput and spectral/energy efficiency, enhanced reliability, massive connectivity, global coverage across terrestrial and non-terrestrial networks, and a growing heterogeneity in the quality of service (QoS) to meet the demands of further-enhanced mobile broadband (FeMBB) for augmented reality (AR) / virtual reality (VR); extremely ultra reliable and low-latency communication (eURLLC) for full automation, control, and operation in industrial environment and connected robotics; and ultra massive machine type communication (umMTC) for Internet-of-Things (IoT) [1]. Importantly, future wireless networks will not only provide the conventional communication functionality but also offer a wide range of new functionalities such as sensing, intelligence, computation, localization/navigation, powering. This tendency for multi-functional networks is well exemplified by the increasing number of research areas studying various forms of wireless systems integration. Indeed, radio waves can be used for communications, but also power in the form of wireless energy harvesting and wireless power transfer, sensing in the form of radar, localization, etc. Those areas have traditionally been studied separately and have led to different disciplines within Electrical and Electronic Engineering but have also evolved into vastly different industry sectors [2]. In the past decade, the community has progressively experienced a paradigm shift in wireless network design, namely, unifying transmission and processing of many quantities and functionalities, such as information, power, sensing, localization etc so as to make the best use of the radio frequency (RF) spectrum and radiation as well as the network infrastructure for the multi-purpose of communicating, energizing, sensing, locating, computing, but also for the synergies that all those disciplines can bring to each other once properly integrated. This quest for integration, convergence, and multi-functionality has led to the new research areas of integrated sensing and communications (ISAC), integrated sensing, localization, and communications, wireless information and power transfer (WIPT), edge computing and intelligence, and integrated artificial intelligence (AI) and communications [1], [3]. Additionally, network intelligence, using AI and machine learning (ML), will become pervasive in the design, control and optimization of the multi-functional networks and the network itself will become the underpinning tool to enable AI applications [4].

The above trend is much reflected in IMT-2030 framework from ITU that sets the tone for the usage scenarios, capabilities, and requirements of 6G [5]. Six main usage scenarios of immersive communication, hyper reliable and low-latency communication, massive communication, ubiquitous connectivity, AI and communication, and ISAC have been identified. The first three are extension from IMT-2020 (5G) on eMBB, mMTC, URLLC, while the last three are new and stress the importance of intelligence and multi-functionality (in the form of sensing and communications) in 6G. Moreover fifteen capabilities have been selected, including nine of which directly enhancing current 5G capabilities, such as security and resilience, reliability, latency, mobility, connection density, area traffic capacity, spectrum efficiency, user experienced data rate, and peak data rate; and six new capabilities such as positioning, interoperability, sustainability, applicable AI-related capabilities, sensing-related capabilities, coverage.

B. The Crucial Role of Multiple Access

Capturing multi-functionality and intelligence in future wireless network design will enable using wireless to its full potential, hence enabling trillions of future intelligent users to sense, compute, connect, energize, analyze anywhere, anytime, and on the move [6], [7]. One major challenge and opportunity that such intelligent multi-functional network brings is that the notion of "wireless networks" and "users" should be understood in a much more broader context compared to 1G–5G era. In multi-functional networks, users refer to communication devices, sensing targets, devices to be charged, AI nodes, training devices, or any other form of services that the network could provide.

At the core of wireless network design lies the multiple access (MA) technique whose pivotal role is to serve and process all these "users" and decide how to allocate them resources, including time, frequency, power, space (e.g., antennas, beams), signal (e.g., messages, codes, etc), in the most efficient way. The design of future intelligent and multi-functional networks brings new challenges and opportunities for wireless network designers and in particular when it comes to MA. It is crucial to comprehend how MA techniques can address these future demands and how they need to be re-thought in light of the network multi-functionality and intelligence paradigms.

Time/frequency domain multiple access (TDMA/FDMA) were popular in 2G, code-division multiple access (CDMA) in 3G, orthogonal frequency division multiple access (OFDMA) coupled with space-division multiple access (SDMA) in 4G and 5G. Though SDMA-OFDMA has remained the dominant MA in the past 20 years, the past decade has also seen a wide interest in other forms of MA schemes, often classified into non-orthogonal multiple access (NOMA) [8], [9]. Classifying MA techniques has been challenging due the proliferation of new schemes in the past decade. Unfortunately, the widely used classification into non-orthogonal MA vs orthogonal MA, i.e., NOMA vs OMA, is over-simplistic and tends to amalgamate many different MA schemes under the non-orthogonal umbrella without contrasting them or truly understanding the essence of all those schemes [10]. Such classification has caused unnecessary confusions and misunderstandings in the past few years [11]. Instead of contrasting orthogonal vs non-orthogonal, [10] suggested that a different classification should be considered in next generation wireless networks and showed that the fundamental question behind MA design should instead be how to manage multi-user interference. Answering this question shed the light on the differences between various non-orthogonal approaches to MA designs and on a new classification of MA schemes based on how the interference is managed. Importantly, this exercise brought to light the powerful and emerging rate-splitting multiple access (RSMA) that unifies into a single MA scheme four seemingly unrelated strategies, namely OMA, power domain (PD)-NOMA, SDMA, and physical-layer multicasting [12]. The capability of RSMA to unify and therefore be more universal than other MA schemes makes practical implementation and operation easier. Indeed, one could claim that a single unified and general MA scheme would be easier to implement and optimize than a combination of multiple MA schemes, each optimized for specific conditions. This can be increasingly important in multi-functional 6G and beyond networks where the range and diversity of services, use cases, and deployments explode.

C. Objectives and Organization

This paper has three objectives: 1) a tutorial paper to educate the readers about the fundamentals of a wide range MA schemes, 2) a survey paper to give the readers access to the state-of-the-art MA schemes and literature, and 3) an outlook paper to guide the readers with new research directions. Specifically the paper contributes in the following ways.

First, we provide a tutorial and survey (in Section II) of a wide range of MA techniques: OMA in the form of TDMA/FDMA/OFDMA, SDMA [13], and NOMA [14]; RSMA [10], [15]; code domain multiple access (CD-MA) departing from traditional 3G CDMA [16] including low-density spreading multiple access (LDS-CDMA), sparse code multiple access (SCMA), multi-user shared multiple access (MUSA), successive interference cancellation amenable multiple access (SAMA); and other MA schemes exploiting other domains such as interleave-division multiple access (IDMA), pattern division multiple access (PDMA), compute-forward multiple access (CFMA), lattice-partition-based multiple access (LPBMA), spatially coupled multiple access, layered-division multiplexing (LDM), index modulation multiple access (IMMA), delay-Doppler domain multiple access (DDMA), etc. Random access finally concludes the overview of MA schemes. The advantages and disadvantages of those MA schemes and the interplay between them are discussed, before drawing observations and conclusions on how to rethink the role and design of MA schemes. This overview departs from the conventional discussion on orthogonal versus non-orthogonal approaches found in 5G [10] and recent tutorials [10], [11], [14], [15], [17]–[21] whose focus was on RSMA and its sub-MA schemes OMA, SDMA, NOMA, but not on code domain and random access approaches nor on other domains MA.

Second, we discuss and motivate (in Section II) the importance of shrinking the knowledge tree of the MA literature in order to identify MA schemes that can exploit multiple dimensions and unify them. Inspired by RSMA that shrinks the knowledge tree by providing a unified and conceptually simple understanding of a morass of results on OMA, SDMA, NOMA, physical-layer multicasting, we motivate future research toward universal/unified multiple access (UMA). UMA should further shrink the knowledge tree of MA schemes by unifying RSMA with all other dimensions, such as code domain MAs, and ultimately provide a unified and conceptually simple understanding of the current and future morass of MA schemes as illustrated in Fig. 1. Clearly, there is no UMA yet that optimally suits all applications and use cases, but such scheme and research directions are expected to become increasingly important in multi-functional 6G and beyond networks where the range and diversity of services, use cases, and deployments explode. This research avenue and vision has not been discussed in prior works.

Third, we discuss MA in the era of network intelligence and the interplay between AI and MA (in Sections III and IV). We first identify how AI can be leveraged to enhance MA designs (in Section III). This includes AI-empowered resource allocation and optimization for different MA schemes, AI-empowered channel estimation for different MA schemes, and AI-empowered receiver designs for advanced MA schemes. We then investigate the converse, namely how to design MA schemes for AI applications (in Section IV). We here touch upon MA for federated learning/edge intelligence and over-the-air computation. We conclude the discussion by identifying further research avenues on the interplay between MA and AI. Such treatment significantly departs from other tutorials [10], [11], [15], [17]–[20]. This is the first tutorial paper providing a comprehensive review of the interplay between AI and MA, addressing both AI for MA and MA for AI.

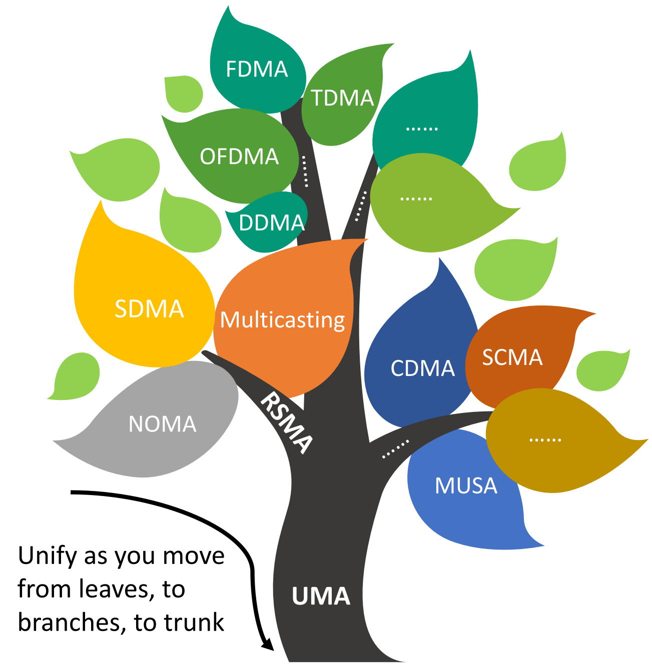


Fig. 1. Shrinking the knowledge tree of MA by unifying MA schemes as we move from leaves, to the branches, to the trunk. UMA, not yet found, would be the holy grail of MA scheme unification.

Fourth, we discuss MA in the era of network multi-functionality and the interplay between MA and integrated sensing, localization, and communications (in Section V). We elaborate on how MA designs should be tailored for joint sensing and communications, for multimodal sensing and digital twin-assisted communications, and for communication-aided sensing/localization systems. We identify the shortcoming of existing MA schemes in multi-functional networks and interesting research topics for future works. This treatment differs from other tutorials such as [19] that focuses on demonstrating RSMA superiority over SDMA and NOMA in ISAC or [22] that does not focus on MA designs for ISAC.

Fifth, we discuss MA designs for emerging intelligence applications (in Section VI) such as semantic communications, virtual reality, and smart radio with reconfigurable intelligent surfaces, before presenting a roadmap toward 6G standardization (in Section VII). This differs from other tutorials such as [10], [15], [17] that focus on RSMA and NOMA for some of those applications.

Table I details the main abbreviations used throughout this work.

TABLE I
LIST OF ABBREVIATIONS.

BC	Broadcast Channel	(u)mMTC	(ultra) massive Machine-Type Communication
CDMA	Code Division Multiple Access	MU–LP	Multi-User Linear Precoding
CD-MA	Code Domain Multiple Access	MU-MIMO	Multi-User Multiple-Input Multiple-Output
CFMA	Compute-Forward Multiple Access	MUSA	Multi-User Shared Multiple Access
CoMP	Coordinated Multi-Point	(PD) NOMA	(power domain) Non-Orthogonal Multiple Access
CSI	Channel State Information	NOUM	Non-Orthogonal Unicast and Multicast
CSIT/R	Channel State Information at the Transmitter/Receiver	OFDMA	Orthogonal Frequency Division Multiple Access
C-RAN	Cloud-Radio Access Networks	OMA	Orthogonal Multiple Access
DDMA	Delay-Doppler Domain Multiple Access	PDMA	Pattern Division Multiple Access
DoF	Degree-of-Freedom	QoS	Quality of Service
DPC	Dirty Paper Coding	RF	Radio Frequency
DPCRS	Dirty Paper Coded Rate-Splitting	RIS	Reconfigurable Intelligent Surfaces
(F)eMBB	(further-)enhanced Mobile Broadband Service	RS	Rate-Splitting
FDD	Frequency Division Duplex	RSMA	Rate-Splitting Multiple Access
FDMA	Frequency Division Multiple Access	SAMA	Successive Interference Cancellation Amenable Multiple Access
F-RAN	Fog-Radio Access Networks	SC	Superposition Coding
GRS	Generalized Rate-Splitting	SCMA	Sparse Code Multiple Access
HRS	Hierarchical Rate-Splitting	SDMA	Space Division Multiple Access
IDMA	Interleave-Division Multiple Access	SIC	Successive Interference Cancellation
IMMA	Index Modulation Multiple Access	SISO	Single-Input Single-Output
ISAC	Integrated Sensing and Communications	SNR	Signal-to-Noise Ratio
IRS	Intelligent Reconfigurable Surface	SWIPT	Simultaneous Wireless Information and Power Transfer
LDM	Layered-Division Multiplexing	SIMO	Single-Input Multiple-Output
LDS	Low-Density Spreading	TDD	Time Division Duplex
LPBMA	Lattice-Partition-Based Multiple Access	TDMA	Time-Division Multiple Access
MA	Multiple Access	UAV	Unmanned Aerial Vehicles
MAC	Multiple Access Channel	UMA	Universal Multiple Access
MIMO	Multiple-Input Multiple-Output	URA	Unsourced Random Access
MISO	Multiple-Input Single-Output	(e)URLLC	(extremely) Ultra-Reliable Low-Latency Communication

II. AN OVERVIEW OF MULTIPLE ACCESS TECHNIQUES

In this section, we provide an overview and discuss pros and cons of OMA, Physical Layer Multicasting, SDMA, NOMA, RSMA, CD-MA, random access, and other MAs exploiting other dimensions. We draw some general observations before identifying possible paths toward UMA.

A. *Orthogonal Multiple Access*

Orthogonal multiple access (OMA) is a fundamental MA technique that has been widely used in mobile communication systems. In OMA, the radio resources at hand are strategically divided into distinct, non-overlapping frequency bands, time slots, or codes, each meticulously allocated to an individual user. OMA adheres to a typical principle of interference management, which is to prevent multi-user interference.

There are four well-established OMA strategies specifically tailored for 1G to 4G communication systems, namely, frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency division multiple access (OFDMA) [23]. Specifically, FDMA divides the available spectrum into non-overlapping frequency bands, each accommodating one user. TDMA partitions time into time slots allocated to different users. CDMA utilizes orthogonal, user-specific codes to spread the modulated user symbols, serving multiple users simultaneously in the same time-frequency resources without causing multi-user interference (under ideal propagation conditions). OFDMA divides the frequency and time resources into narrow subcarriers and time slots, which are grouped into resource units and allocated to the users.

While OMA has garnered widespread acceptance in past communication systems, its efficacy becomes restricted in light of the explosive expansion of wireless communication worldwide. Here we outline the advantages and disadvantages of OMA:

- *Advantages:*
  1) **Widespread adoption:** OMA is well-established and extensively utilized in current communication systems, making it easier for facilitating smooth and effortless network expansion.
  2) **Simplicity:** OMA simplifies transceiver design, implementation, and management.
  3) **Interference free:** OMA prevents interference between users, enabling interference-free transmissions and thereby enhancing the overall quality and reliability of communications. It excels in managing low to moderate user loads effectively. It is effective for handling low to moderate user loads.
- *Disadvantages:*
  1) **Inefficient spectrum utilization:** In OMA, there is a risk that even low-rate users, such as IoT sensors with minimal resource requirements, may occupy an entire resource block, leading to inefficient spectrum utilization.
  2) **Low capacity:** OMA allocates each orthogonal radio resource to an individual user. The system capacity is restricted by the total number of available radio resources. This limitation hinders its ability to accommodate the surging user demand experienced in modern communication systems.
  3) **High signaling overhead:** To enhance system per-

formance of OMA, it is essential to implement well-designed user scheduling, which typically results in a significant increase in signaling overhead.

B. Physical Layer Multicasting

Multicasting usually refers to the transmission of a message intended to multiple users, i.e., one-to-all. Popular example are radio and television where the message of interest to multiple users is decoded by those users. However, multicasting can also be understood in a wider context where the transmitter transmits multiple unicast (i.e., one-to-one) messages, each intended to one user, by encoding them jointly into one stream to be decoded by all users. Users would then retrieve from the decoded stream the part intended to them. The encoding and transmission over the air is effectively a physical layer multicasting since all messages are encoded into one multicast stream to be decoded by all users. The difference with conventional multicasting is that only part of the multicast stream is intended to a given user, instead of the entire stream. Physical-layer multicasting contrasts with OMA where messages are encoded in independent streams and transmitted on orthogonal resources.

- *Advantages:*
  1) **Coding gain:** By combining messages and encoding jointly multiple medium-size packets together into a single stream, the encoded stream is longer and the coding gain is increased, which leads to higher reliability. Such feature is particularly useful in non-terrestrial systems, such as geostationary satellite communications based on DVB-S2X technology [24], where each spot beam of the satellite serves more than one user simultaneously by transmitting a single coded frame. Since different beams illuminate different group of users, such satellite system follows a physical layer multigroup multicast transmission [25].
  2) **Low latency:** Since all messages are encoded jointly into one stream to be decoded by all, users can decode their messages simultaneously and do not have to wait for their message to be transmitted consecutively as in TDMA, therefore reducing latency in the network.
  3) **Interference free:** Multicasting prevents interference between users since there is only one stream transmitted, enabling interference-free transmission and thereby enhancing the overall quality and reliability of communications.
- *Disadvantages:*
  1) **Inefficient spectrum utilization:** In multicasting, all users, from the weakest to the strongest, need to be able to decode the stream. Consequently the transmission rate of the stream is always determined by the weakest user, therefore leading to inefficient spectrum and resource utilization. This can be frustrating to stronger users who could receive at higher rates but are constrained by the weakest user in the pool.

C. Space Division Multiple Access

In response to the limitations inherent in OMA and the growing demand for higher data rates, improved QoS, and increased network capacity, a new resource dimension - space - has been introduced in modern wireless networks. This gives rise to the widespread integration of multiple antennas in most wireless access points, signifying the advent of the multiple-input multiple-output (MIMO) paradigm since 4G networks. MIMO has become indispensable in modern and future wireless networks, finding inclusion in nearly all high-rate wireless standards such as WiMAX, 4G LTE, IEEE 802.11n, 5G NR, etc. By leveraging the spatial dimension, MIMO networks introduce a novel MA known as SDMA [23], [26], [27]. SDMA empowers multiple users to share the same time-frequency resources, adhering to the interference management principle of precanceling interference at the transmitter and treating interference as noise at the receivers.

To mitigate interference at the transmitter, SDMA introduces precoding techniques, which are typically classified into two primary categories: non-linear and linear precoding. One of the most renowned non-linear techniques is dirty paper coding (DPC) [26], [28]. It attains the capacity region of MIMO Gaussian broadcast channel (BC) when perfect channel state information is available at the transmitter. However, its application is hindered by the impracticality arising from its high computational demands. In contrast, multi-user linear precoding (MU-LP) offers a more practical alternative by leveraging linear precoding techniques at the transmitter while regarding multi-user interference as noise at the receivers [26], [27]. Although MU-LP cannot achieve the capacity region achieved by DPC, it is particularly useful when users possess semi-orthogonal channels and comparable signal strengths (and also perfect channel state information at the transmitter-CSIT). Therefore, it plays a pivotal role in many transmission techniques underpinning both 4G and 5G networks, including multi-user MIMO (MU-MIMO), massive MIMO, networked MIMO, and other advanced techniques.

While SDMA remains crucial in modern wireless networks, the ongoing evolution of wireless technologies compels us to scrutinize SDMA critically. In the following, we delineate the advantages and disadvantages of SDMA:

- *Advantages:*
  1) **Enhanced spectrum efficiency:** By utilizing the spatial domain, SDMA allows multiple users to efficiently share the same time-frequency resources, thereby enhancing the spectrum efficiency when CSIT is perfect and the network is underloaded.
  2) **Interference mitigation:** With perfect CSIT and an underloaded network, SDMA effectively eliminates or suppresses multi-user interference, achieving the maximum degrees-of-freedom in underloaded multi-antenna BC [29].
  3) **Low transceiver complexity:** Thanks to the utilization of linear precoding at the transmitter and each receiver's ability to directly decode the intended message while treating interference as mere noise,

SDMA exhibits a relatively low level of hardware complexity at both the transmitter and receivers.

- *Disadvantages:*
  1) **High sensitivity to CSIT inaccuracy:** The effectiveness of SDMA is highly affected by the inaccuracies in CSIT [30], [31]. While SDMA excels with perfect CSIT in underloaded network, its performance significantly decreases with imperfect CSIT due to residual interference from imprecise interference mitigation strategies that are originally designed for perfect CSIT.
  2) **Limited network load tolerance:** SDMA performs effectively in underloaded networks but experiences degradation in overloaded systems due to the constraints posed by limited spatial resources. To address this issue, user grouping is usually employed in overloaded scenarios, but at the cost of reduced QoS and increased latency.
  3) **Limited user deployment flexibility:** SDMA is sensitive to user deployment such as the angles and strengths of user channels. It works well for users with orthogonal channels and similar signal strengths. However, its performance significantly degrades when user channels become nearly aligned or exhibit substantial variations in signal strengths.
  4) **Complex scheduling:** Due to the limited user deployment flexibility, precise scheduling is imperative for SDMA, leading to extra complexity in user scheduling. Moreover, the scheduling algorithms to achieve the (near) optimal performance can be challenging to implement and maintain.
  5) **High signaling overhead:** The requirements for channel estimation, scheduling, and interference management in SDMA leads to substantial signaling overhead, particularly in dynamic and densely populated network environments or when employing a massive number of antennas at the transmitter.

D. (Power-domain) Non-Orthogonal Multiple Access

NOMA leverages the concept of superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receivers to facilitate simultaneous sharing of common resources, i.e., time, frequency, code, or space among users [14], [32], [33]. This is achieved by allocating users with varying power levels, and enabling signals from different users superposed in the power domain. NOMA ensures the effective decoding of these signals at users by empowering those with weaker power levels to decode the messages of users with stronger power levels. This approach is also referred to as SC-SIC, adhering to the interference management principle of decoding interference.

It is well-established in the literature that NOMA based on SC–SIC achieves the capacity region for single-input single-output (SISO) BC [34], and for the SISO multiple access channel (MAC) [23], it is also capacity-achieving with time-sharing. The performance merits of NOMA over OMA demonstrated in SISO BC/MAC has driven research into MIMO NOMA. In the multi-antenna BC, to exploit the spatial domain, MIMO NOMA typically separates users into distinct groups. Interference from users within the same group is managed by SC-SIC while interference from users in different groups is treated as noise. When there exists only a single user group, MIMO NOMA reduces to SC-SIC, requiring users to decode and remove all interference [33].

NOMA has not yet been incorporated into emerging wireless standards, though it has been investigated as part of a study item in 5G but was not considered any further in 5G because its gains compared to SDMA/MU-MIMO were not found convincing [35]. As highlighted in [11], there exist many confusions and misconceptions about NOMA, which impels us to look into its advantages and disadvantages in the following:

- *Advantages:*
  1) **Enhanced spectrum efficiency:** By utilizing the power domain and advanced receiver techniques, NOMA allows multiple users with closely aligned channels and diverse channel strengths in the same time-frequency resources to efficiently share the same time-frequency resources, thereby enhancing the spectrum efficiency when the network is extremely overloaded.
  2) **Enhanced user fairness:** As NOMA requires power allocation favoring users with weaker channel strength to enable successful interference cancellation, it is capable of enhancing user fairness than SDMA in extremely overloaded network with closely aligned users and large channel strength disparities.
- *Disadvantages:*
  1) **Inefficient use of spatial dimensions:** As shown in [11], the sum multiplexing gain of MIMO NOMA always falls below or equals that of SDMA. This implies that the slope of the sum-rate of MIMO NOMA at high signal-to-noise ratio (SNR) will be lower than that of SDMA. This phenomenon indicates that NOMA makes an inefficient use of the multiple antennas.
  2) **Inefficient use of SIC:** In MIMO NOMA, the number of SIC deployed at each user scales proportionally with the number of users in that user group. Consequently, compared to SDMA, MIMO NOMA introduces two notable issues: a loss in multiplexing gain and an increase in receiver complexity. This observation underscores MIMO NOMA's inefficient utilization of SIC.
  3) **High sensitivity to CSIT inaccuracy:** As the inter-group interference management in MIMO NOMA follows the approach employed in SDMA, akin to SDMA, MIMO NOMA is also sensitive to CSIT inaccuracy.
  4) **Limited network load tolerance:** NOMA has been demonstrated as a capacity-achieving scheme in SISO BC, rendering it effective in extremely overloaded scenarios. However, as the number of transmit antenna increases, MIMO NOMA suffers

from performance degradation primarily due to its inefficient use of spatial dimensions. Therefore, NOMA is sensitive to the network load.

5) **Limited user deployment flexibility:** With perfect CSIT and an underloaded network, SDMA effectively eliminates or suppresses multi-user interference, achieving the maximum degrees-of-freedom in underloaded multi-antenna BC.
6) **High transceiver complexity:** In addition to the hardware complexity introduced by SIC receivers, NOMA imposes substantial computational complexity at both transmitter and receivers. Specifically, at the transmitter, the joint optimization of user scheduling, user grouping, decoding orders, and precoders is imperative for enhancing performance. However, solving such resource allocation problem is typically challenging with high computational complexity.
7) **High signaling overhead:** In NOMA, extra signaling overhead is required to convey information such as the decoding order, user grouping result, and power levels each user should use.

E. Rate-Splitting Multiple Access

RSMA has recently emerged as a promising non-orthogonal transmission strategy for multi-antenna wireless networks, owning to its capability to enhance the system performance in a wide range of network loads, user deployment, and CSIT qualities [15], [36]–[38]. RSMA is a generalized MA scheme originally proposed in [36]–[38] for downlink and in [39] for uplink. The key concept of RSMA is splitting each user message into sub-messages at the transmitter. For downlink RSMA, sub-messages are categorized into common and private with the common sub-messages designed to be decoded by multiple users, whereas private sub-messages are intended to be decoded exclusively by their respective users. For uplink RSMA, sub-messages of a given transmitter can be decoded at the receiver in a non-consecutive manner. This message split capability endows RSMA with a flexible interference management strategy of partially decoding the interference and partially treating the interference as noise.

Depending on the specific approaches used for message splitting and combining, RSMA has a range of transmission schemes, including linearly precoded 1-layer RS, 2-layer hierarchical RS (HRS), and generalized RS (GRS), each of which can be further extended to their respective non-linear counterparts [15], [38], [40], and even to space-time designs [41].

The most straightforward and practical downlink RSMA scheme is 1-layer RS [36], [37], which is also the basic building block of almost all existing RSMA schemes. In 1-layer RS, each user message is split into one common sub-message and one private sub-message (Fig. 2). All the common sub-messages are combined and jointly encoded as a single common stream to be decoded by all users, whereas the private sub-messages are encoded individually as private streams. 1-layer RS requires only one layer of SIC at each receiver. User grouping and decoding order design is unnecessary since all users decodes the unique common stream before decoding its private stream.

HRS represents a more encompassing scheme compared to 1-layer RS, as it introduces additional common streams tailored for specific user groups (Fig. 2). In HRS, all users are divided into separate groups, each comprising one or multiple users. At the transmitter, each user message is split into three sub-messages: an inter-group common sub-message, an inner-group common sub-message, and a private sub-message [42]. The inter-group common sub-messages and private sub-messages respectively follow the common and private submessages in 1-layer RS. The primary difference between HRS and 1-layer RS lies in the inner-group common sub-messages. The inner-group sub-messages for users within the same group are merged into a group-specific common message, subsequently encoded into a inner-group common stream. The inner-group common stream is decoded exclusively by users within the corresponding user group, following their decoding of the inter-group common stream. HRS requires two layers of SIC at each receiver. Decoding order design in unnecessary since each user follows the decoding order of inter-group stream, inner-group stream, and private stream. In contrast to 1-layer RS, the user grouping requires to be considered in HRS to achieve a more flexible interference management capability than 1-layer RS.

Unlike 1-layer RS and 2-layer HRS, which maintain a constant number of message splits for each user message regardless of the number of users, GRS takes a more comprehensive approach by utilizing all possible message splitting and combining strategies to achieve generalized transmission framework [38]. In GRS, during the message splitting and combining process, all potential user grouping are explored and the groups containing different number of users are categorized into different layers (Fig. 2). For example, the group in the first layer contains all users while each group in the last layer only contains a single user. Each user message is split into a number of common sub-messages. The common sub-messages in the same group are combined and encoded following the basic principle of HRS. At the user sides, a number of SIC layers are employed at each user to decode all common streams and the intend private stream. As all possible user grouping has been considered in the message splitting and combining process, only the decoding order among the groups in the same layer need to be carefully designed. In Fig. 2, a straightforward comparison of the three-user transmission frameworks among 1-layer RS, HRS, and GRS is illustrated.

RSMA has been shown to generalize multicasting, OMA, SDMA, and NOMA schemes [12], [15]. Specifically, when all transmit power is allocated exclusively to the common stream or one private stream, RSMA respectively simplifies to the classical multicasting or OMA. When the transmit power is solely allocated to the private streams, RSMA reduces to SDMA. Furthermore, in the case where each user message is totally encoded into distinct layers of common streams and private stream within the GRS framework, RSMA reduces to NOMA. For this reason, RSMA is a superset and therefore always achieves equal or better performance compared to OMA,

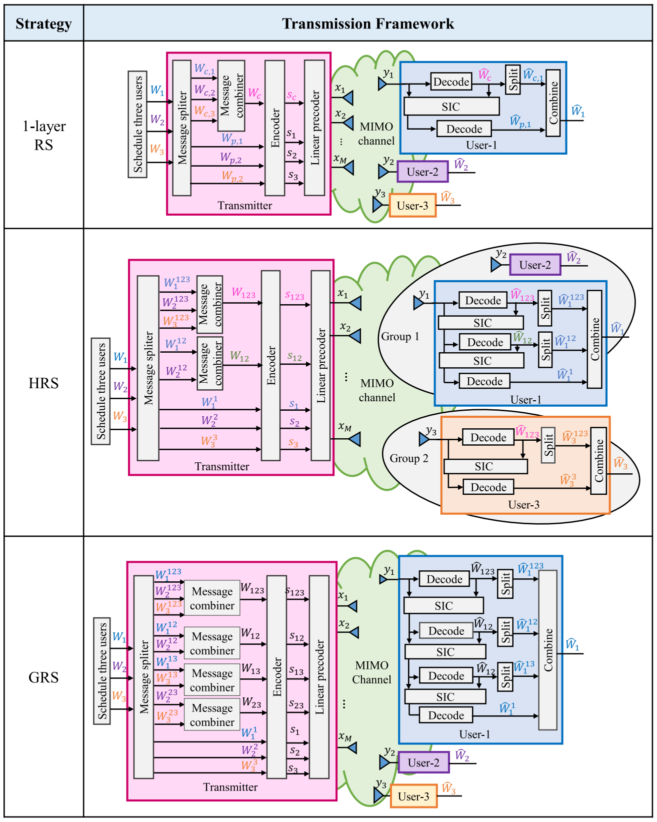


Fig. 2. Illustration of the downlink transmission frameworks for 1-layer RS, HRS, and GRS with 3 users.

SDMA, and NOMA. A simple two-user downlink transmission framework comparison among multicasting, OMA, SDMA, NOMA, and RSMA is delineated in Fig. 3, and the corresponding uplink transmission frameworks are illustrated in Fig. 4.

Fig. 5 illustrates the performance gain of RSMA over MU-MIMO using BBC's 3GPP compliant link-level simulations in a two-user high density demand (HDD) area [43]. HDD areas are important deployments for future 6G [44]. Though only around 1% of UK geographical areas served by mobile networks could be characterised as HDD, HDD areas drive an estimated 20% of total mobile traffic. HDD scenarios are dense urban areas, airports, sports venues, railway and subway stations and major public events. Most HDD scenarios occur in situations where a lot of users and devices are present in a specific place or event, putting intense demand on the network. In Fig. 5 simulations, an HDD with 20m spacing among the two users at target spectral efficiencies of 3.4 bits per resource element is considered. LDPC with finite block length and finite constellations are used along with various RSMA receivers such as hard and soft codeword interference cancellation (CWIC), soft symbol level interference cancellation (SLIC), joint demapping, with tailored RSMA precoders for common and private streams [45]. For MU-MIMO, zero-forcing (ZF), maximum ratio transmission (MRT) or regularized zero forcing (RZF) precoders are used and the receivers apply single user detector (only de-map the desired signal) with QPSK or 16QAM for modulation (MU-MIMO+ZF+QPSK/16QAM+SD), or jointly de-map (JD) with QPSK for modulation (MU-MIMO+MRT/RZF+QPSK+JD). The latter uses SDMA transmission strategy but uses an advanced receiver that demaps the desired signal and interference. MU-MIMO with advanced receivers is outperformed by RSMA since RSMA can be viewed as a joint transmit-side and receive-side interference management/cancellation strategies where the contribution of the common stream is adjusted according to the level of interference that can be canceled by the receiver, instead of a full transmit-side interference management strategy (as in SDMA) aided by an advanced receiver [11].

Fig. 6 displays experimental results of sum and minimum throughput achieved by RSMA, SDMA and NOMA using software define radio implementations and over-the-air measurements in a 2-user setting with finite block length polar codes and finite constellations [46]. Nine indoor locations are measured with various disparity of the channel strengths and angle between user channel directions. The number beside each data point indicates a measurement location. RSMA can achieve higher sum-throughput but also better fairness among users (minimum throughput of RSMA is higher). This double advantage of RSMA in boosting jointly the sum and minimum throughput originates from the split into common and private streams and from adjusting the content of the common stream so that better fairness is achieved while maximizing sum throughput.

A general observation of the many existing simulation results on RSMA shows that the benefit of RSMA in comparison with MU-MIMO/SDMA increases as users are closer together, e.g., as in high density demand areas, and/or experience a diversity of channel strengths, and/or as the system becomes more overloaded, and/or as the CSIT accuracy degrades, and/or as more fairness among users and QoS per user are captured in the design objective and constraints, and/or as the system is multi-functional such as ISAC - see [10], [11], [15] and therein. NOMA can have some performance benefits over SDMA when users are closely aligned (and potentially experience a disparity of channel strength) leading to correlated channels, or when the system is overloaded and the performance metric accounts for QoS or user fairness. Otherwise, SDMA outperforms NOMA due its higher multiplexing gain capability [11]. This is visible in Fig. 6 where SDMA can outperform or be outperformed by NOMA depending on the measurement location. RSMA, on the other hand, always outperforms SDMA and NOMA since NOMA and SDMA are particular instances of RSMA framework.

The main advantages and disadvantages of RSMA are summarized in the following:

- *Advantages:*
    1) **Universality:** RSMA is a comprehensive MA framework encompassing OMA, SDMA, NOMA, and multicasting as its constituent sub-schemes [12], [38]. The universality of RSMA obviates the need for a system to consider switching between OMA, SDMA, NOMA, and multicasting.

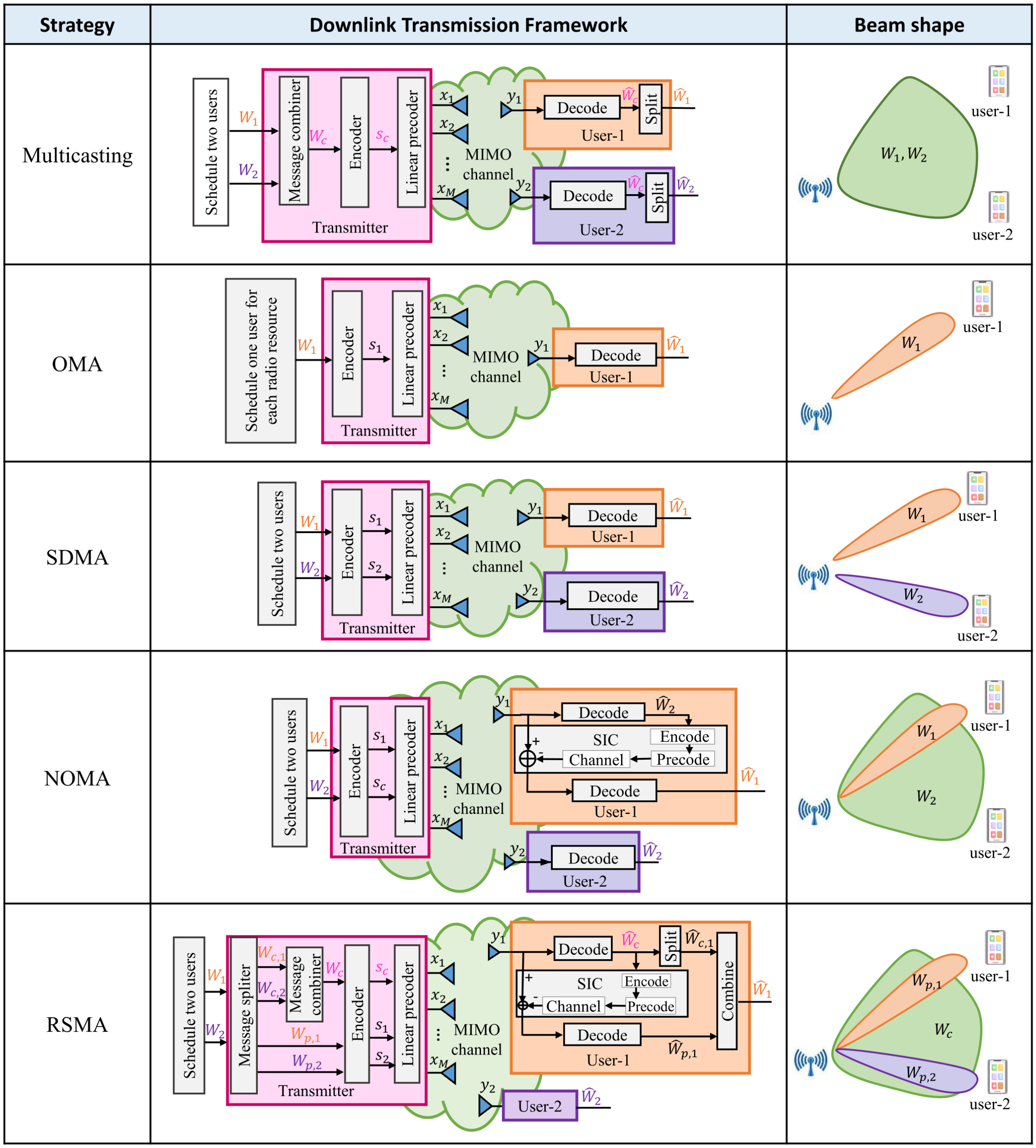


Fig. 3. Illustration of the two-user downlink transmission frameworks and beam shapes for multicasting, OMA, SDMA, NOMA, and RSMA.

2) **Flexibility:** RSMA is highly adaptable to varying network loads, whether they are underloaded or overloaded, as well as to diverse user deployments characterized by varying channel directions and strengths [38], [47], [48]. This remarkable advantage of RSMA comes from its capability to flexibly handle multi-user interference through partial interference decoding and treating the remaining interference as noise. An additional feature of RSMA is its flexibility to handle mixed services and simultaneous unicast and multicast transmission [49]. A multicast message W_0 can indeed be encoded jointly with the common parts $W_{c,1}, W_{c,2}$ of the unicast messages W_1, W_2, as illustrated in Fig. 7, to enable very efficient non-orthogonal unicast and multicast (NOUM) transmission. The SIC is used for the double purpose of separating unicast from multicast and managing interference among unicast[1].

3) **Robustness:** The rapid growth of RSMA in multi-antenna networks is primarily attributed to its capac-

[1]SDMA-assisted NOUM is a particular instance of the RSMA architecture of Fig. 7 where W_1 and W_2 are encoded into s_1 and s_2, respectively, without being split. SIC would then be used only to separate multicast from unicast but cannot manage interference between unicast due to lack of message splitting - hence less flexibility in dealing with NOUM compared to RSMA approach.

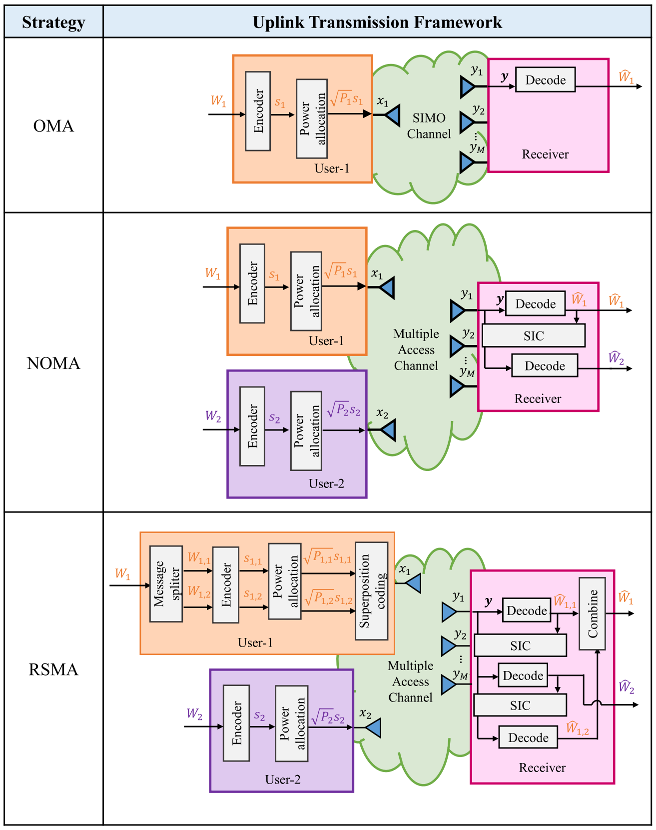


Fig. 4. Illustration of the two-user uplink transmission frameworks for OMA, NOMA, and RSMA.

ity to achieve an optimal spatial multiplexing gain in the multi-antenna BC with imperfect CSIT [37], [41], [50]–[53]. A multitude of extant research endeavors has substantiated RSMA's resilience when confronted with CSIT uncertainties stemming from diverse sources of impairment, such as quantized feedback [41], [54], pilot contamination [55], [56], channel estimation inaccuracies, latency in feedback, user mobility [57], and even RF impairments such as phase noise [58].

4) **Enhanced spectrum efficiency:** RSMA is guaranteed to achieve the best sum-degree of freedom (DoF) in both perfect and imperfect CSIT, which translates into superior spectral efficiency as well. In other words, RSMA outperforms other MA schemes in terms of spectral efficiency in both perfect and imperfect CSIT [37], [38]. When CSIT is perfect, the achievable rate region of RSMA is larger than other MA schemes, and it approaches the capacity region achieved by DPC [38]. When CSIT is imperfect, RSMA can achieve a larger rate region than DPC [31]. RSMA's enhanced spectrum efficiency is not limited to conventional multi-user multiple-input single-output (MISO) unicast transmissions, but also holds for MIMO unicast [59], [60], multigroup multicast [25], [47], [61], [62], non-orthogonal unicast and multicast [49], non-terrestrial [25], [63], [64], network slicing [65]. The benefits have been demonstrated using link-level simulations [45], [60] and real-world experimentation [46].
5) **Enhanced energy efficiency:** RSMA not only boosts spectral efficiency but also energy efficiency and their tradeoffs across various applications with different network loads and user deployments, as evidenced in many existing works [49], [66].
6) **Enhanced QoS and user fairness:** RSMA achieves the optimal max-min fair spatial multiplexing gain in multi-antenna BC with both perfect and imperfect CSIT. This advantage is reflected in the finite SNR regime, where RSMA demonstrates its superior max-min rate compared to the aforementioned MA schemes. Furthermore, the substantial spectral efficiency gain achieved by RSMA becomes even more pronounced when stringent QoS rate constraints are imposed on each individual user.
7) **Low complexity:** 1-layer RSMA stands out for its simplicity at both transmitter and receivers design. Its adaptability to varying network loads and user deployments obviates the need for user ordering, grouping, and scheduling at the transmitter. Each user only necessitates a single SIC layer to decode and cancel the common stream. This contrasts with MIMO NOMA, which imposes stringent requirements for user grouping and ordering at the transmitter along with multiple layers of SIC at each user.
8) **Coverage extension:** RSMA with user relaying, also known as cooperative rate-splitting (CRS) [67], [68], has been demonstrated as a promising strategy to amplify the data rates for users located at the cell edge, offering substantial coverage extension benefits [69]. CRS involves collaborative user relaying, enabling one user to decode and relay the common stream to other users, thereby enhancing user fairness particularly when jointly serving users with substantial discrepancies in channel strengths.
9) **Low latency:** RSMA has demonstrated its prowess in improving throughput compared to the aforementioned MA schemes when employing finite-length (e.g. polar) codes, as highlighted in [70]–[72]. In other words, RSMA achieves the same transmission rate as SDMA and NOMA but with shorter block lengths, resulting in reduced latency. Therefore, RSMA is a promising enabling technology for significantly reducing latency in URLLC services.
10) **Security enhancement:** RSMA can adjust the level of confidentiality of its messages and consequently trade off spectral efficiency with secrecy [73], [74]. This is not possible with NOMA since the message of the weaker user is always decoded by a stronger user, which creates a secrecy threat.

- *Disadvantages:*

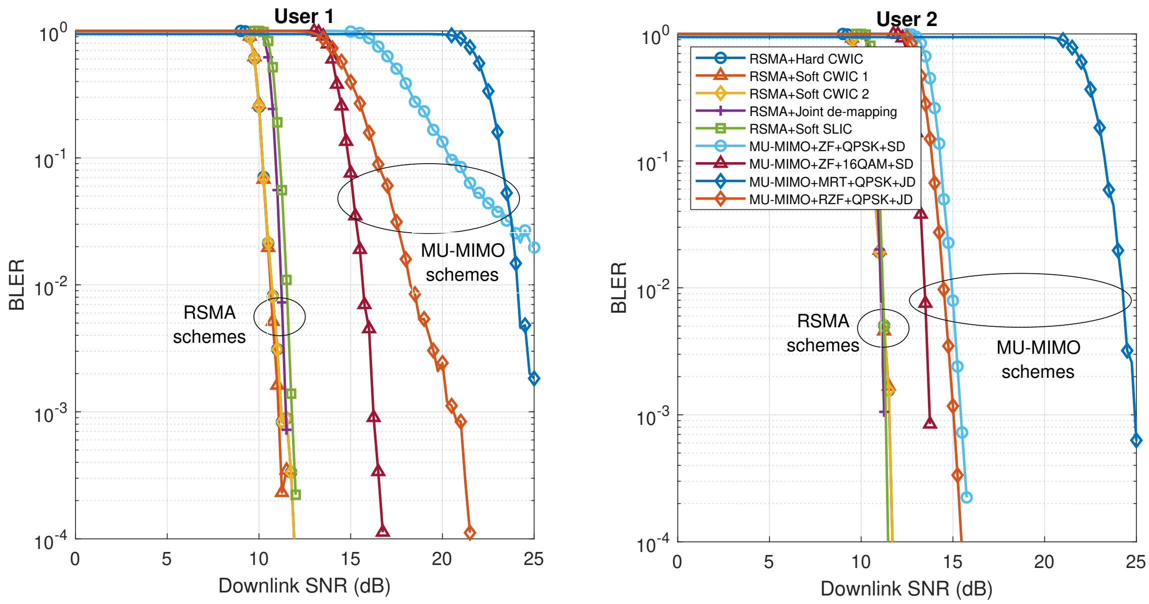


Fig. 5. Downlink link-level unicast (PDSCH) performance of RSMA and MU-MIMO at target spectral efficiencies of 3.4 bits/RE. QPSK for RSMA, QPSK/16QAM for MU-MIMO. Transport block error rate vs. SNR (dB) with realistic channel estimation [43].

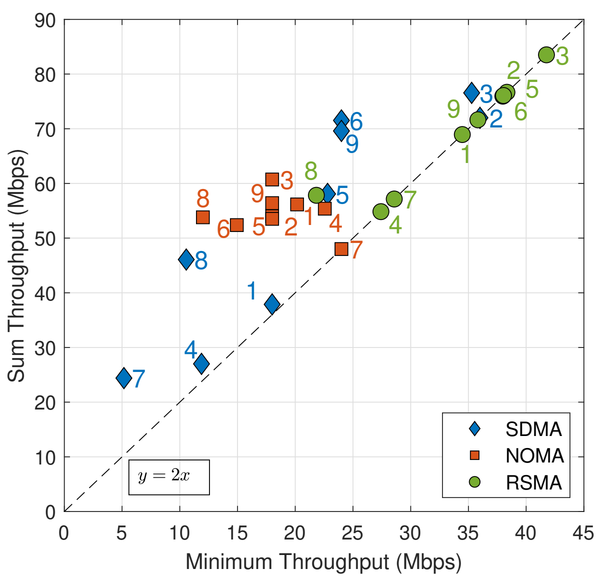


Fig. 6. Experimental results of downlink sum and minimum throughput achieved by RSMA, SDMA and NOMA. The dashed line (y = 2x) corresponds to max-min fairness and points that are closer (in terms of Euclidean distance) to this line represent more fair outcomes [46].

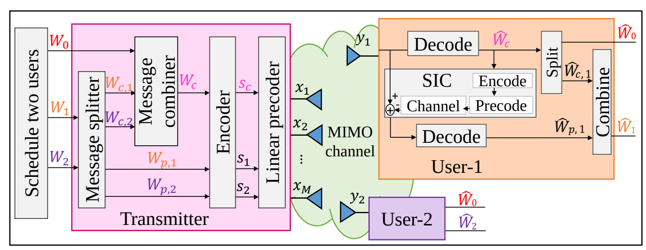


Fig. 7. RSMA as an enabler of efficient non-orthogonal unicast and multicast (NOUM) transmission [49].

1) **SIC requirement at receiver:** Some RSMA schemes, such as GRS, exhibit a characteristic where the number of SIC layers at each user increases with the number of users. This trend not only leads to considerable burden on receiver complexity but also introduces the error propagation issues. Promising SIC-free RSMA architectures have recently been developed [45].
2) **High encoding complexity:** Compared with the aforementioned MA schemes, RSMA requires more date streams to be encoded due to the additional common streams obtained from message splitting and recombining.
3) **High signaling burden:** In RSMA, additional signaling overhead is necessary to facilitate alignment between the transmitter and receivers, ensuring they possess a shared understanding of the methodology for splitting and combining user messages.
4) **High optimization burden:** In terms of resource allocation and precoder design, RSMA requires the precoders (accounting for beamformer directions and power allocation) of the common and private streams to be jointly optimized with the common rate allocation. This joint optimization is instrumental in unlocking the complete benefits of RSMA. However, it is worth noting that this expanded optimization space places a more demanding computational burden on RSMA in comparison to conventional MA schemes.

F. Code-Domain Multiple Access

Code-domain multiple access (CD-MA) is inspired by the traditional CDMA [16]. It allows multiple users to share the same time/frequency resources while dedicated interleavers and/or code sequences are employed to multiplex users. However, unlike CDMA, the spreading sequences in CD-MA are usually sparse or non-orthogonal low cross-correlation sequences. In general, code-domain multiple access can be divided into *sparse CD-MA* and *dense CD-MA* [75], [76], depending on the sparsity of the spreading sequences. In this section, we review some popular CD-MA schemes, compare their performance, and discuss their main advantages and disadvantages.

1) Low-Density Spreading (LDS) MA: One of the early CD-MA techniques is the LDS-CDMA [77], an extension to the classical CDMA. In an LDS-CDMA system, the data symbols of each user are spread by a unique LDS with a small fraction of non-zero entries and then superimposed before transmission. Compared to the classical CDMA, the interference on each chip can be reduced in LDS-CDMA. In particular, the multi-user interference pattern at the receiver entails a low-density graph, where message passing algorithm (MPA) or belief propagation (BP) [78] can be adopted for efficient symbol detection [79]. To design LDS sequences, a structural approach was proposed in [80]. Furthermore, the capacity region of LDS-CDMA was investigated in [81], where the impacts of spreading sequence density factor and the maximum number of users associated with each chip on the capacity were revealed. In addition, the idea of LDS-CDMA was extended to OFDM (LDS-OFDM) [82] where each user's data symbols are spread across a number of carefully selected subcarriers. Significant improvement of peak-to-average-power ratio (PAPR) the link-level performance was reported for LDS-OFDM over OFDM [83], [84].

2) Sparse Code Multiple Access (SCMA): The concept of code domain multiplexing in LDS-CDMA was extended to SCMA [85]. In SCMA, each user is assigned a sparse codeword according to its message. In other words, the modulation symbol mapping and the spreading sequences are merged together such that the bits are directly mapped to a sparse vector of a multi-dimensional constellation. Hence, SCMA improves the spectral efficiency of LDS through shaping gains of multi-dimensional constellations. Meanwhile, it still inherits the benefits of LDS in terms of overloading and moderate complexity of detection. An example of a SCMA system with 6 users and 4 resources is illustrated in Fig. 8. Each user maps its two bits to its SCMA codeword such that the superimposed transmitted block spread over the 4 resource blocks as shown in Fig. 8(a). The resultant SCMA system can be represented by a factor graph depicted in Fig. 8(b), where every circle represents a user or a variable node (VN), every block represents a resource or factor node (FN), and the edge between VN i and a FN j means that user i's data is mapped to the j-th resource block j for $i \in \{1, \ldots, 6\}$ and $j \in \{1, \ldots, 4\}$. The factor graph is regular in the sense that each VN is of degree-2 while each FN is of degree-3. The SCMA codebook design based on lattice constellations [86] was investigated in [87],

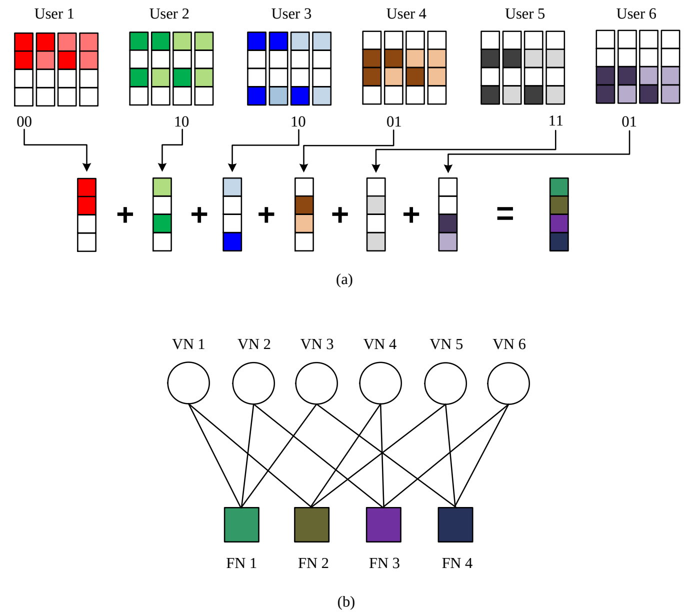


Fig. 8. Illustration of a SCMA system with 6 users and 4 resource blocks: (a) mapping between each user's two bits and SCMA codewords; (b) factor graph corresponding to the SCMA system.

[88]. Similar to LDS-CDMA, the SCMA receiver adopts MPA. To reduce the detection complexity of MPA, various detection algorithms such as discretized MPA and sphere decoding can be employed [88], [89]. Potential applications of SCMA in 6G wireless communications systems were discussed in [90]. However, the overloading performance is limited by the number of sparse codewords and the sparsity.

3) Multi-User Shared Access (MUSA): MUSA was introduced in [91] to support grant-free Internet of things [92]. In MUSA, the data symbols of each user are spread with a short-length spreading sequence. The key principle is that non-orthogonal complex spreading sequences are chosen by multiple users autonomously to enable grant-free transmissions on the same resources. It is also worth noting that the same user can choose different spreading sequences for different symbols to benefit from interference averaging. The receiver exploits the low cross-correlation properties of spreading sequences and uses SIC decoding. However, unlike LDS multiple access schemes and SCMA, the spreading sequences of MUSA are dense.

4) Successive Interference Cancellation Amenable Multiple Access (SAMA): SAMA is based on the joint design of the system signature matrix and the SIC-based MPA [93]. The symbols of different users are judiciously spread in the frequency, which can be effectively exploited by the SIC-based MPA and to obtain the diversity gain. Different from the aforementioned LDS schemes, the spreading sequences in SAMA have variable sparsity.

Let us now compare the performance of sparse CD-MA and dense CD-MA. To see how they perform, we compare their information-theoretic limits and link-level performance.

First, the spectral efficiency in the large-system limit versus the system load β in users per resources for dense CD-MA and sparse CD-MA is shown in Fig. 9, where Gaussian signaling,

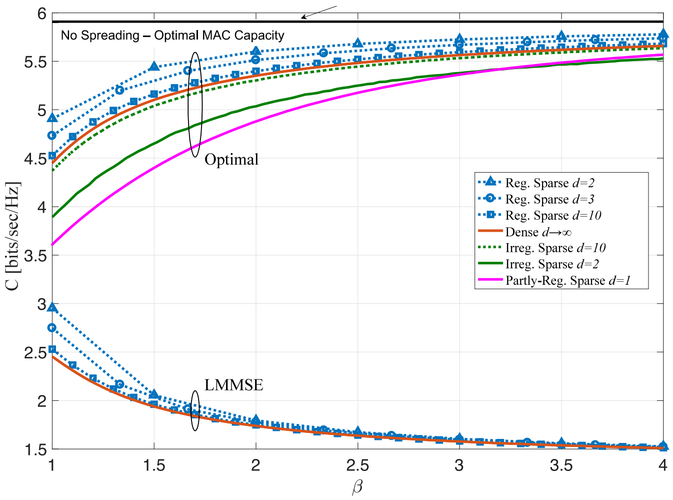


Fig. 9. Spectral efficiency of dense CD-MA and sparse CD-MA as a function of system load β for fixed $\frac{E_b}{N_0} = 10$ dB [76].

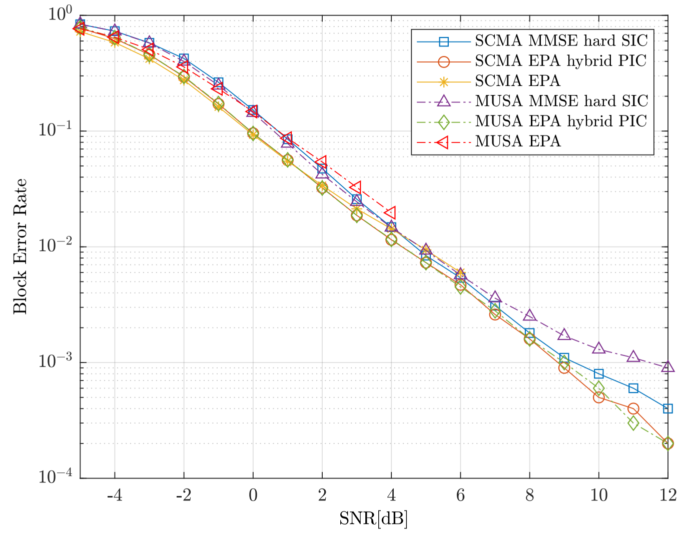


Fig. 10. Link level performance of MUSA and SCMA in the uplink with various receivers [8].

random spreading sequences, symmetric channels, and fixed power are assumed [76]. Moreover, sparse CD-MA can be further divided into *regular* sparse CD-MA and *irregular* sparse CD-MA depending on whether each user occupies a fixed number of resources or not. Observe that regular sparse CD-MA achieves larger spectral efficiency than dense CD-MA under the optimal receiver and sub-optimal linear minimum-mean square error (MMSE) receiver. Compared to regular sparse CD-MA, irregular sparse CD-MA exhibit degraded performance due to that some resources may be left unused. With the increase of sparsity parameter d, regular sparse CD-MA and irregular sparse CD-MA approach dense CD-MA from the above and below, respectively.

Next, we consider the representative schemes of dense CD-MA and sparse CD-MA, i.e., MUSA and SCMA, and compare their link-level performance. Both schemes operate in the uplink OFDM system with 10 equal SNR users, where three receivers namely MMSE with hard SIC, expectation propagation algorithm (EPA) with hard and soft parallel interference cancellation (PIC), and EPA are considered. The block error rate versus SNR is shown in Fig. 10 [8]. For detailed system parameter settings, please refer to Table 8.2-1 and Table 8.2-7 of [8]. It can be seen that SCMA slightly outperforms MUSA under the same receiver. In addition, both schemes achieve their best performance under EPA with hybrid PIC.

In light of the above, one can see that the class of sparse CD-MA schemes demonstrates better theoretical and link-level performance than dense CD-MA schemes. In addition, receiver complexity reduction can be achieved by utilizing MPAs due to the sparsity.

We summarize the main advantages and disadvantages of CD-MA schemes as follows.

- *Advantages:*
  1) **Enhanced spectrum efficiency:** CD-MA utilizes the code domain to allow multiple users to share the same resources, e.g., time and frequency, efficiently. With advanced multi-user detectors, CD-MA improves the spectrum efficiency when CSIT is perfect and the network can be either underloaded or overloaded.
  2) **Enhanced user fairness:** User fairness in CD-MA can be achieved by detecting strong users' signals first in the multi-user detection. This is because the early detected users only collect less extrinsic information compared to the late detected users [94]. In this way, more extrinsic information together with interference cancellation performed in the early detection steps can improve the detection performance of weak users.
  3) **Flexibility:** CD-MA is adaptable to various network loads and channel conditions. This is owing to the effective design of spreading code sequences in terms of correlation and sparseness for reducing inter-user interference.
  4) **Low signaling overhead:** CD-MA can allow users to choose their spreading sequences autonomously to reduce the signaling overhead and latency. In this case, the spreading sequences need to be carefully designed for the underlying multi-user detectors to mitigate collisions.
- *Disadvantages:*
  1) **High receiver complexity:** CD-MA requires complex multi-user detection techniques. In particular, to achieve the promised gain of CD-MA, iterative receiver architecture is often required. The complexity scales proportionally with the number of users and iteration numbers.
  2) **High design complexity:** The key ingredient of CD-MA is the spreading code sequence. However, finding the optimal design for the spreading sequence is still a difficult problem, particularly when the number of users is large.
  3) **High channel estimation complexity:** Most existing CD-MA schemes assume perfect CSI knowl-

edge. To accurately estimate the channel for each user, channel estimation is often performed jointly with multi-user detection, which incurs a high complexity.

4) **High sensitivity to synchronization:** Most existing CD-MA schemes assume perfect synchronization at the receiver. However, this is not always the case in practice. A delayed signal from a user could disturb the message exchange in the iterative multi-user detection, leading to performance loss.

G. Multiple Access in Other Domains

In addition to the prominent SDMA, NOMA, RSMA, and CD-MA, a range of MA schemes in other domains have been proposed in the literature. In particular, a full list of MA schemes proposed during the study phase of 5G NR were included in [8]. In what follows, we will review some typical examples from [8] as well as some newly proposed MA schemes.

1) Interleave-Division Multiple Access (IDMA): IDMA was introduced in [95] in 2003 and has gained considerable interest in recent years. IDMA can be considered as a spectral case of CDMA with all signature sequences reducing to a single chip [96]. In other words, IDMA relies on user-specific interleaving as the only means to distinguish the signals from different users [97]. Moreover, the spreading operations in CDMA can be replaced by low-rate forward error correction codes in IDMA to provide increased coding gain. It is also possible to have non-orthogonal interleavers. For a K-user IDMA system, the receiver adopts a chip-by-chip elementary signal estimation and K single-user a posteriori probability (APP) decoders, such that the total receiver complexity only increases linearly with K [98]. Advantages of IDMA over CDMA in terms of performance and complexity under practical considerations were demonstrated in [99]. Integration of IDMA with other technologies such as OFDM and massive MIMO were investigated in [100] and [101], respectively.

2) Pattern Division Multiple Access (PDMA): In PDMA [93], [102], the transmitted data symbols are mapped to a resource group that can consist of time, frequency, and spatial resources or any combination of these resources according to a pattern. Data of multiple users can be multiplexed onto the same resource group with a different pattern to realize non-orthogonal transmission. In addition, the pattern is designed with disparate diversity order and sparsity such that the BP algorithm can be efficiently employed for detection. To achieve the best possible performance, an iterative turbo receiver architecture [103] can be adopted, where the outer iteration process between the detector and decoder is performed on top of the inner iteration of the detection and decoder themselves. A gain of 500% in terms of the number of supported users under the given system packet drop rate of 1% was reported for PDMA over OFDMA in the uplink [104]. In [105], the patterns of different users are judiciously designed to exhibit appropriate diversity disparity at the symbol level and power disparity at the physical resource element level. In this way, the appropriate disparity in diversity and power can be effectively

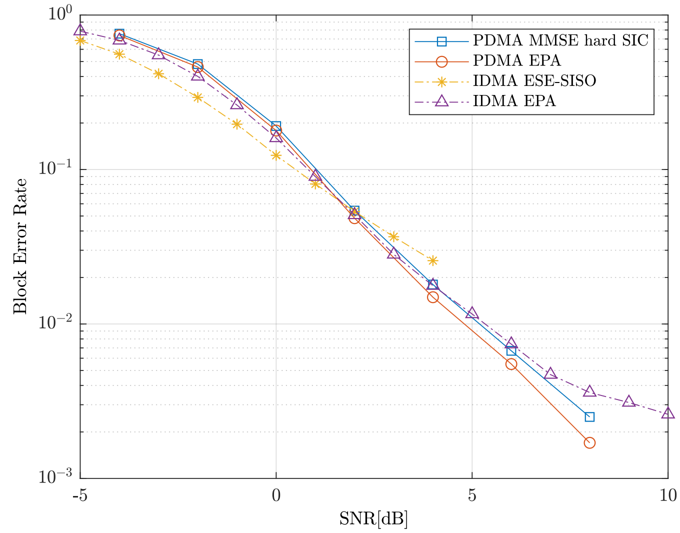


Fig. 11. link-level performance of IDMA and PDMA in the uplink with various receivers [8].

exploited by the low-complexity SIC-based BP detector with reduced error propagation during interference cancellation. The principle of PDMA has later been exploited to design pattern division random access [106]. The link-level performance of PDMA with MMSE hard SIC and EPA in the uplink OFDM system with 10 equal SNR user was demonstrated in [8]. The block error rate versus SNR is shown in Fig. 11. For comparison, the performance of IDMA with iterative elementary signal estimation and decoding (ESE-IDD) and EPA is also included in the figure. It can be seen that PDMA slightly outperforms IDMA in the high SNR regime. Note also that the performance of hard interference cancellation does not degrade much compared to soft interference cancellation.

3) Compute-Forward Multiple Access (CFMA): Compute-and-forward (C&F) is a relaying strategy introduced in [107], where the relay decodes a noisy linear combination of lattice codewords. In contrast to the classic relaying strategies such as amplify-and-forward and decode-and-forward, C&F exploits interference to obtain significantly higher rates between users in a network. Motivated by the benefits of C&F, [108] introduced CFMA based on a modified C&F technique for the Gaussian MAC. The receiver first decodes the sum of the linear combination of lattice codewords. Upon recovering the sum codewords, other users' codewords can be successfully decoded by using the sum as side information. It was proved in [108] that whole capacity region of the two-user Gaussian MAC is achievable by CFMA under lattice decoding [109], provided that the SNR of both user is larger than $1+\sqrt{2}$. It is worth mentioning that any rate points on the dominant face of the Gaussian MAC capacity region can be achieved by CFMA, without the need of time-sharing [110, Ch. 15] or rate-splitting [39]. However, random lattice codes were used in [108] as the proof techniques, which can be difficult to implement in practice. Practical implementation of CFMA based on off-the-shelf binary low-density parity-check (LDPC) codes and sum-product decoding was conducted in [111].

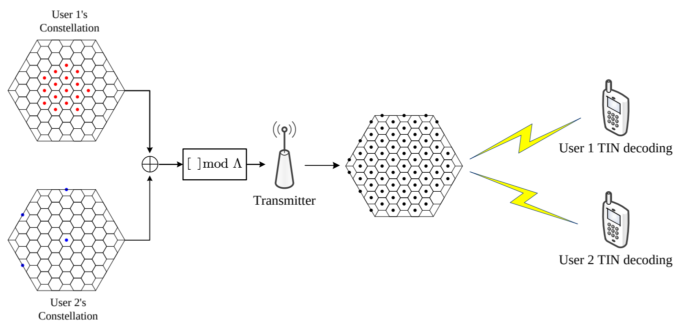


Fig. 12. An illustration of the lattice-partition-based multiple access with receivers adopting single-user TIN decoding.

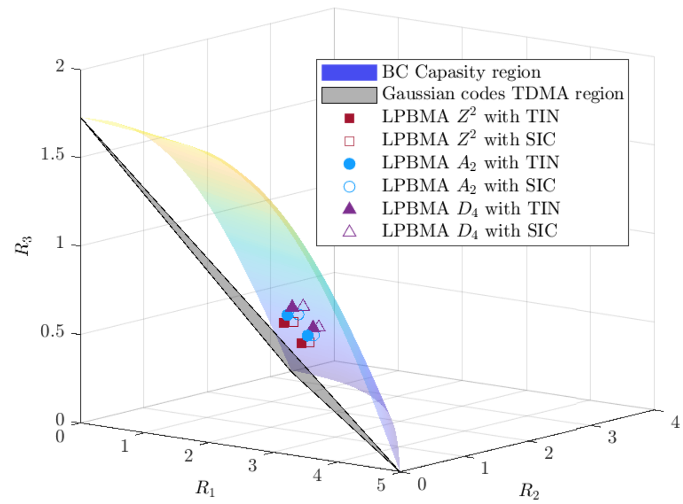


Fig. 13. The achievable rate tuples of lattice-partition-based multiple access with and without SIC by employing two-dimensional base lattices $\mathbb{Z}^2$ and A_2, and four-dimensional base lattice D_4.

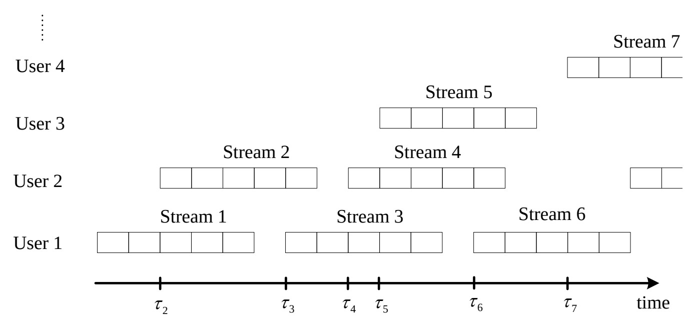


Fig. 14. Examples of spatially coupled multiple access where data streams are transmitted with different delays.

4) Lattice-Partition-Based Multiple Access (LPBMA) without SIC: In [112], A lattice-partition framework of MA without SIC, i.e., with single-user treating interference as noise (TIN) decoding, was introduced for the K-user Gaussian BC. Each user adopts a linear code [113] in conjunction with an appropriately designed constellation carved from a multi-dimensional lattice which includes the commonly adopted one- and two-dimensional constellations, i.e., pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM), as special cases. Essentially, the design ensures that the superposition of all users' signalings still preserves the lattice structure, which can be exploited to hardness inter-user interference in TIN decoding. A two-user example is illustrated in Fig. 12, where both users' signal constellations and the superimposed constellation preserve the same lattice structure. It was proved that the scheme based on discrete signaling and TIN is capable of achieving the whole capacity region within a constant gap independent of the number of users and channel parameters. Consider the three-user BC as an example. The achievable rate tuples of LBPMA with three different base lattices, TDMA with Gaussian coding, and the capacity region in bits per channel use are shown in Fig. 13. Observe that LBPMA without SIC approaches the capacity region and outperform Gaussian TDMA. Further performance gain is obtained by employing higher-dimensional base lattices. In addition, several new MA schemes evolving from lattice-partition-based multiple access were proposed for various channels, e.g., the Gaussian interference channel, the Gaussian BC with heterogeneous blocklength and error probability constraints, etc. [114]–[117], where carefully designed discrete signalings with TIN decoding are shown to be close to Gaussian signalings with perfect SIC decoding.

5) Spatially Coupled Multiple Access (SC-MA): Spatial coupling [118] is a code construction technique to boost the performance of uncoupled (weak) codes, e.g., regular LDPC codes and turbo codes, all the way to capacity-achieving [119]–[122]. Motivated by this, [123], [124] introduced a new signaling formats for the Gaussian MAC, where the modulated data streams are repeated, permuted, and transmitted with regular time offsets (delays). Since the relations between the data bits and modulation symbols transmitted over the channel can be represented by a sparse graph, the receiver observes spatial coupling of the individual graphs which enables efficient demodulation/decoding. It was proved in [123], [124] that coupling data transmission with a two-stage demodulation/decoding in which iterative demodulation based on symbol detection and interference cancellation followed by parallel decoding achieves the Gaussian MAC capacity asymptotically such that the gap to capacity vanishes as the system's SNR increases. Note that the feedback between the decoder and the demodulator is not required. In contrast to most NOMA receivers that have to demodulate/decode the whole symbol block, sliding window demodulation/decoding can be adopted thanks to the spatially coupled graph structure [125]. As a result, the demodulation/decoding latency is limited, which can be suitable for streaming services. In [126], the spatially coupled transmission scheme has been extended to the case of transmitted blocks with irregular delays. An example for the transmission scheme is illustrated in Fig. 14, where each user divides its transmission frame into sub-blocks and delays the transmission by τ. However, the spatially coupled transmission scheme can be sensitive to synchronization errors and fractional delay.

6) Layered-Division Multiplexing (LDM): LDM is a non-orthogonal multiplexing technology adopted in the Advanced Television Systems Committee standard (ATSC 3.0) [127], which has never been implemented in previous broadcast and broadband systems [128]. In an LDM system, a layered transmission structure is used to simultaneously transmit multiple signals with different power levels and technologies (channel coding, interleaving, modulation, multiple-antenna, etc.) for different services [129]. At the receiver, the decoding of signals from multiple layers requires SIC. The principle of LDM is similar to NOMA. Hence, its gain over the conventional OMA schemes has been well understood. Note that LDM is implemented on top of OFDM such that the superposition of signals take place before OFDM modulations.

7) Index Modulation Multiple Access (IMMA): Index modulation is a technique of using the indices of resources to carry extra information bits [130]. It can achieve higher spectral and energy efficiency than conventional modulation schemes. Motivated by this, IMMA was proposed to further improve the spectral and energy efficiency of the traditional NOMA scheme [131]. In IMMA, each user transmits additional information bits (index bits) by partial resources, where the index patterns can be the activation status of time slots, subcarriers, transmit or receive antennas, spreading codes, power levels, etc [132]. However, the detectors of these IMMA schemes are joint maximum-likelihood detectors which can be challenging to implement.

8) Delay-Doppler Domain Multiple Access (DDMA): Orthogonal time-frequency space (OTFS) modulation [133]–[135] was introduced for high-mobility wireless applications, which multiplexes data in the delay-Doppler (DD) domain rather than the time-frequency domain as in conventional multi-carrier modulations. OTFS has also stimulated new research on delay-Doppler plane modulations. For instance, the newly proposed orthogonal delay-Doppler division multiplexing (ODDM) [136] was proved to achieve orthogonality on the DD plane's fine resolutions with realizable pulses. Owing to the orthogonality, ODDM itself can be potentially employed as an OMA scheme on the DD domain. Another promising research area is that other MA schemes such as SDMA and RSMA can be combined with OTFS/ODDM, which can offer high spectral efficiency with robust performance in high-mobility channels.

9) Other Multiple Access: All the above discussion on MA techniques is not exhaustive with new MA strategies disclosed on a regular basis such as fluid antenna multiple access (FAMA) [137], location division multiple access (LDMA) for near-field communication [138], or orbital angular momentum (OAM) multiplexing and multiple access. Additionally, tailored MA schemes have been introduced to support specific technologies such as model division multiple access (MDMA) for semantic communications [139] and AirComp MA scheme for federated learning [140].

H. Random Access

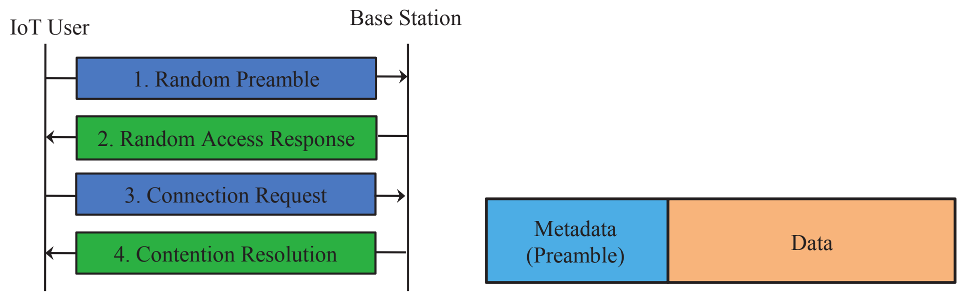
One last type of MA is random access, specifically tailored for Internet of Things (IoT) and massive connectivity. For IoT communication network, the key is how to coordinate the transmission of small packets between the base station and a huge number of IoT devices in a reliable and low-latency manner. This is in sharp contrast to the human-type communication, which targets at high-speed transmission to/from a medium number of users. Such a difference poses new challenges for the MA design under IoT applications. Specifically, if the number of users is huge, then access collision is more likely to occur, which leads to long delay. In IoT network, it is of paramount importance to design innovative MA schemes that can greatly reduce the access delay so as to facilitate the small-packet transmission from a massive number of IoT devices. Generally speaking, in the literature, there are two solutions for MA in IoT network: the grant-based MA technique and the grant-free MA technique, as shown in Fig. 15.

Grant-based MA technique are conventional for eMBB-like services and relies on the base station to grant users the access of the radio spectrum for data transmission.

In grant-free MA, or also called grant-free random access (RA), the devices can access the channel without any prior resource requests [92], [141]. In addition, the set of active users is unknown to the receiver. Grant-free RA is suitable for supporting a massive number of devices with sporadic traffic, e.g., mMTC as in IoT, and can be divided into two different paradigms, namely sourced and unsourced RA [142], [143]. For sourced RA, the receiver is interested in both messages and user identities. In the literature, various techniques, such as the compressed sensing approach [144] and the covariance approach [145], have been proposed to detect the active devices and/or estimate their channels. For unsourced RA, the receiver is interested in the transmitted messages only (user identity is recovered at the higher layers of the communication protocol rather than the physical layer).

In the sequel, we discuss grant-based MA, grant-free sourced random access (based on compressed sensing and AI), and grant-free unsourced random access.

1) Grant-based Random Access: First, with human-type communication, the contention-based grant-based access schemes, e.g., ALOHA, are widely used. Under this scheme, there is a preamble pool consisting of lots of orthogonal preambles. When a user becomes active, it will randomly select a preamble from the poor. If an active users selects a unique preamble that is not selected by other active users, the base station will grant this user the access for data transmission. Otherwise, if some active users select the same preambles, the base station will detect their access collision and not grant them the access for data transmission. Despite of their simplicity, such schemes do work quite well in current cellular network, where the number of users is moderate. Therefore, it is not surprised that some efforts have been paid in the literature to investigate how to modify the existing grant-based access schemes according to the requirements of IoT applications. Here, the main issue to directly apply grant-based schemes in IoT network lies in delay - when there is a huge number of IoT devices to compete for the transmission grant from the base station, the possibility for collision is very high. As a result, in the literature, some works have been done to propose efficient methods for resolving the collisions. For example, in [146], a strongest-user collision resolution scheme was proposed for grant-free random access in IoT network. The idea is that after a collision occurs, each colliding user will make a decision about whether it is with the strongest channel



(a) Grant-based Random Access

(b) Grant-Free Random Access

Fig. 15. Grant-based versus grant-free random access schemes.

among all the colliding users. If a colliding user believes that its channel is the strongest, it will send the access request to the base station. Otherwise, it will be quiet for some time and re-compete for the access. It was shown in [146] that under the massive MIMO regime, the strongest-user collision resolution scheme can resolve most of the access collisions in massive connectivity.

2) Grant-Free Sourced Random Access based on Compressed Sensing: Second, one may ask whether it is possible to reduce the collision probability, rather than resolving the collision, under IoT networks. The goal can be achieved via the grant-free random access scheme [141]. Under the grant-based random access scheme, collision arises from the case when multiple users select the same preamble for transmission. To mitigate collision, a natural idea is that each user is allocated with a fixed preamble, while the preambles allocated to different users are different. In this case, the preamble of each user serves as the identification of this user. In other words, after receiving the preambles sent from the active IoT devices, the base station can detect which preambles are received, and then know which users are active. This scheme is called grant-free scheme because the base station just detects which users are active, instead of sending grants to permit some users to be active. However, detecting which preambles are received is quite challenging in massive IoT connectivity, because it is not possible to assign orthogonal preambles to a huge number of users. Recently, a breakthrough was made to tackle the above challenge arising from grant-based random access based on the compressed sensing technique. Specifically, although there is a huge number of IoT devices, most of them are in the sleeping model to save the energy at each time slot. Thanks to this activity sparsity, it was shown in [144] that detecting the active devices, i.e., deciding which preambles are received, can be cast into a compressed sensing problem. More importantly, based on the state evolution, [144] rigorously proved that in the asymptotic massive MIMO regime where the number of antennas at the base station goes to infinity, the activity detection error probability can go down to zero when the approximate message passing algorithm is used. This result theoretically justifies the effectiveness of the compressed sensing based grant-free random access scheme for IoT networks.

3) Grant-Free Sourced Random Access based on Covariance Approach: Under the compressed sensing based grant-free sourced random access scheme, the base station can jointly detect the active devices and estimate their instantaneous channels. In many IoT applications, we may be only interested in device activity detection, because channel estimation is not that important to short packet transmission, but will induce long estimation delay. Recently, [145] pointed out that the covariance matrix of the base station received signals is a sufficient statistic for device activity detection, and proposed a covariance approach that exploits the sample covariance matrix of the base station received signals to detect the active devices without estimating their channels. The covariance approach is suitable for massive MIMO systems, because in such systems, the sample covariance matrix of the base station received signals is a good estimation of the covariance matrix. Under the covariance approach, the fundamental scaling law among the numbers of the users, active users, base station antennas, and the length of the user preambles were rigorously characterized in [145] for the case when the large-scale fading coefficients of the users are known, termed as restricted version of the maximum likelihood estimation, and in [147] for the case when the large-scale fading coefficients of the users are not known, termed as unrestricted version of the maximum likelihood estimation. It is not surprising that more active devices can be detected under the covariance approach compared to the compressed sensing approach, because channel estimation is not performed.

To summarize, under both the compressed sensing based scheme and the covariance based scheme, there is a theoretical gain of the grant-free random access scheme over the grant-based random access scheme in IoT networks, because the former scheme can mitigate almost all the collisions in the massive MIMO region, but the latter scheme can merely tackle some of the collisions. But the theoretical gain is achieved with complicated device activity detection algorithms, which need to deal with high-dimension matrices in massive connectivity systems.

4) Grant-Free Unsourced Random Access: Unsourced random access (URA) is a novel communication paradigm introduced in [148]. URA differs from the conventional random access models in several ways and is well-suited to meet the demands of massive connectivity in mMTC. In this new paradigm, a wireless network has K_{tot} users, out of which only a small number $K_{\text{a}} \ll K_{\text{tot}}$ users are active and sending finite blocklength packets at any given time. The total number

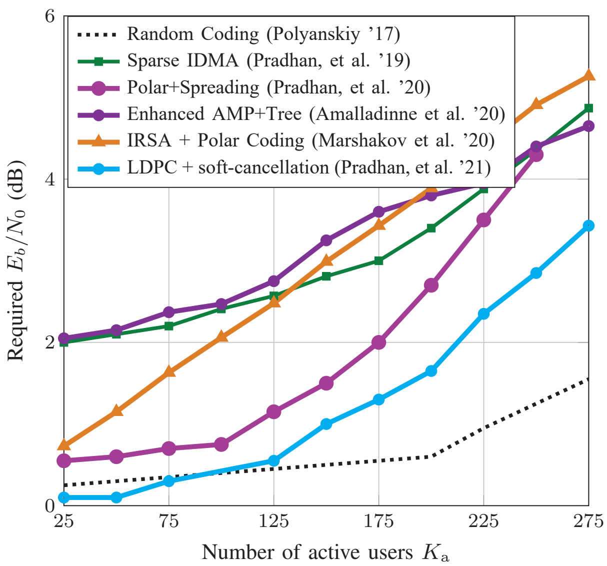


Fig. 16. Comparison between different achievable schemes for URA [152].

of users K_{tot} can be very large (potentially infinite) and grows linearly with the blocklength. Moreover, all users share a common codebook, which is motivated by the practical scenario where millions of low-cost IoT devices have their codebook hardwired at the moment of production. In this case, the receiver is interested in the transmitted messages only, whereas the user identity is recovered at the higher layers of the communication protocol rather than the physical layer. However, an active device can embed its identity as a part of the payload. The decoder's job is to produce a permutation of the transmitted messages. As opposed to the global error probability in the conventional RA, only per-user error probability is of interest in URA.

Since the message lengths for typical IoT applications are small, the fundamental limits of the URA channel should be reinvestigated rather than relying on the classic information-theoretic results based on infinite blocklength and a fixed number of users. The achievability and converse bounds for URA with single antenna transceiver and without fading were first given in [148] as the minimum required energy-per-bit $\frac{E_{\text{b}}}{N_0}$ for given K_{a}, blocklength, and the per-user error probability requirement. The results were later generalized to URA with quasi-static fading [149] and MIMO with quasi-static fading fading [150]. Interestingly, all these results show that near-perfect multi-user interference cancellation can be achieved for user densities below a critical threshold. Very recently, it was shown that the achievability of binary coding is very close to that of Gaussian coding [151].

Ever since the introduction of URA, there has been significant effort in designing achievable schemes to approach the performance bound while maintaining low computational complexity. The first line of works adopt coded compressed sensing (CS) based approaches [153], [154], which split the payloads into smaller pieces, perform CS recovery on the sub-problems, and subsequently stitch pieces together using an outer code. Another line of works focuses on sparsifying collisions to reduce multiuser interference, such as schemes in the spirit of T-fold ALOHA [155], [156] and schemes leveraging spreading sequences [152], [157]. Notably, the best achievable scheme based on LDPC codes with random spreading and soft interference cancellation [152] even slightly outperforms the random coding achievability bound in [148] as shown in Fig. 16. The detailed list of URA schemes can be found in [158, Ch. 2]. The coded compressing sensing approaches were later extended to massive MIMO URA. It was proved that as the number of receive antennas is sufficiently large, one can detect $K_{\text{a}} = O(n^2)$ active users with n being the channel coherent blocklength [159] and the required transmit $\frac{E_{\text{b}}}{N_0}$ for reliable communication can be made arbitrarily small [160]. Recently, [161] demonstrated that employing spreading sequence for interference mitigation together with data-aided channel estimation gives the best performance so far.

Current URA schemes focus sorely on uplink transmission. Incorporating downlink feedback, e.g., by informing decoding failures to users, can further improve system performance. However, without user identities, it is challenging for the base station to provide feedback to unsourced active users. Hence, how to integrate feedback into URA schemes would be a worthwhile research direction.

I. Technology Outlook and Future Works

We draw several observations from the past subsectionvs on common challenges in MA designs and on MA dimensions before giving an outlook of MA technology development toward UMA.

1) Common Challenges in MA Designs and Performance Evaluations: Modern and emerging MA schemes share a number of common challenges, namely their optimization complexity (for power allocation, precoders, and resource allocation), channel estimation complexity, receiver complexity (involving highly complex operations at the receivers such as SIC), and sensitivity to various impairments such as RF impairments and synchronization. The latter challenge is particularly common in IoT where most of the works assume that all the IoT devices are perfectly synchronized, such that when they become active, they transmit their preambles at the same time. In practice, it is impossible to perfectly synchronize a huge number of low-cost IoT devices. In this case, how to make the strongest-user collision resolution scheme work under the grant-based random access scheme? How to formulate the device activity detection and synchronization error estimation problem as a compressed sensing problem under the grant-based random access scheme? Whether the sample covariance matrix of the received sequence is still a sufficient statistic for device activity detection? All these questions should be carefully addressed before the MA schemes can work in IoT networks. How to overcome those challenges will be discussed in more details in Section III.

Additionally there is a critical lack of comprehensive performance evaluations and comparison among MA schemes. The literature on NOMA has been primarily built upon NOMA vs OMA comparisons, without consideration for RSMA and other schemes. The RSMA literature has thrived to compare with OMA, NOMA and SDMA, but not with code-domain schemes. Evaluations of CD-MAs compared performance among various types of codes. What is missing is a consistent performance evaluation methodology that enables a comprehensive performance comparison of MA schemes across all dimensions. This is crucial to obtain a holistic understanding of the performance benefits of MA schemes and the role played by the various MA dimensions for both uplink and downlink.

2) MA Dimensions: Previous discussions highlight that there are numerous dimensions exploitable by a MA scheme, starting from the usual time, frequency, space, power, code, but new dimensions have also appeared in recent years including message, index, pattern, interleaver, etc. We have decided to classify them all into five categories: time, frequency, power, space, signal. Signal is a broad category that encompasses multiple sub-dimensions such as message split and combiner, channel coding, modulation, interleaver, spreading sequence, etc. Table II displays a non-exhaustive list of dimensions exploited by the aforementioned MA schemes and how aforementioned MA schemes exploit those various dimensions.

Table II also highlights that MA techniques exploit one or multiple of those dimensions. For instance, SDMA and NOMA exploit the space domain and the power domain, respectively, while RSMA by opening the door to the message dimension exploits the message combiner, message split, power, and space domains to generalize and unify SDMA, NOMA, and physical layer multicasting. PDMA exploits space, time and frequency domains and CD-MA schemes exploit the spreading code domain together with time and frequency. All five domains would be exploited by OFDM(A)-RSMA for instance [162], [163].

3) RSMA - A First Step Toward Unification in 6G: The beauty of RSMA is that it *unifies* into a single MA scheme the seemingly unrelated strategies of NOMA, SDMA, and physical layer multicasting. This unification capability translates in RSMA to be a superset of those three strategies and can boil down to any of them by turning off some of the streams. This was illustrated in Fig. 3 for a two-user scenario and has been discussed extensively in the NOMA and RSMA literature [10], [11], [15] and demonstrated by the well known message-to-stream mapping of [10], [12]. Consequently, RSMA can softly bridge those three strategies, explore operating points that are not achievable by any of them, and outperforms them all. This gives RSMA an edge over other MA schemes as demonstrated in over 40 applications in 6G [10], [15], including in multi-functional networks such as ISAC [164].

The capability of RSMA to *unify* MA schemes is crucial for the long term research in the theory and practice of communications and wireless systems. Recall indeed the wise words in the acknowledgments section of the book [23] by David Tse and Pramod Viswanath: "Bob Gallager's research and teaching style have greatly inspired our writing of this book. He has taught us that good theory, by providing a unified and conceptually simple understanding of a morass of results, should shrink rather than grow the knowledge tree.". In our context, as illustrated in Fig. 17 and by the RSMA branch in Fig. 1, *RSMA, by providing a unified and conceptually simple understanding of a morass of results on SDMA, NOMA, physical layer multicasting, shrinks rather grows the knowledge tree of MA schemes based on space, power, signal dimensions.* The capability of RSMA to unify and therefore be more universal than other MA schemes makes practical implementation and operation easier. Indeed a single unified and general MA scheme would be easier to implement and optimize than a combination of multiple MA schemes, each optimized for specific conditions. This is increasingly important in multi-functional 6G and beyond networks given the wide diversity of services, use cases, and deployments.

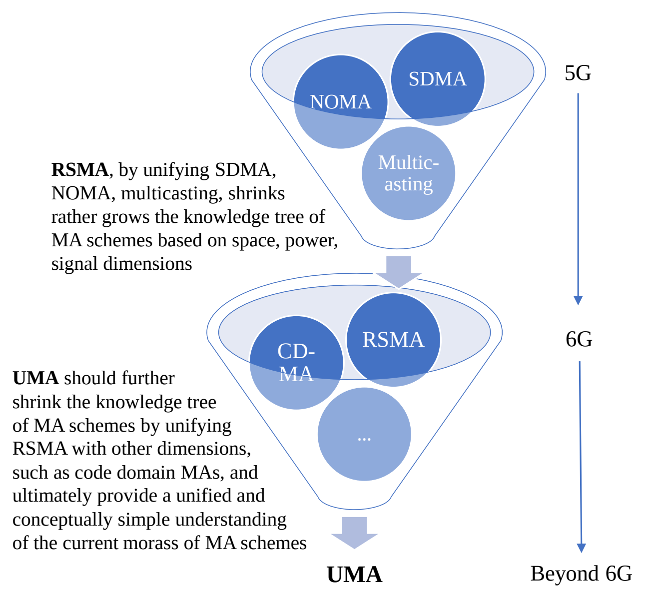


Fig. 17. From RSMA to UMA: How to further shrink the knowledge tree.

4) Toward Universal Multiple Access in Beyond 6G: Though RSMA provides a good example of a unified theory of MA schemes, RSMA does not unify all MA schemes and therefore does not exploit all dimensions of time, frequency, power, space (e.g. antennas, beams), signal (e.g. messages, codes, etc). This would then bring the central question for future research on MA in beyond 6G: "*What is Universal Multiple Access (UMA)?*". Following the same wise philosophy of Gallager, Tse, and Viswanath, *UMA should further shrink the knowledge tree of MA schemes by unifying RSMA with all other dimensions, such as code domain MAs, and ultimately provide a unified and conceptually simple understanding of the current and future morass of MA schemes.* This question, vision, and research philosophy is illustrated in Fig. 17 and Fig. 1. Following the above question, come the natural and important questions (and related challenges) "*How to design UMA?*" and "*Why and when do we need UMA?*"[2].

There are no answers to those questions yet since UMA does not exist. Those questions nevertheless aim to trigger discussions, give research directions, and motivate further research. Addressing them is very timely for 6G and beyond for two main reasons. *First*, the literature and the number of

[2]Those questions include the design of the transmitter and receiver architectures, resource allocation, optimization, channel estimation, performance analysis, sensitivity to impairments, performance evaluations, etc.

TABLE II
MA DIMENSION CLASSIFICATION AND MAPPING OF MA SCHEMES TO MA DIMENSIONS. RSMA BASIC SCHEMES EXPLOIT SPACE, POWER, AND MESSAGE SPLITTING DIMENSIONS, THOUGH SPACE-TIME AND SPACE-FREQUENCY RSMA SCHEMES EXIST [10], HENCE EXPLOITING A FOURTH DIMENSION. IMMA HAS PRIMARILY TWO DIFFERENT SCHEMES, AND THE CURRENT DESIGNS UTILIZE THREE DIMENSIONS AT THE SAME TIME TO DISTINGUISH USERS.

Dimension / MA	Time	Frequency	Space	Power	Signal					
					Message combiner	*Message split*	*Channel coding*	*Modulation*	*Interleaver*	*Spreading sequence*
TDMA	✓									
FDMA		✓								
OFDMA	✓	✓								
SDMA			✓							
NOMA				✓						
PHY Multicasting					✓					
RSMA	✓	✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓				
CD-MA	✓	✓								✓
IDMA							✓		✓	
PDMA	✓	✓	✓							
CFMA				✓			✓	✓		
LPBMA				✓			✓	✓		
SC-MA									✓	✓
LDM	✓	✓		✓						
IMMA	✓✓	✓	✓					✓✓		
DDMA	✓	✓						✓		

MA schemes have exploded in the past decade and grown non-organically, fueled by the sheer interest in making new or better use of MA dimensions and the limited resources but also by the large number of new wireless services offered by multi-functional and intelligent 6G. Many of those schemes claim to exploit new dimensions and presenting new ideas, though it remains to be seen whether this is true or whether those are a rebranding or twist of known concepts. *Second*, the multi-functionality of wireless networks calls for a more efficient, flexible and robust use of resources, better management of interference, better handling of heterogeneity of wireless services, unified and simplified hardware and software architectures. UMA should shrink the knowledge tree by truly understanding the essence of a unified MA design at exploiting in the most simple way all five dimensions (time, frequency, space, power, signal) to efficiently provide intelligence and multi-functionality (communications, sensing, localization, computation, energy transfer and harvesting) in 6G and beyond network.

Many MA schemes exploiting the code domain have been classified as CD-MA with LDS and SCMA being instance of sparse CD-MA and MUSA and SAMA of dense CD-MA. However, CD-MA here only stands for a collection MA schemes exploiting various properties of codes but CD-MA is not a MA scheme itself that unifies LDS, SCMA, MUSA, SAMA, i.e., there is no MA scheme that unifies LDS, SCMA, MUSA, SAMA and enables to softly bridge them all. Hence CD-MA does not have the same unification capability as RSMA. Nevertheless, it is important to observe that RSMA and CD-MA span different MA dimensions, with e.g., RSMA not exploiting the spreading code dimension. Shrinking the knowledge tree of MA schemes to identify UMA would require a better understanding of how all CD-MA schemes are linked to each other and whether they form particular instances of a more general class of MA schemes, and a better understanding of the interplay between RSMA and CD-MA. Unfortunately such research avenues remain largely unexplored. Indeed, a number of works have attempted to combine CD-MA with other MA schemes. The combination of NOMA and SCMA was investigated in [165]–[167]. The motivation is to allow the same SCMA codeword to be used by multiple users simultaneously, thereby increasing the number of supported users. At the receiver, MPA combined with SIC decoder is employed. However, the complexity would be higher than that of a single SCMA or a NOMA scheme. In [168], CD-MA is combined with SDMA in the massive MIMO setting, where the cases of CD-MA offering spectral efficiency improvement were identified. The integration of CD-MA into RSMA was suggested in [10]. Since the split in common and private streams in RSMA can be seen as virtual users, each virtual user can be assigned a dedicated spreading sequence for enabling code-domain multiplexing. However, such an idea has not been explored further. One major arising challenge of integrating RSMA and CD-MA, hence UMA, will lie in the receiver complexity.

III. Artificial Intelligence for Multiple Access

In this section, we demonstrate how AI techniques can be used to address some of the MA challenges highlighted in the previous section. We start by providing an overview of AI methods and delve into how they can be applied for resource allocation and optimization for different MAs. We then discuss AI-empowered channel estimation, receiver design, and user behavior predictions, and finish the section by providing some outlook of AI for UMA.

A. AI-empowered MA Resource Allocation and Optimization

In every wireless communication system, there is a constant challenge: harnessing limited resources to achieve improved performance while satisfying various QoS requirements such as rate, latency, and reliability. Therefore, the allocation of wireless resources is of critical importance in the pursuit of optimized network performance and the fulfillment of the multifaceted needs inherent to wireless communication systems. Different MA schemes introduce distinct resources for allocation. For instance, OFDMA introduces the challenge of subcarrier allocation among users, TDMA involves time slot allocation, NOMA introduces user pairing, grouping, decoding order, and power allocation, SDMA incorporates beamforming, power allocation, and user scheduling, RSMA centers on beamforming, power allocation, common and private rate allocation design, and SCMA deals with sparse code allocation among users. Essentially, the choice of the optimal resource allocation approach relies on various factors, including the specific MA scheme, available resources, system requirements, and the intricate interplay of these key factors.

Conventional resource allocation algorithms typically rely on optimization theory and game theory. While these methods can yield mathematically optimal, sub-optimal, or Nash equilibrium solutions, they often come with a high computational cost and burden, especially in large-scale (many devices, antennas) systems. The rapid advancement of AI in the past decades has opened up a new approach to tackling the complex, non-convex problems often encountered in resource allocation. This approach allows the resource allocation scheme to be learned directly from data samples or environment, eliminating the need for complex mathematical models. Fig. 18 shows a non-exhaustive search of the potential AI methods that can be applied for resource allocation. Broadly, AI-based resource allocation methods fall into three categories: traditional ML, deep learning (DL), and reinforcement learning (RL). In the following, we provide a concise overview of these three methodologies and how those AI-based approaches can be applied to diverse resource allocation scenarios, followed by a summary of the advantages and disadvantages for using AI-based approaches.

1) Machine learning: The term "machine learning (ML)" originally referred to the development of algorithms that enabled machines to learn and solve specific problems. With the growing development of neural networks, ML now primarily refers to traditional learning approaches that do not rely on neural network. Common ML methods include support vector machine (SVM), K nearest neighbors (KNN), and K-means clustering and principal component analysis (PCA).

SVM is often used for resource allocation scenarios that involve binary data classification. For instance, it can be applied to spectrum sensing in dynamic spectrum access (DSA) systems. In this application, SVM is used to classify segments of the spectrum as either occupied or available for secondary users based on received signal characteristics, enabling efficient spectrum utilization [169]. *KNN* works by finding the K closest data points in the training data to a test data point, then predicting or classifying based on the majority label or aggregated values of those nearest neighbors. It's non-parametric, so it doesn't make any assumptions about the data distribution. In applications like user grouping, KNN helps clustering or categorizing users based on their similarity in terms of channel conditions or interference patterns [170]. SVM and KNN are both supervised learning algorithms, and they share common challenges when dealing with large datasets due to their high computational complexity. For KNN, it suffers the "curse of dimensionality" wherein data points tend to become equidistant from each other in high-dimensional spaces, posing difficulties in identifying meaningful nearest neighbors. In contrast, both *K-means clustering and PCA* are unsupervised learning algorithms. K-means clustering can be applied to wireless resource allocation problems such as user grouping, spectrum allocation. For example, a K-means based user clustering algorithm is developed in [171] for NOMA by exploiting the channel correlation among users. PCA is a dimensionality reduction technique and itself is not a direct resource allocation algorithm for wireless networks. However, it plays a crucial role as a preprocessing and analysis tool that enhances the quality of data used by resource allocation algorithms [172].

All the aforementioned non-neural network-based ML algorithms offer straightforward interpretations of results. However, their effectiveness is pronounced when dealing with data characterized by straightforward relationships. They may encounter difficulties in capturing intricate hierarchical representations within complex data relationships. In such case, DL or neural works demonstrate superior performance.

2) Deep Learning: The DL approach, powered by neural networks, is one of the most renowned techniques in the field of AI for its success in addressing various challenges, such as image recognition and natural language processing. Within the realm of wireless resource allocation, DL has shown to be valuable thanks to its capacity to efficiently process vast and intricate datasets. It facilitates intelligent and adaptive resource allocation, continually optimizing network performance in real-time. Popular neural networks mainly consist of dense neural networks (DNNs), convolution neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs). Each of these architectures has demonstrated its prowess in addressing specific resource allocation challenges for different MA schemes.

DNN: DNN is the most fundamental neural network architecture, often referred to as multi-layer perception (MLP). Within this architecture, neurons are associated with input weights and incorporate an activation function to produce outputs. The computation unfolds layer by layer, with parameter updates facilitated through the process of backpropagation. It's worth noting that this structural framework finds application in various other types of neural networks as well. DNNs have been employed to approximate the weighted minimum mean square error (WMMSE) algorithm, a versatile non-convex optimization technique for resource allocation and optimization of MA schemes, in an end-to-end manner [173]. This pioneering approach signifies a paradigm shift in optimization, known as "learn to optimize", since a three-layer DNN can closely approximate WMMSE while achieving significant computational time reductions, often spanning orders of magnitude. Considering the widespread adoption of WMMSE in wireless resource allocation tasks such as beamforming and power allocation for different MA schemes, this achievement implies that DNNs can effectively address a broad range of resource allocation challenges [174].

CNN: In contrast to DNNs, CNNs leverage convolutional

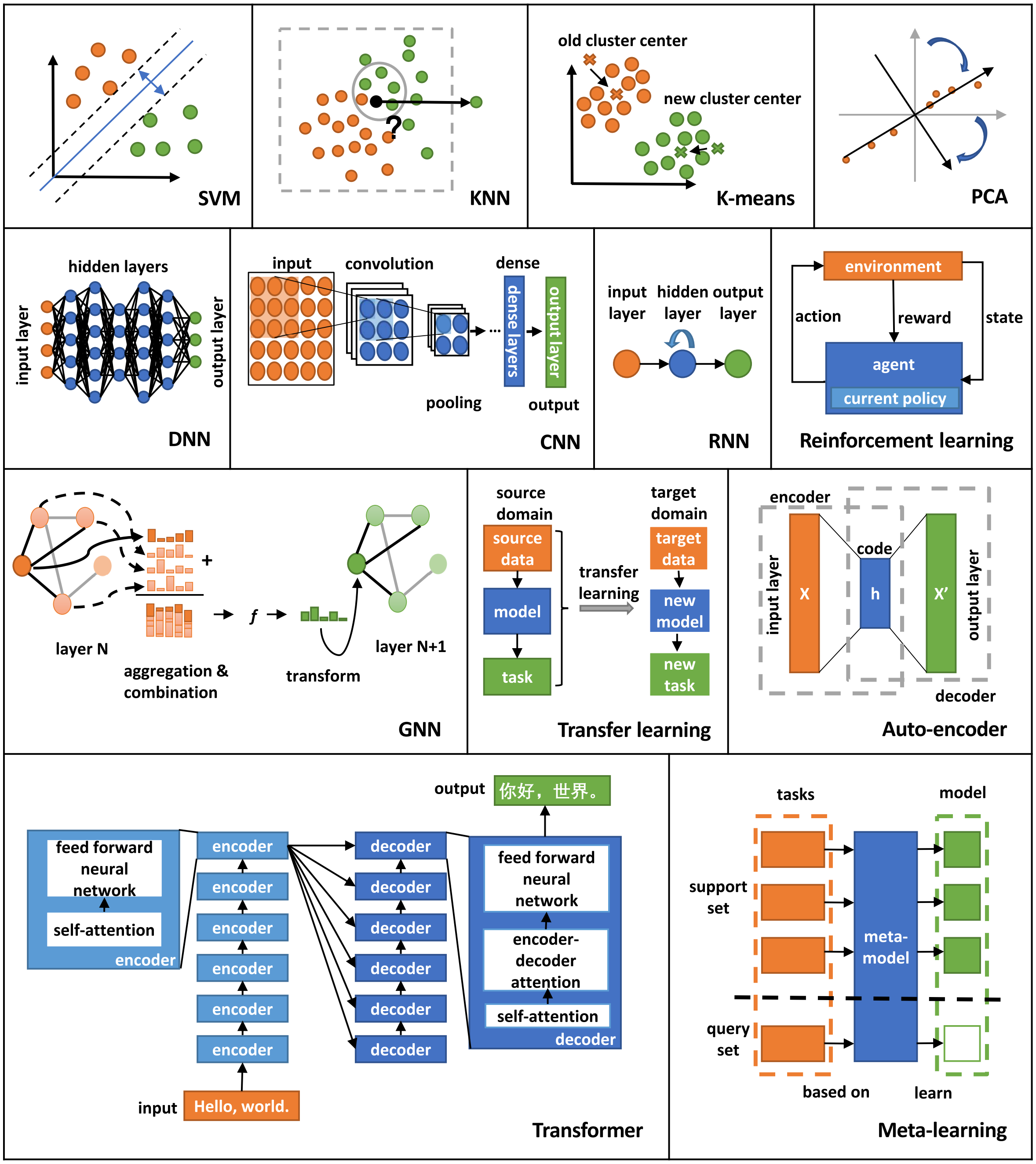


Fig. 18. A non-exhaustive search of AI methods that can be applied for resource allocation.

and pooling layers for specialized operations. By stacking these convolutional and pooling layers alongside dense layers, CNNs significantly reduce the number of parameters required to achieve equivalent performance to that of DNNs. CNNs excel in feature extraction and are notably easier to train compared to DNNs. These advantages have led to the widespread adoption of CNNs in addressing resource allocation challenges. CNN has been trained to approximate WMMSE for power allocation in [175], and for SDMA beamforming design in [176]. In [177], the power control problem of NOMA is also addressed by CNN.

RNN: RNNs take into account the impact of inputs from the previous time steps. This distinctive feature allows RNNs to maintain memory of prior inputs through their unique network architecture. Therefore, RNNs are well-suited for handling time-dependent data. This characteristic of RNNs make them highly valuable to capture the temporal correlations of data. In [178], RNN is used by each base station to predict the spectrum allocation. Additionally, in [179], a graph attention RNN is proposed for traffic prediction. The RNN family includes various common variants, including long short-term memory (LSTM) and gated recurrent unit (GRU), among others. These variants have been found highly effective in tackling various resource allocation challenges. LSTM is used to forecast the user mobility, and enhance the performance of resource allocation [180]. In terms of beamforming design for SDMA, LSTM emerges as a powerful tool for integrating information from multiple preceding steps in order to refine the current step, as evidenced in [181]. This contrasts with conventional alternative optimization algorithms, which typically rely solely on information from the present and the immediately preceding step to influence their output. For GRU, it can serve as a robust predictive model capable of consistently delivering high levels of accuracy in forecasting resource requests [182].

GNN: GNNs are specifically built upon graph structures, offering the capability to handle non-Euclidean data. GNNs exhibit remarkable flexibility in accommodating input data of varying dimensionality, making them a natural fit for the dynamic nature of the wireless communication domain, where the number of users and antennas can vary significantly. By constructing directed/undirected graphs based on the structure of the considered wireless networks with devices as nodes and channels as edges, GNNs are capable to learn the wireless network and enhance resource allocation [183], [184]. As an alternative, GNNs can construct graphical models based on the mathematical formulation of specific resource optimization problems by defining node sets and incorporating parameters as features, as evidenced in [185], [186]. By such means, GNNs are promising for addressing large-scale MA resource allocation problems while enjoying a high computational efficiency.

Other Methods: In addition to the previously mentioned standard neural network architecture, numerous advanced DL approaches have gained widespread adoption in addressing resource allocation problems. These encompass a range of techniques, such as the *transformer* [187], *autoencoder* [188], *transfer learning* [189], *meta-learning* [181]. The transformer takes the advantage of the attention mechanism, which sets it apart from RNNs. It is suitable for parallel computing, leading to substantial performance enhancements when compared to RNNs. An autoencoder, a neural network for unsupervised learning, contains two parts: an encoder that compresses input data into a simpler form, and a decoder that recreates the original input from this simpler representation. In wireless resource allocation, autoencoders are used to learn efficient data representations that capture important patterns in communication data. These learned representations improve the efficiency of resource allocation methods like power control, channel assignment, or user scheduling by giving decision-making algorithms a more concise and informative input. The fundamental concept behind transfer learning involves extracting essential features from the source domain and fine-tuning the pre-trained model for application in the target domain. By leveraging its capability to transfer valuable prior knowledge to new scenarios, transfer learning offers a promising solution to address the challenge of task mismatch encountered in practical wireless communication systems. The key to meta-learning is the ability to "learn-to-learn" so that the trained model can improve its learning ability and adapt to new tasks or domains with minimal or no human intervention. This

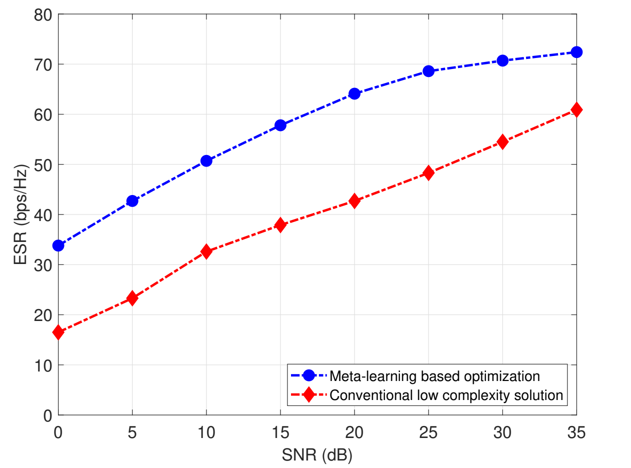


Fig. 19. Ergodic sum-rate of RSMA with meta-learning based optimization and conventional precoder designs in a massive MIMO setting with 100 antennas serving 12 users [190].

approach facilitates rapid adaptation to new environments, thus enabling more effective resource allocation decisions. In this context, meta-learning was found to be a powerful computationally efficient alternative to WMMSE-based optimization techniques to optimize precoders and resources allocation in RSMA [190]. The performance obtained with meta-learning is comparable to convex optimization-based approaches but achieves a significantly lower running time. This is particularly helpful for large scale RSMA network settings with many antennas and users, where WMMSE approaches are not feasible due to the astronomic computational time, and low complexity precoding algorithms achieve poor and clearly sub-optimal performance, as illustrated in Fig. 19 where HRS (recall Fig. 2) is used in a massive MIMO setting with 100 antennas serving 12 users divided in four groups. Meta-learning provides significant performance enhancements over the conventional low complexity precoder of [42] at reasonable complexity and computational time [190]. Meta-learning also enables to optimize and explore the full performance benefits of MA schemes in large scale settings, with many antennas and many users, which could not be achieved so far by convex optimization techniques. For instance, Fig. 20 shows the ergodic sum-rate (ESR) of RSMA and SDMA in a massive MIMO deployment with 100 antennas and 40 users grouped into four disjoint (non-overlapping) groups (12 users per group) [191]. The relative gain of RSMA over SDMA, both optimized using meta-learning, increases as the CSIT error σ_e^2 increases[3]. This shows that the potential performance benefits of RSMA over SDMA hold in fully optimized large scale systems. Further investigations are needed to understand how those gains change when using a 3GPP-compliant evaluation methodology.

Additionally, algorithm-driven methods like deep unfold-

[3]The partial CSIT for user-k is given by $\hat{\mathbf{h}}_k = \mathbf{R}_g^{1/2}\left(\sqrt{1-\sigma_e^2}\mathbf{g}_k + \sigma_e^2\mathbf{z}_k\right)$, where $\mathbf{z}_k$ is the CSIT error with i.i.d. entries drawn from the distributions $\mathcal{CN}(0,1)$, $\sigma_e^2 \in [0,1]$ denotes the CSIT error variance, and $\mathbf{R}_g$ is the channel spatial covariance matrix of group g.

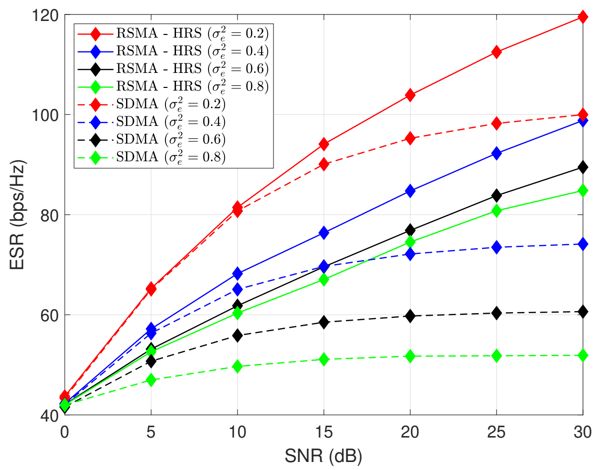


Fig. 20. Ergodic sum-rate with QoS rate constraints of 1 bps/Hz per user of RSMA and SDMA with meta-learning based optimization in a massive MIMO deployment with 100 antennas and 40 users [191].

ing/unrolling and learning to branch and bound (BB) have also played significant roles in MA resource allocation problems. As the problems are typically non-convex, one approach is to approximate the original non-convex problem to a sequence of convex problems, and use an iterative algorithm to solve the approximated problems. Deep unfolding/unrolling extends this idea by unfolding the iterations of an iterative resource allocation algorithm into a neural network architecture, allowing for end-to-end training and optimization. Such approach has been widely studied in beamforming design, user scheduling, power allocation, etc [192], [193]. Another standard method to deal with the non-convex optimization problem is to find the global optimal solution based on BB, which is of high computational complexity. To address this issue, a learning-based BB is proposed that use neural network to learning the BB algorithm [194].

3) Reinforcement Learning: RL involves the interaction between agents and the environment, with the goal of optimizing a long-term objective. Among the most widely embraced model-free RL algorithms, Q-learning is the most well-known algorithm for computing an optimal policy that maximizes the long-term reward. However, as system complexity deepens, particularly when there exists hidden system states, calculating and maintaining all Q-values becomes exceedingly impractical. This challenge, often referred to as the "curse of dimensionality", makes Q-learning less suitable for the intricate demands of ultra-dense and complex wireless networks [195]. To address this issue, deep RL (DRL) emerges as a powerful solution that integrates traditional RL with deep learning techniques.

Recent advances in DRL has made it a promising frontier for resource optimization in wireless network for different MA schemes [196]–[199]. DRL has found applications in a wide range of resource allocation problems, such as user scheduling, power control, beamforming design, and bandwidth allocation. DRL demonstrates remarkable capability in tackling the intricate challenges of sequential decision-making within dynamic and large-scale networks, all without relying on explicit models of the transmission environment. This unique feature empowers agents to continually refine their decision policies through interactions with the unknown environment.

4) Advantages and Disadvantages: A non-exhaustive search of existing works that use AI methods to allocate resources for difference MA schemes is summarized in Table III. The advantages and disadvantages of these AI-empowered algorithms are summarized as follows.

AI-empowered resource allocation algorithms offer the following advantages:

Superior Performance: Thanks to their data-driven features and the well developed neural networks, AI-empowered resource allocation can discover communication patterns and relationships that are too complex to identify precise mathematical model. Therefore, AI-empowered algorithms often achieve better performance compared to traditional model-based algorithms.

Generalization: AI-empowered resource allocation algorithms can generalize from historical data to make predictions or decisions on uncharted data. This capability empowers these algorithms to flexibly adapt to rapidly changing wireless environment and efficiently allocate resources.

Scalability: AI-empowered resource allocation algorithms can efficiently handle and process larger volumes of data. They possess the ability to accommodate growing demands without compromising performance or efficiency. Therefore, these algorithms are generally proficient in tackling resource allocation challenges of diverse sizes and complexities.

Low computational complexity: AI inference demands only a limited set of basic operations and can be executed in real-time. This appealing characteristic allows these algorithms to reduce processing time and require fewer computational resources, making them practical and cost-effective solutions for both online and offline resource allocation problems.

Robustness: As AI-empowered algorithms are data-driven, they are typically robust to handle uncertainty in data or environment, allowing them to make reliable decisions despite incomplete or noisy information. This robustness is particularly valuable in resource allocation problems because it ensures that the algorithm can adapt and make sound decisions even when operating in unpredictable or dynamic environments.

Although AI-empowered MA resource allocation solutions offer numerous advantages, they also come with certain disadvantages and challenges:

Lack of Interpretability: Directly replacing the conventional resource allocation algorithm with a generic neural network in an end-to-end fashion is often referred to as a "black-box" approach, which lacks interpretability regarding the decisions it makes. To address the issue, one can combine data-driven and model driven algorithms, i.e., by unfolding an iterative algorithm, it becomes possible to enhance interpretability.

Data Dependency: AI-empowered algorithms have a strong dependence on data for both training and inference. When the training data is biased, incomplete, or not representative, it can result in suboptimal resource allocation decisions. As the wireless environment is changing rapidly, user channels, typically

TABLE III
AI-EMPOWERED RESOURCE ALLOCATION FOR DIFFERENT MULTIPLE ACCESS SCHEMES.

MAs	AI Method	Reference	Resource Allocation Problem
OFDMA	Deep Transfer Reinforcement Learning	[200]	beamforming and subcarrier assignment
	DNN	[201]	subcarrier assignment
		[202]	power allocation
		[203]	user scheduling
	DRL	[204]	subcarrier assignment and power allocation
TDMA	RNN (LSTM)	[205]	radio resource assignment
	DRL	[206]	time-frequency resource block allocation
	DRL (Q-learning)	[207]	power allocation
SDMA	K Nearest Neighbors	[208]	beamforming
	K-means	[209]	user grouping
	CNN	[210]	hybrid beamforming
		[211]	beamforming
		[176]	
	GNN	[212]	
		[213]	
		[214]	
NOMA	GNN	[215]	user scheduling
	DQN	[216]	user pairing and power allocation
	DNN	[217]	joint beamforming and SIC decoding
RSMA	Transformer	[218]	beamforming
	DNN and autoencoder	[219]	
	Meta-Learning	[190]	
	DRL	[220]	power allocation

used as input data, may shift to different distributions over time. This can significantly impact the ability of the trained neural network to generalize effectively to new conditions and scenarios.

Training Overhead: Developing and training AI models for wireless resource allocation can be time-consuming and resource-intensive. This is especially evident when utilizing traditional resource optimization algorithms, known for their high computational complexity, to generate the training data, leading to substantial time requirements. Furthermore, to maintain optimal performance, periodic model updates may also be a necessity.

B. AI-empowered MA Channel Estimation

CSI plays a pivotal role in wireless communications and MA designs and optimization. Precise CSIT facilitates the implementation of precoding, adaptive resource allocation and user scheduling, leading to substantial improvements in performance and efficiency, while accurate CSI at the receiver (CSIR) empowers coherent detection and decoding processes. For acquiring CSIR, downlink channel estimation is a necessity, where the transmitter sends pilot signals to the receivers. This task becomes particularly vital and challenging in OFDM systems due to the temporal variability and frequency-selective characteristics of wireless channels. In terms of CSIT acquisition, in time division duplex (TDD) systems, CSI is obtained at the transmitter through the exploitation of reciprocity. In this case, uplink channel estimation is required, where users send pilot signals, enabling the base station to estimate the uplink CSI. However, tackling this task in massive access is challenging due to the limited number of orthogonal pilot resources [221]. In frequency division duplex (FDD) systems, acquiring CSIT involves sending the CSI estimated at the receiver back to the transmitter through a feedback link. However, this becomes challenging in massive MIMO setups due to the large dimensions of the CSI, resulting in a substantial increase in feedback overhead. To mitigate the overhead, it becomes imperative to efficiently compress the CSI estimate at the receiver or employ a codebook-based approach for quantizing the CSI [222], [223].

AI can play a pivotal role in tackling the aforementioned challenges and improving the quality of CSI obtained at either the transmitter or receiver. This can be achieved through two primary directions [224]:

1) Channel estimation: By modeling the channel estimation problem as a regression task, we can create a neural network-based channel estimator.
2) CSI compression: By modeling CSI compression as a dimension reduction problem, we can develop AI architectures specifically tailored to address this issue.

These two directions in general apply to different MA techniques. However, for different MA enabled channel estimation, some additional problems may arise. In the following, we will respectively discuss these challenges for different systems employing different MA schemes.

AI-enabled channel estimation: Deep learning enabled channel estimation is first studied in [225] for a point-to-point OFDM system, where a DNN is employed to learn the characteristics of frequency selective wireless channels. After that, many existing works focus on using neural networks, such as CNN, generative adversarial network (GAN) to tackle channel estimation of MIMO networks [226]–[228]. As summarized in [229], most of existing works on channel estimation consider point-to-point communication networks. However, these approaches often face limitations when applied to multi-user MIMO systems due to the presence of multi-user interference. In uplink multi-user MIMO, one way to mitigate inter-user interference is by assigning different users orthogonal pilots. However, due to the limited availability of these sequences, users often end up using the same pilot for training, leading to pilot contamination [230]. To tackle this problem, uplink NOMA can be considered. It allows non-orthogonal pilot transmission and uses SIC for channel estimation [231]. However, since multiple users sharing the same pilots in uplink NOMA, it introduces additional power control issues. Hence, it is necessary to jointly design pilot and power allocation for better performance compared to traditional methods without SIC [232]. In addition to uplink NOMA, RSMA also shows promise in resolving pilot contamination problems. An interesting finding in [56] suggests that downlink RSMA can assist in reducing the negative impacts of pilot contamination in TDD massive MIMO. Such appealing benefit of RSMA makes it highly promising for simplifying the channel estimation process in the uplink.

AI-enabled CSI compression: In FDD massive MIMO systems, to reduce the CSI feedback overhead, one line of works focuses on CSI compression methods and codebook designs. While compression theories such as compressive sensing have been applied to simplify the CSI feedback or codebook design [233], [234], the feedback signaling overhead in these approaches remains heavy since the overhead typically increases linearly with the number of antennas, imposing practical limitations. Deep learning has shown to be highly effective in addressing data compression challenges across various domains, including images, audio, and video [235], [236]. Recognizing CSI as an additional data source for compression, deep learning techniques have been successfully applied to the compression and feedback of CSI for massive MIMO systems [237]–[243]. Among various MA schemes, NOMA exhibits greater sensitivity to CSI quantization/compression errors. These errors not only affect the decoding order design but also lead to practical issues of imperfect SIC and error propagation during downlink transmission [244]. Thanks to its significant advantages in interference management, RSMA has demonstrated appealing spectral efficiency gains compared to NOMA and SDMA in FDD massive MIMO systems with limited feedback [42], [245]. It holds great promise for simplifying deep learning based CSI compression or codebook design in FDD massive MIMO setups while maintaining equivalent or superior performance to conventional MA schemes.

C. AI-empowered MA Receiver Design

Aforementioned MA schemes often require advanced receivers, such as SIC for NOMA and RSMA, though other types of receivers can be used to better trade performance with receiver complexity [45]. One commonality in the design of receivers for MAs is the reliance on models and assumption for the noise, interference, signal propagation [246]. It is nevertheless challenging to mathematically relax those assumptions while maintaining a tractable model. Indeed interference is non-Gaussian with finite constellations, SIC and CSIR are imperfect and propagates error, decoding delay and latency increase chance of errors, etc. Those model-based receiver algorithms have so far performed quite well [45] but take the risk to rely on accurate prior model knowledge and potentially perform poorly if it is not accurately acquired [246].

DL-based receivers are able to directly extract meaningful information from the unknown channel solely on observations, which is a major advantage. Therefore, DL is well suited for scenarios in which the underlying mathematical channel model is unknown, its parameters cannot be acquired with precision, or when it is too complex to be studied by model-based algorithms with low computational resources [247]. Nevertheless, employing DL in MA receivers presents substantial challenges. Firstly, the appeal of model agnosticism in DL approaches necessitates a complex network with numerous nodes and layers, along with a sizable training set to acquire a specific mapping. Consequently, this results in a substantial computational burden at the receiver during the training phase. Secondly, sharing a training set between the transmitter and receiver introduces significant training overhead, potentially causing delays in transmitting actual useful data to communication users. Lastly, due to the time-varying and dynamic nature of wireless channels, periodic training of the DNNs at the receivers becomes essential to accommodate channel variations. However, this requirement renders the sharing of large training sets highly impractical.

To tackle those challenges and leverage the advantages of both traditional model-based algorithms and model-agnostic DL approaches, model-based deep learning (MBDL) has been introduced [246]. MBDL systems are implemented by substituting specific steps and computations in model-based algorithms, which rely on precise channel model knowledge, with compact neural networks. These neural networks demand smaller training data sets compared to conventional DL systems. In this context, adaptations of the SIC receiver using MBDL have been proposed for uplink and downlink NOMA and downlink RSMA [248]–[250], such that specialized DNNs are employed for tasks such as interference cancellation and symbol classification, as illustrated in Fig. 21 for RSMA. Interestingly, the MBDL approach surpasses the performance of conventional model-based SIC receivers. Fig. 22 illustrates the throughput performance of RSMA receiver architectures of Fig. 21. Eight transmit antennas serve eight users using 1-layer RS scheme (recall Fig. 2 - hence one common stream and eight private streams) which features finite alphabet modulation, finite-length polar coding, Adaptive Modulation and Coding (AMC), and rate backoff. We observe that the throughput loss of the SIC receiver with imperfect CSIR increases with the SNR. In comparison, the MBDL receiver is able to achieve a similar throughput to that of the SIC receiver with perfect CSIR, with an almost negligible loss, and comes close to the performance of the optimum (but highly complex - hence not practical for large number of streams and modulation order) joint decoding receiver, the maximum a-posteriori probability (MAP) receiver. The significant margin of the MBDL receiver over the SIC receiver is due to its ability to generate on demand non-linear symbol detection boundaries in a pure data-driven manner [250].

D. AI-empowered Grant-Free Random Access

Another MA area where AI can play an important role is the grant-free random access scheme discussed in Section II-H. In massive IoT connectivity setup, the number of IoT devices is huge. As a result, low-complexity MA design suitable for

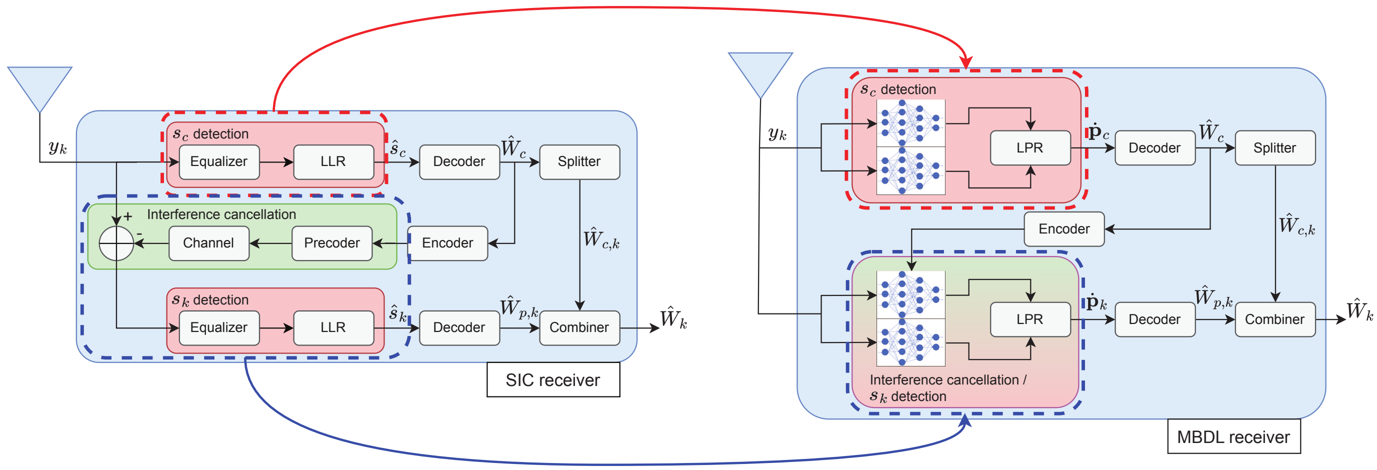


Fig. 21. 1-Layer model-based deep learning RSMA receiver and relationship with 1-layer SIC RSMA receiver [250]. The bit log-probability ratio (LPR) can be used as substitutes for the bit log-likelihood ratios (LLRs) that are used in conventional channel decoders.

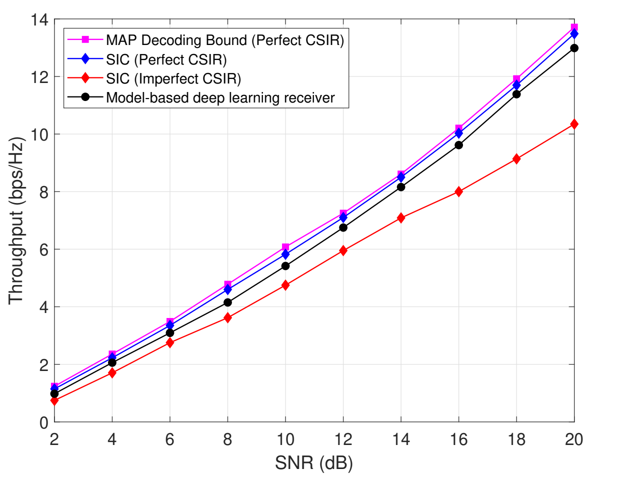


Fig. 22. Throughput performance of RSMA with model-based deep learning receiver and SIC receiver [250].

large-scale IoT network is appealing. In this sense, recently, AI-empowered MA schemes have attracted more and more attention in the area of grant-free random access under massive machine-type communications.

On one hand, the performance of compressed sensing algorithms introduced in Section II-H heavily depend on the choice of the sensing matrix, while most of the works on compressed sensing based device activity detection assume that the sensing matrix, which is determined by the pilot signals assigned to IoT devices, is given. For example, many works assume Gaussian pilots [144], [159], [251] such that the approximate message passing (AMP) algorithm can be adopted to detect the active devices and analyze the detection performance based on the state evolution. Moreover, Reed-muller sequences based pilots have been utilized for device activity detection in [252]. However, how to design the pilot signals that work the best for device activity detection is still an open problem. This difficulty lies in the unknown relation between the sensing matrix and the detection performance. It is well-known that AI techniques are a good option to learn the complicated and unknown relations. Therefore, recently, there is a growing interest in applying AI to design the pilot signals for massive IoT connectivity. For example, it was found in [253], [254] that if the deep learning technique is used to design the pilot signals, the device detection performance can be improved compared to using Gaussian pilots and Reed-muller sequences based pilots.

On the other hand, deep learning technique can also be applied to design efficient device activity detection algorithms, when the dimension of the problem is huge due to the large number of IoT devices. In general, these deep learning based device activity detection methods can be classified into data-driven based methods and model-driven based methods. Under data-driven based methods, basic neural network architectures, e.g., fully connected neural networks, CNNs, are utilized to detect active devices [254], [255]. Under model-driven based methods, existing algorithms, e.g., AMP, group LASSO, etc., are utilized to train the neural networks for better activity detection performance [256]–[259].

E. Technology Outlook and Future Works

AI has been investigated for some relatively simple MA schemes such as OMA and SDMA, but more needs to be done for more advanced MA schemes, such as RSMA, CD-MA and random access, for which the literature still remains quite limited with a lot of new research challenges and opportunities on the horizon. AI is likely to play an even bigger role for UMA for which the optimization space is enlarged due to the exploitation of all five dimensions (and sub-dimensions). Moreover, AI could also be helpful to complement communication and information theory to discover and identify key ingredients of UMA designs and architectures and also help with receiver design for the same reasons as RSMA but also because of the additional complexity increase incurred by the exploitation of the code dimension.

IV. Multiple Access for Artificial Intelligence

We now switch the role and assess how MA schemes can be designed and tailored to serve AI applications. We focus on MA for federated learning (FL) and edge intelligence.

A. *MA for Federated Learning and Edge Intelligence*

FL is a decentralized method for training ML models. Specifically, FL training consits of four main steps: 1) each device uses its raw dataset to train its ML model, 2) each device sends its local FL model parameters (i.e., gradient vector) to a parameter server, 3) the parameter server aggregates the received local FL model parameters of devices and generate a global FL model, 4) the parameter server sends the global FL model parameters back to all devices, 5) repeat steps 1)-4) until model converges. From the training process, we can see that FL eliminates the need for transferring data from client devices to parameter servers since the devices exchange FL model parameters. Hence, FL enhances data privacy. One challenge of deploying FL [260] over wireless networks is the communication bottleneck, which arises from several edge devices transmitting large sized model parameters to a parameter server or a base station. Moreover the heterogeneity of edge devices computing capabilities, communication rates, and amount and quality of data can affect the training performance in terms of accuracy, fairness and convergence time. Researchers have attempted to reduce the resultant communication latency using different approaches such as excluding slow devices (“stragglers”) [261] or compressing FL model parameters by exploiting their sparsity. To further improve the number of FL participating devices, in this subsection, we provide a review on the use of MA schemes for supporting FL deployment over realistic wireless networks, with a specific emphasis on OMA/SDMA/NOMA/RSMA-based networks and an alternative MA approach – over-the-air computation (AirComp) - for wireless FL performance optimization.

1) OMA for FL: Using OMA, devices must use different spectrum resource for FL parameter transmission. Due to limited resources in wireless networks, the wireless resources (i.e., power, computational power, spectrum) allocated to each user are limited, which may significantly limit the number of devices that can participate in FL and increase FL convergence time [262]–[264]. Consequently, it is necessary to optimize resource allocation such that devices can efficiently complete the FL training process. The current researches have optimized device selection [265]–[267], local and global FL model updates [268], [269], spectrum allocation [270], and used advanced technologies such as RIS [271], [272], MIMO [273], compression [274] to improve training loss, convergence speed, or training complexity of FL in OMA based networks. However, the main drawback of OMA is that it does not scale well with the number of devices. Specifically, the required radio resources increase linearly with the number of devices, or else the latency will grow linearly. This calls for the integration of FL with other MA schemes.

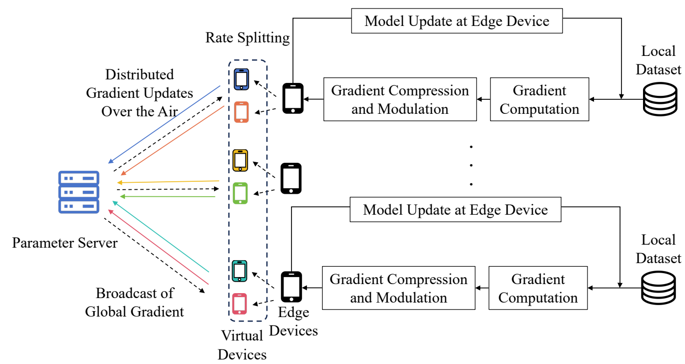


Fig. 23. An illustration of an RSMA-based FL.

2) SDMA for FL: To solve OMA drawback and enhance the applications of FL, one can use MIMO and spatial domain techniques, which enable edge devices to use the same time-frequency resources for FL parameter transmission thus improving wireless resource utilization and FL convergence speed. However, using MIMO for FL requires to optimize transmit and receiver beforming matrices. In particular, the parameter server requires to know the future gradient vectors of devices to optimize its beamforming matrix since beamforming matrix depends on FL model parameter aggregation.

3) NOMA for FL: To further improve the FL parameter transmission efficiency, one can use NOMA to separate edge devices into distinct groups and managed the interference in each group by SC-SIC. Here, current works have studied the combination of NOMA with gradient quantization and sparsification schemes [275], power allocation schemes [276], device scheduling [277] to improve FL performance. However, since NOMA decodes the interference of all other users, it lacks in flexibility and the designed FL algorithms in NOMA based networks may not be used for scenarios where each user needs to transmit with a strict rate demand due to the high dimension of model gradient. Moreover, since the FL process requires multiple transmission rounds and accurate knowledge of the CSI in all those time slots is unlikely, the use of MA schemes robust to imperfect CSI is important in FL applications. However, NOMA cannot be well adapted to the imperfect CSI scenarios [11].

4) RSMA for FL: To address the issues introduced by NOMA techniques and benefit from the space and power domains, one can use RSMA to efficiently balance the signal and interference and cope with imperfect CSI. In particular, due to the flexibility of signal design and decoding, RSMA can achieve robust performance with imperfect CSI, which is suited to the long-term transmission in FL. Fig. 23 shows an illustration of an RSMA-based FL. In this figure, each device is considered as two virtual devices that jointly transmit local FL model with different data rates. Meanwhile, RSMA is appropriate for supporting the deployment of clustering FL. The work in [278] has shown that RSMA can not only significantly reduce the latency of clustering FL in comparison to systems employing TDMA and NOMA, but also be more energy efficient, reaching a lower latency with less power than the latter alternatives. Another promising scenario of RSMA for FL is in a fog radio access network (F-RAN). F-RAN leverages proximity edge nodes (ENs) to collect local model parameters of wireless devices [279]. By splitting the message in two parts in uplink RSMA as in Fig. 23, one part is intended to be decoded by the ENs and the other part by the cloud. By adjusting the content and the transmit power of each part, one can softly combine the edge-decoding and cloud-decoding approaches and consequently achieve a better tradeoff between fronthaul overhead (communication cost) and learning accuracy.

5) Over-the-air Computation for FL: An alternative MA approach – over-the-air computation (AirComp) can also be considered for wireless FL performance optimization. In particular, AirComp enables edge devices to achieve over-the-air FL parameter aggregation via the waveform superposition property of a multi-access channel [140]. Hence, when edge devices implement AirComp, the FL transmission delay per iteration will not depend on the number of edge devices thus improving FL convergence speed. Note that AirComp uses "interference" among device transmission for functional computation via uplink devices' cooperation. This contrasts with the usual uplink MA schemes that aim at suppressing interference among transmitted messages as devices transmit independent data and do not collaborate.

A key requirement for implementing AirComp based FL is time synchronization of devices' transmissions. To address this challenge, one can use timing advance methods which have been used in several existing practical communication systems (e.g., LTE). The timing advance technique requires each edge device to estimate the FL parameter propagation delay and then transmitting in advance to "cancel" the delay. Thereby, different signals can overlap with sufficiently small misalignment. The signal distortion of AirComp stems from channel noise and interference of analog modulated signals. Therefore, channel noise and interference will affect the model updates of AirComp based FL. The impacts of channel noise and interference on FL model updates can be evaluated by the corresponding FL performance metrics.

AirComp can be classified into two categories: 1) Analog AirComp and 2) Digital AirComp. Analog AirComp uses discrete-time analog transmission for FL parameter transmission and aggregation. However, the performance of analog AirComp is sensitive to the phase offsets among the super imposed signal. In particular, if the signals cannot be synchronized, the errors can also be accumulated thus leading to significant aggregation errors. To address this problem, digital AirComp is proposed which enables multiple devices to use non-orthogonal subcarriers for data transmission. Different from analog AirComp that relies on accurate channel precoding, digital AirComp leverages digital modulation and channel codes to combat channel impairments and phase misalignments, thereby achieving accurate model aggregation even when the signal phases of multiple edge devices are misaligned at the parameter server.

B. Technology Outlook and Future Works

Different MA techniques have their own advantages and disadvantages as discussed in Section II. For FL, the selection of MA techniques depends on the number of participating edge devices, wireless resource (i.e., spectrum, transmit power), computational power, and hardware (i.e., CPU or GPU), FL model architecture of each device. To find the optimal MA schemes, one must first analyze the convergence of FL and figure out how different MA schemes affect FL convergence. However, in practice, the selection of MA schemes must be performed before FL implementation. Hence, the central controller (i.e., base station or parameter server) may not be able to know channel state information and FL model setting information (i.e., gradient information), which may increase the difficulty of MA scheme selection. Therefore, MA scheme selection also depends on the available wireless and FL setting information. The above discussion further motivates the design of unified and universal schemes such as RSMA and ultimately UMA, which would completely eliminate the need for selection of MA schemes and whose design and optimization could be fully integrated with FL. Also, the use of over-the-air computation opens the door to other types of uplink MA schemes that brings back analog computation into the MA design picture.

V. Multiple Access for Integrated Sensing and Communications

In this section, we discuss the use and design of MA schemes for ISAC, including joint sensing and communications, multimodal sensing-aided communications, digital twin-assisted communications, and communication-aided sensing/localization systems.

A. MA Assisted ISAC

ISAC is primarily motivated by the potential of sharing the same spectrum and, hence, providing access to more bands for both the communication and sensing systems. Operating at the same band, however, requires careful coordination of the various wireless resources and necessitates the development of new communication and sensing systems as well as joint processing frameworks. When a base station is jointly serving communication users and performing sensing functions, MA schemes are critical to efficiently allocate the available resources in time, frequency, space, code, power, etc., and optimize the waveform for both the communication and sensing objectives. In addition, MA schemes are also required to mitigate inter-user interference and sensing-to-communication interference. In what follows, we review several design examples and highlight some key directions on MA-assisted ISAC.

There are two main approaches to the design MA-assisted ISAC schemes. In the first approach, the communication signals of a multiple access scheme are exploited for sensing. Multiple access is employed within the communication functionality to mitigate inter-user interference. This approach can be deemed as a straightforward extension of employing MA from the conventional communication system to the ISAC system, such that the key features of the MA can be retained. For this approach, research mainly focuses on the optimization of the communication performance subject to additional sensing performance constraints, where the underlying MA schemes

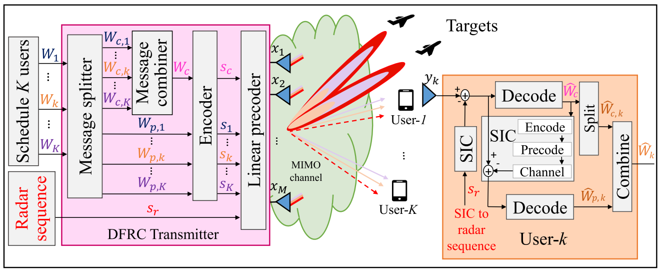


Fig. 24. RSMA-assisted ISAC system with an additional radar sequence [164].

can be SDMA, NOMA, and RSMA [19], [280]. Compared to the first approach, the second approach uses additional sensing waveforms (may or may not carry information) to improve sensing quality. An example of such an approach is shown in Fig. 24, where the underlying MA scheme is RSMA. In this case, MA is employed between the communication and sensing functionalities to mitigate both inter-user interference and sensing-to-communication interference. For SDMA-assisted ISAC [281], additional radar waveforms improve the communication and sensing performance tradeoff than the first approach [280]. However, this does not provide any performance enhancement for RSMA-assisted ISAC as demonstrated in [164].

In summary, various works have demonstrated that the superior performance of the MA schemes in communication applications can be carried over to ISAC. Hence, the MA scheme that has the best communication performance would also achieve better communication and sensing performance tradeoff than other MA schemes in ISAC. This has been verified in [19] that RSMA-assisted ISAC outperforms the SDMA and NOMA counterparts, as illustrated in Fig. 25. RSMA benefits from the fact that the common stream is not only used to manage multi-user interference but also helps enhancing target sensing. On the other hand, NOMA-assisted ISAC achieves the poorest tradeoff due to the degrees of freedom (DoF) loss in multi-antenna NOMA [11]. At the leftmost points which correspond to prioritizing the radar functionality, the optimized precoders are designed to radiate the highest power towards the target angle. This forces the precoders to become linearly dependent and SDMA-assisted ISAC can no longer exploit spatial DoF provided by multiple antennas, which leads to lower MFR compared with the RSMA-assisted and NOMA-assisted ISAC that employ SIC at user sides to manage the multi-user interference. In summary, pursuing the best-achieving MA scheme is meaningful in both pure communication applications and ISAC.

B. Multimodal Sensing-aided Communications

Two key trends in current and future communication systems are (i) the use of large numbers of antennas and (ii) the increasing dependence on high frequency bands. These trends, while promising high data rates, pose multiple challenges for the design and operation of future communication systems. For example, the use of large numbers of antennas increases

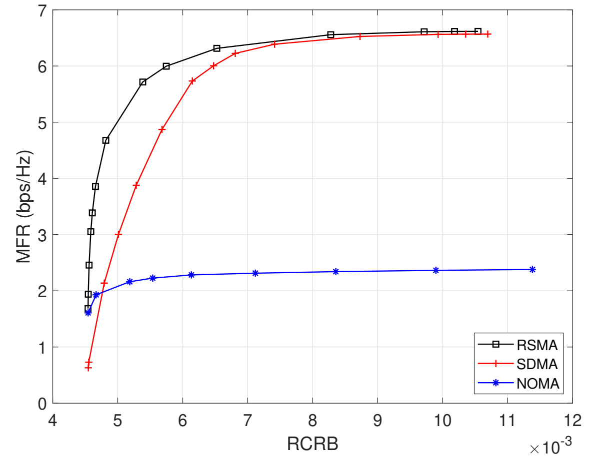


Fig. 25. Minimum fair rate (MFR) versus root Cramér-Rao bound (RCRB) of target direction in a terrestrial ISAC system with 8 transmit antennas, 9 receive antennas, 4 single-antenna communication users and one sensing target [19].

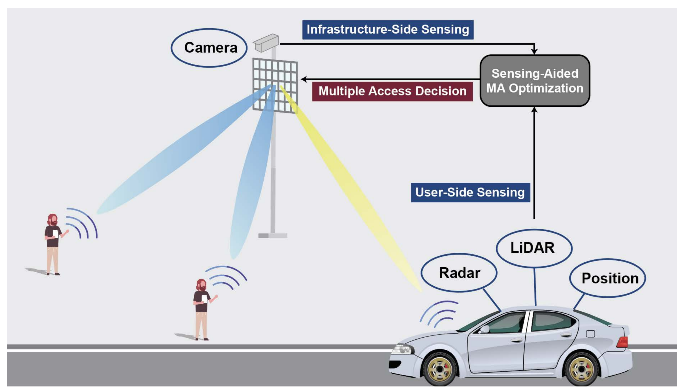


Fig. 26. The figure illustrates the key idea of multi-modal sensing-aided multiple access, where infrastructure-side and user-side multi-modal sensing information from cameras, LiDARs, radars, position sensors, etc., is leveraged to optimize the multiple access decisions.

the overhead of estimating the channels or training the beams, making it hard for these systems to support highly mobile applications. Further, the use of high frequency bands makes the signal propagation very sensitive to blockages, which challenges the reliability and latency requirements for these networks. To address these challenges, a new direction that has been recently gaining increasing interest is the use of multi-modal sensing data, collected for example from visual cameras, LiDAR, radar, position sensors, to aid the communication network decisions [282]. This is motivated by the increasing dependency of the large antenna array and high-frequency channels on the user location and on the geometry of the surrounding environment. Over the last few years, researchers investigated the potential gain of leveraging multi-modal sensing data in multiple problems, such as beam selection/tracking [283]–[292], proactive blockage prediction [293]–[299], channel statistics prediction [300], codebook design [301], among other interesting applications.

Multi-modal sensing provides some perception about the location and geometry of the user devices, infrastructure nodes, and various elements of the environment. This perception can enable more efficient MA operation in several ways. First, multi-modal sensing may enable the base stations and access points to intelligently select and schedule users to maximize the MA system performance. For example, leveraging the absolute or relative user position information, obtained by GPS, radar, LiDAR, or camera, the base station may schedule the users that are well-separated on the same time-frequency resource to have less interference. In dense deployment, the sensing information may also be utilized to assign users to base stations/access points or to proactively predict handover, which yields higher reliability, less latency, and better session continuity. Beyond user selection and scheduling, the underlying information captured by multi-modal sensing can guide the multiple access resource allocation decisions. For example, in NOMA, multi-modal sensing data can potentially aid the optimization of the transmit power of the different devices. Further, in OMA, this sensing data may enable adaptive allocation of the time, frequency, and spatial resources based on the users relative positions, mobility patterns, and dynamic scatterers in the environment. All that motivates investigating the various promising applications of multi-modal sensing in emerging MA schemes.

Realizing the potential gains of multi-modal sensing aided MA in practice, however, requires overcoming a number of challenges. First, a fundamental challenge in using sensors such as cameras, LiDARs, and radars is identifying and differentiating the communication users in the sensing scene. This task, which is defined as user identification [302], is a key for enabling multi-modal sensing-aided communication in realistic environments. Initial approaches for user identification rely on augmenting other user-specific measurements such as sparse/limited channel or beam power measurements [284], [303]. In addition to user identification and differentiation, the coverage mismatch between the multi-modal sensing and communication systems raises critical challenges. Overcoming that motivates the extension to distributed multi-modal sensing which involves the design and coordination of distributed sensors to optimize the MA objectives. Addressing these important challenges defines interesting directions for future research in the interplay between MA techniques and multi-modal sensing systems.

C. From Multimodal Sensing to Digital Twin-Assisted Communications

Building upon the advances in multi-modal sensing aided communications discussed in Section V-B and targeting a more comprehensive perception of the communication environment, the concept of (near) real-time digital twins of the physical environment has been proposed [304]. In these digital twins, the multi-modal sensing information is fused with the 3D maps to construct real-time or near real-time maps of the surrounding environment. Running real-time ray-tracing, which could potentially be machine learning enhanced, on these 3D maps leads to (near) real-time digital twins of the physical communication system. Using these digital twins, the communication systems can pre-train ML models for the different communication tasks, optimize the various network procedures, or directly make predictions for the real system. This is particularly valuable when these digital twins are sufficiently calibrated with the physical world.

The relationship between wireless communication and digital twins is mutually beneficial. With more efficient communication of the sensory data, more accurate and updated digital twins can be constructed. Further, as presented in [304], the digital twins with real-time 3D maps of the environment can be jointly leveraged to optimize the communication networks and to assist other applications such as autonomous driving, proactive collision prediction, traffic management, surveillance, and many other applications. Towards this objective, MA techniques play a central role in communicating the data from the distributed sensors in autonomous vehicles, IoT devices, distributed infrastructure side units, etc., needed for constructing and maintaining the digital twins. For example, it is important to develop digital twin-specific MA approaches that optimize the resource allocations among different sensors and modalities with the required digital twin fidelity and freshness as objectives [305].

On the other side, these environment digital twins could also be very beneficial for the design of MA solutions. In particular, the digital twins provide the communication network with comprehensive perception, that is potentially near real-time, about its surrounding environment. This perception can assist the various resource allocation decisions. For example, the digital twins could be leveraged to decide on what users to schedule in a specific MA transmission based on their relative locations, mobility patterns, and interaction with the nearest blockages. They can also be leveraged to develop adaptive and site-specific MA techniques that dynamically change their parameters based on the specific site characteristics. This highlights the promising gains when leveraging environment digital twins to assist MA designs and optimization in future communication systems.

D. Communication-Aided Sensing/Localization Systems

Localization technologies have been embedded in mobile communication systems for decades [306]. It is well known that satellite-based positioning technologies, e.g., Global Navigation Satellite Systems (GNSS), have been popular. However, they suffer from severe degradation in indoors and urban areas. To address these challenges, cellular-based localization methods have been considered due to their natural advantages of wide coverage, easy deployment, and low cost [307]. As we have seen in the previous sections, MA schemes are capable of providing larger spectral and energy efficiencies in multi-user communications. Hence, under limited resources, MA techniques can contribute to the potential improvement of positioning accuracy and coverage for multiple users via efficient resource and interference management.

The use of RSMA to assist direct localization was investigated in [308], where multiple base stations cooperatively communicate with multiple users and directly estimate a target's 3D locations simultaneously. An example with N cooperative base stations and K users is illustrated in Fig.

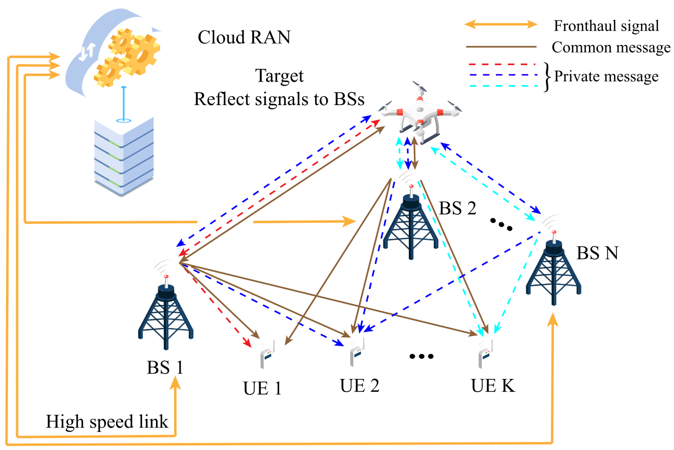
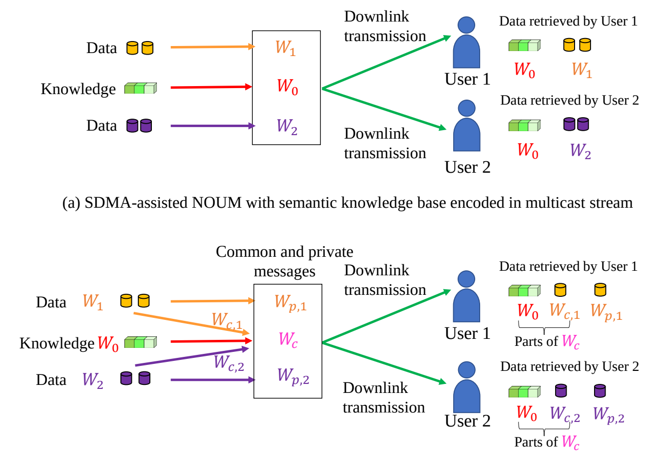


Fig. 27. RSMA assisted direct localization with N cooperative base stations serving K users while estimating a target's location [308].



(a) SDMA-assisted NOUM with semantic knowledge base encoded in multicast stream

(b) RSMA-assisted NOUM with semantic knowledge base and parts of unicast data encoded in common stream

Fig. 28. RSMA/SDMA-aided NOUM for SeCom.

27, where the common and private messages are used for both communication and localization. Due to the efficient resource allocation and interference management, RSMA-assisted localization achieves better sum rate and localization accuracy than TDMA and FDMA based localization schemes. When the location accuracy requirement is high, RSMA can achieve a larger sum rate than SDMA even though the number of transmit antennas is larger than the number of users. As pointed out in [309], localization can be performed in both channel estimation and data transmission. In this case, the duration of pilot and data transmission periods can be optimized to enlarge the rate-accuracy region for a given channel coherence time. Such an approach, however, has only been investigated for SDMA [309].

E. Technology Outlook and Future Works

The design of efficient MA techniques for ISAC systems is a challenging task. First, the joint optimization of the resources and waveforms is typically associated with high computational complexity. This complexity mainly stems from the differences in the communication and sensing objectives and from the various hardware and operation constraints on these systems. This becomes more complicated in large-scale MIMO systems which are main components in 5G, 6G, and beyond. Going from nodes to networks, it becomes critical to design multiple access approaches for networks of integrated sensing and communication nodes. This requires accounting for the possible interference between different nodes in both the sensing and communication signals. With that, it becomes interesting to explore the tradeoff between coordination performance gains and complexity in multiple access ISAC networks. This results in a set of approaches that range from no coordination to fully coherent ISAC operation where the full network acts as one cell, leading to cell-free ISAC systems [310]. How to design MA schemes in these coordinating ISAC networks is an interesting research direction. Finally, the fundamental tradeoff between sensing and communication performance of general ISAC systems with MA has not been fully understood. Only a few prior works have characterized the tradeoff for very basic channel models with small numbers of users, e.g., [311]. Hence, to gain more insights into coding, signaling, and waveform designs for more general channels, it is also important to study the fundamental tradeoff in terms of achievability and converse for these channels. Last but not least, though RSMA-assisted ISAC has been shown to outperform SDMA/NOMA-assisted ISAC thanks to RSMA universality and flexibility advantages, it would be important to explore how other MA domains could be exploited for ISAC and get a complete picture of what UMA-assisted ISAC would look like.

Using MA for assisting localization is a new research area. In light of the above works, a number of challenges are identified. First, the above works assume static users. However, the impacts of user movements and synchronization errors in MA-assisted localization have not been fully investigated. Second, the performance metric for measuring positioning accuracy is generally based on the Cramér–Rao bound. Explicit designs of practical localization estimators to achieve this bound are still lacking. Third, the MA-assisted localization schemes based on SIC could cause some decoding latency, which may not be desirable for delay-sensitive applications. To fully exploit the benefits of MA-assisted localization in wider communication scenarios, these challenges need to be addressed in future works. Additionally, it is crucial to understand how one can exploit all MA dimensions in the design of MA-assisted localization beyond the SDMA/NOMA/RSMA approaches investigated so far and ultimately aim for a UMA-assisted localization.

VI. Multiple Access for Emerging Intelligent Applications

A. Semantic Communications

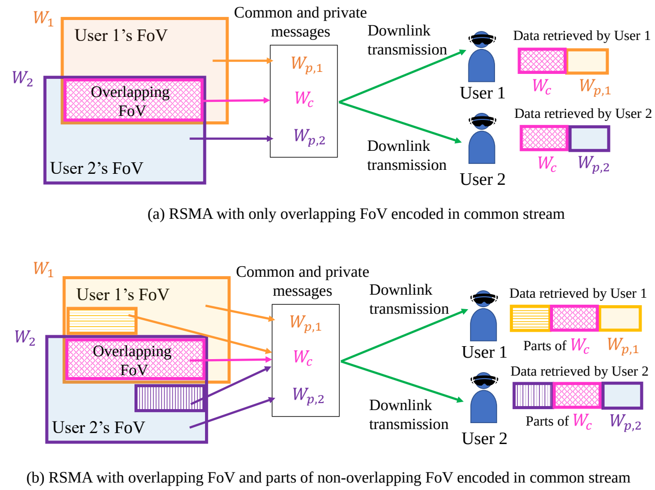
Semantic communications (SeCom), regarded as an advancement beyond the Shannon paradigm, strives to successfully transmit semantic information from the source. Unlike the conventional Shannon paradigm that focuses on precision in receiving individual symbols or bits regardless of the meaning, SeCom greatly enhances the communication efficiency and reliability by prioritizing the transmission of meaningful information [312]–[315].

In SeCom, it is typical for both the transmitter and receiver to possess a shared knowledge base. This shared knowledge base enables the extraction of compact semantic information at the transmitter and the reconstruction of the original information from compressed semantic signal at the receiver [316]. To handle additional knowledge information, it is important to design the communication system to smartly synergize MA schemes and SeCom. Naturally, OMA, NOMA, SDMA, RSMA, and others can be used for SeCom and the respective performance benefits of the different schemes would still carry on in SeCom. However, the presence of the knowledge base brings some changes. Indeed in a multi-user setting, if the knowledge base is the same for all users and is transmitted and received in a unicast way at each user, a spectral efficiency loss is incurred. A more efficient way is to use a NOUM transmission for SeCom as illustrated in Fig. 28. Using multicast, the resource allocated for the knowledge base can be greatly reduced and SDMA/RSMA-assisted NOUM of Fig. 7 can be leveraged to deliver SeCom to multiple users in a natural and efficient way. Indeed, the knowledge base intended to multiple users and parts of the user data can be encoded into the common stream, while the remaining parts of the individual data intended to a specific user are encoded into respective private streams. As a result, RSMA-aided SeCom outperforms NOMA and SDMA counterparts [317].

Future directions of MA schemes for SeCom should explore the integration of MA design in semantic information extraction and resource allocation for RSMA and CD-MA, and ultimately UMA to exploit all MA dimensions. Another aspect to consider is energy efficiency of SeCom. Compared to conventional communication, the total energy of SeCom involves both computation and communication energy. The total energy of SeCom can be optimized through optimizing the semantic compression ratio, defined as the ratio of the bits of the extracted semantic information to the bits of the original information [318], [319]. Re-thinking MA scheme design to maximize energy efficiency of SeCom is also of much importance.

B. Virtual Reality

Three MA schemes are commonly used in the literature for VR video streaming, namely OMA, NOMA, and RSMA. OMA supports a number of users no larger than the number of available time-frequency orthogonal radio resources. NOMA cannot exploit the full potential offered by the multi-antenna domain. RSMA manages multi-user interference by decoding a part of the interference, considering it as the common message of VR users, while treating the remaining part as noise, representing the private messages of other VR users. As expected, RSMA has been demonstrated to outperform other MA schemes, such as OMA and NOMA, in VR applications [320]–[323]. Examples of using RSMA for VR applications is shown in Fig. 29. Two different schemes are illustrated. For



(a) RSMA with only overlapping FoV encoded in common stream

(b) RSMA with overlapping FoV and parts of non-overlapping FoV encoded in common stream

Fig. 29. Examples of using RSMA for VR applications. Users 1 and 2 have overlapping field of view (FoV). The video tiles with overlapping FoV are sent via the common message. The video tiles with non-overlapping FoV are sent via the private messages in (a) and partially via common and private messages in (b).

both schemes, the video tiles corresponding to overlapping field of view (FoV) between users are transmitted via the common message. The video tiles with non-overlapping FoV are sent exclusively via the private messages or partially via common and private messages. Following RSMA principles, the second approach has more flexibility to manage interference among non-overlapping FoV video tiles.

Some potential research challenges in MA design for VR streaming systems encompass several areas. Firstly, there is a need to develop low-complexity receivers for VR users that account for the characteristics of their head-mounted displays (HMDs), including limited battery and computational resources. Secondly, higher frequency bands, such as mmWave and THz, can be employed in VR streaming systems to fulfill the bitrate requirements of high-resolution 360-degree videos [324], [325]. The use of RSMA in combination with beamforming design in these high-frequency bands represents an interesting research area. Thirdly, the integration of RSMA with other MA schemes, such as OFDMA, has not yet been thoroughly investigated to determine whether RSMA performance can be further enhanced in VR video streaming systems. Fourthly, in VR streaming systems, certain information, such as users' head movements, needs to be transmitted from the users to the server. Utilizing RSMA, users can share the uplink channel to send their information simultaneously. Finally, FL can be applied within a VR streaming system to obtain a viewport prediction model for the users [326], [327]. RSMA can be incorporated into FL to facilitate model aggregation without compromising model accuracy in such systems. The high data rate and low latency requirements for VR video streaming position it as one of the key FeMBB services provided by 6G systems. RSMA is one of the potential MA schemes that can meet the hybrid requirements of VR users.

C. Smart Radio and Reconfigurable Intelligent Surfaces

Reconfigurable intelligent surface (RIS) has garnered significant recognition as a revolutionary technique for 6G. By incorporating multiple passive reconfigurable elements with tunable amplitudes and phases, RIS is capable of effectively manipulating the direction and strength of the scattered signals, thereby exerting a transformative impact on the propagation environment. Thanks to its powerful capability for controlling wireless propagation environments and manipulating signals spatially, considerable research efforts have been dedicated to RIS-assisted MA schemes.

RIS was initially explored in OMA for point-to-point communication using one or more resource blocks [328]–[333]. This was later expanded to multi-user scenarios, where RIS assists in SDMA [334]–[338], NOMA [21], [339], and RSMA [340]–[349]. Indeed, RIS can significantly improve the performance of these MA schemes, offering benefits like signal strength enhancement, multi-user interference reduction, and improvements in spectral and energy efficiencies, as well as coverage extension. It also reduces hardware complexities in various MA schemes, enabling a reduction in the number of antennas at transceivers without affecting system performance. Moreover, for RSMA, integrating RIS empowers the use of simplified schemes, i.e., 1-layer RS, to achieve performance comparable to HRS/GRS without RIS [350]. RIS contributes to reduce the receiver complexity of RSMA. Nevertheless, the integration of RIS also brings unique research challenges that must be tackled across different MA schemes.

In multi-user scenarios, unlike RIS-assisted OMA, maximizing the channel gain is not the only objective due to the introduction of multi-user interference. This brings extra research challenges regarding beamforming design and channel estimation. To achieve objectives such as maximizing spectral or energy efficiency or minimizing energy consumption, transceiver design in RIS-assisted SDMA/NOMA/RSMA typically involves the joint optimization of passive beamforming at the RIS and active beamforming at the transmitter. These require tailored optimization methods for continuous or discrete diagonal-RIS (D-RIS) phase shifts [334], [342], [346], or for beyond diagonal-RIS (BD-RIS) [335]–[337], [348], [349]. However, this task becomes more complex for NOMA. RIS introduces additional resource allocation challenges to NOMA since the optimal SIC decoding order and user grouping depend on both active beamforming at the transmitter and passive beamforming at the RIS [21]. In terms for performance comparison, an interesting observation from [339] is that in the presence of RIS, NOMA is not always a better option than OMA. Indeed the minimum power needed to achieve a target rate may be lower for TDMA than NOMA, i.e., NOMA may perform worse than TDMA. This contrasts with the non-RIS deployment where NOMA is superior to TDMA.

Given the absence of an RF chain at RISs, channel estimation has been a widely recognized challenge in RIS-assisted links. Considering the resilience of RSMA to CSI inaccuracies and user mobility as shown in many existing works, RIS-assisted RSMA emerges as a promising new paradigm to effectively address and compensate for the channel estimation limitations inherent in RIS [341]. RIS-assisted RSMA therefore offers a mutually beneficial solution to both RSMA and RIS. This advantage is distinctive to RIS-assisted RSMA, as RSMA uniquely demonstrates robustness to CSI, which is not shared by other conventional MA schemes.

VII. Roadmap to 6G Standardization

3GPP is the standards body that introduced UMTS/W-CDMA for 3G, LTE for 4G and NR for 5G. 3GPP also introduced GSM for 2G even before being called 3GPP. While 6G work has not yet started in 3GPP, it is expected to quickly intensify in the 2nd half of 2025.

At the highest conceptual level, 3GPP has delivered over its history systems based on TDMA (2G GSM), CDMA (3G UMTS) and (O)FDMA (4G LTE and 5G NR) closely following any textbook description of basic MA schemes. In this section we look ahead into what 3GPP may consider for MA for 6G by looking back first into previous generations.

The work of 3GPP is structured in so-called Releases, the first version of the 3G specifications was introduced in Release 99 towards the end of last millennium. Release 4 was the subsequent Release after which Releases have been labeled with increasing index. For example, the first version of 4G LTE was introduced in Release 8 while the first version of 5G NR was introduced in Release 15. Release 18, the first Release of 5G-Advanced, was finalized late 2023 and Release 19 started in early 2024. The study phase for 6G is expected to launch in Release 20 in the second half of 2025. In turn, the first 6G specifications are expected in Release 21 with a completion date anticipated to be between the end of 2028 and the end of 2029 aiming at commercial 6G deployments sometime around 2030.

MA is one of the main characteristics of any wireless system. Indeed, it is one of the first decisions made when the design of a new generation takes place as it relates closely to the underlying waveform. It is important to note that, typically, there is one main MA scheme which is supplemented by other schemes for certain specific functions as we will discuss shortly.

A. MA in 3G and 4G

In 3G systems, the MA scheme in the downlink is based on CDMA with Hadamard or Walsh codes applied to coded modulation symbols and orthogonally separating the various channels in a given cell. Different codes can be used to transmit data of different users and multiple codes can be used to increase the data rate for a given user. The application of these codes effectively expands the transmission bandwidth by a factor called spreading factor. The spreading factor offers the ability to boost the SINR at the receiver helping to cope with multiple sources of interference, i.e., inter-cell, intra-cell. However, the spreading factor decreased over time as the target operating data rates increased. This trend increased the need to use equalizers at the receiver, which, in turn, made OFDM based transmissions more appealing as they would not require equalization at the receiver. The application of a cell-dependent scrambling based on long pseudo-noise (PN) sequences onto the aggregated transmissions of a given cell makes it appear as noise when received at neighboring cells in full frequency reuse deployments.

For 4G, OFDMA was introduced in the downlink maintaining intra-cell orthogonality while being more resilient to the inter-symbol interference (ISI) caused by the delay spread of the channel. In full frequency reuse deployments, transmissions from different cells are randomized through the application of a cell-dependent scrambling based on long PN sequences similar to 3G.

In the transition from 3G to 4G, for the uplink we went from a non-orthogonal MA with CDMA in 3G to orthogonal MA with DFT-spread OFDM (DFT-S-OFDM) in 4G. Transitions are typically somewhat blurred with inroads into new schemes by way of extensions of the current scheme before an entirely new scheme is introduced. For example, in the second half of 3G days we experienced the migration from circuit switch (CS) to packet switch (PS) communications. With that migration, we also moved from continuous downlink and uplink transmissions to packetized transmissions which would need to be scheduled by the network. Therefore, Release 5 introduced scheduled downlink with much lower multiplexing capability in the code domain (lower spreading factor) and aiming more at time-domain-multiplexing (TDM'ing) transmissions to multiple users with the goal to increase the instantaneous individual user data rate with the so-called high speed downlink packet access (HSDPA). Similarly, Release 6 introduced at the time scheduled UL with what was called enhanced uplink (eUL) or high speed uplink packet access (HSUPA). Release 7, in turn, introduced discontinuous uplink transmissions which, for the first time, enabled uplink transmissions to be gated in time and hence favoring TDM'ing of transmissions from different users supplemented with the possibility of non-orthogonally multiplexing multiple users' transmissions in the code domain (CDMA).

The main driver for the switch of MA from 3G to 4G was the uplink performance for cell-edge users for which the introduction of frequency domain multiplexing (FDM'ing) brought important performance gains in terms of uplink user throughput [351]. The realization of FDM in the uplink of 4G was done via the introduction of DFT-S-OFDM, which due to its single carrier properties was key to reduce the user transmissions' peak-to-average-power-ratio (PAPR) which directly entails a coverage gain for the uplink. Also, because of its underlying block-based construction of OFDM, subframe transmissions of 1ms simply meant transmissions of 14 symbols facilitating TDM'ing of transmissions with a built-in time-domain gating function which did not exist in 3G days.

While FDMA/TDMA was the main MA scheme for the uplink enabling orthogonal access for users with the same serving cell, it is not the only MA scheme used in 4G LTE.

Indeed, random access is performed via the Physical Random Access Channel (PRACH) which, while orthogonalized in time and frequency with the other uplink Physical Layer (PHY) channels, PRACH transmissions from different users overlap with each other yielding a CDMA-like MA for this physical channel with possibility of recoverable or unrecoverable inter-user collisions.

Furthermore, SDMA was envisioned to enable MU-MIMO operation from the start of 4G. In this scenario, the spatial dimensionality of the channel is exploited by enabling data transmissions from different users to overlap in time and frequency with each other relying on spatially separating them at the multi-antenna base-station receiver. As a result, while the data portion of the uplink transmission is merely spatial-division-multiplexed (SDM'd) across users, their reference signal transmissions are orthogonalized by virtue of different cyclic shifts of the underlying Zadoff-Chu sequences used to modulate the demodulation reference signals (DMRS). In turn, that orthogonalization enables the possibility to perform channel estimation for each user's transmission off cleaner DMRS samples without inter-user interference.

Moreover, uplink control channel transmissions typically bear few bits of information which attempting to orthogonalize solely in the frequency domain would very rapidly run out of dimensions in the frequency domain (resource limitation). In order to alleviate that limitation, the possibility to orthogonally multiplex within the same time/frequency resources via different cyclic shifts of underlying Zadoff-Chu sequences (similar to DMRS) or time-domain orthogonal covers applied to each uplink symbol was also introduced to increase the capacity of the uplink control channel. Release 10 introduced the possibility for SU-MIMO in the uplink of LTE still resorting to DFT-S-OFDM for each of the transmitted layers.

Towards the latter part of LTE, multi-user superposition transmission (MUST) was standardized for the downlink of LTE in Release 14 after the conclusion of the corresponding study item in the preceding release. The project attempted to jointly optimize multi-user operation from both transmitter and receiver's perspective to improve the MU system capacity even if the transmission/precoding was non-orthogonal. The outcome of the study is captured in the technical report (TR) for MUST in [9]. Three MUST categories were identified during the study: 1) MUST Category 1: Superposition transmission with adaptive power ratio on component constellations and non-Gray-mapped composite constellation; 2) MUST Category 2: Superposition transmission with adaptive power ratio on component constellations and Gray-mapped composite constellation; MUST Category 3: Superposition transmission with label-bit assignment on Gray-mapped composite constellation.

B. MA in 5G and 6G

With the exception of MUST, all the schemes mentioned for 4G LTE were also adopted for 5G NR. Indeed, for the uplink data channel, the earlier mentioned DFT-S-OFDM is only used for single layer transmissions, while multi-layer transmissions always resort to OFDMA to improve link efficiency at the cost of an increased PAPR. The DMRS of the uplink data channel for different layers of the same user and from different users in the same cell are orthogonalized in time/frequency domain by corresponding orthogonal covers.

5G NR while introduced in Release 15, had a project in Release 16 to study NOMA [352] as a follow up to the discussions during the 5G NR study of Release 14 [353]. The TR for NOMA is a good reference [8] to review the schemes that were considered during the study with some of them presented in this paper. The Release 16 NOMA study investigated the application of NOMA techniques to mMTC, URLLC, and eMBB communications.

The uplink NOMA transmitter was generalized so that it would include all the different flavors being considered. Also, different receivers were considered tailored to the different schemes and incurring various levels of complexity.

The outcome of the Release 16 study was deemed to be non-conclusive and a follow up work item was not approved at the time. Part of the reason was the proliferation of schemes during the study without a clear winner amongst them which made it impossible to define the scope of a normative project for only one scheme. Perhaps the scope of the project was too broad aiming at NOMA for mMTC, URLLC and eMBB use cases. Note that the mMTC use case of IMT-2020 is actually served by LTE CatM (eMTC) and NBIOT originally developed in Release 13 and evolved thereafter.

Indeed, it was discussed that NOMA could be enabled in NR without specification impact provided that there was adequate receiver processing. The possibility to expand the number of orthogonal DMRS ports was discussed which, in combination with multi-user detectors at the receiver, would enable the possibility for processing fully overlapping data transmissions from different users. Those transmissions from multiple users could be resulting from dynamic scheduling from the network (with individual grants to each of the users) or from overlapping configured grants (pre-configured via L2 or L3 allocations).

In the absence of normative projects for NOMA specification, a minor improvement to the Random Access procedure which used to take 4-steps was approved with the goal to shorten it to 2-steps [354]. It is, however, likely that the investigations that took part during Release 16 will be revisited when NOMA is considered again in 3GPP in the future.

Clearly, there is no universal MA (yet) that optimally suits all applications and use cases. The MA scheme aiming at optimizing individual user's high throughput operation (eMBB-like use case) is not the same as the one aiming at maximizing multiplexing capability for low-rate applications of many, many users (mMTC-like use case). To some extent, that tradeoff is already seen by looking into the MA scheme used today for random access, uplink control channels, uplink data channels, etc.

6G will see a resurgence of MA discussions. Though traditionally there is one main MA scheme supplemented by other schemes for certain specific functions, it is unclear whether this approach will still hold for 6G with the proliferation of schemes to cope with eMBB and mMTC but also and mainly to enable new use cases fueled by emerging intelligent and multi-functional applications. Aside the ones already discussed as part of 5G, more recent MA schemes have made their way toward standardization. This is the case of RSMA whose "universal" capability has been found attractive by various industries and recently been proposed for the first time for 6G for various use cases such as 1) to boost the performance of unicast transmissions beyond what has been achieved traditionally in 4G and 5G by SDMA-multi-user MIMO [355], and 2) to enable non-orthogonal unicast and multicast transmissions where multicast traffic are superimposed over unicast traffic instead of being served in orthogonal resources as in 4G and 5G [43]. Though from a theoretical point of view, a unifying MA scheme like RSMA provides performance benefits (as discussed in Section II-E), it will be interesting to see in the coming years whether 3GPP will favour a single unified and general MA scheme as main MA scheme instead of a combination of multiple MA schemes, each optimized for specific conditions.

C. AI/ML in 5G and 6G

A Study Item for AI/ML applied to the NR Air Interface was agreed for Release 18 [356]. The TR for this study is available at [357] and constitutes a good reference to review what is being studied (project completion planned for December 2023). While AI/ML first appeared in 3GPP for the purpose of users' data collection, this project is the first time in 3GPP where the direct application of AI/ML is sought for some fundamental air interface problems. In order to avoid too abstract, high level conceptual discussions and to give some shape to the project, a number of pilot use cases was identified to be analyzed for this project. Namely, CSI enhancements, beam management (BM) and positioning accuracy enhancements. Each of those three use cases was further refined into two sub-use cases. Namely, CSI (spatial-frequency) compression and CSI (temporal) prediction for CSI enhancement, Spatial beam prediction and Temporal beam prediction for BM, direct AI/ML positioning and AI/ML assisted positioning for Positioning accuracy enhancements. All sub use cases but CSI compression rely on single-sided models at the user terminal or the network, while CSI compression resorts to two-sided AI/ML models with one part of the model running at the user terminal side and the other at the network side. For the CSI compression sub use case, the user terminal side performs the encoding or compression of the CSI information, and the network side performs the decoder or decompression to obtain the terminal's CSI.

It is expected that normative work to address the specification impact of these use cases will be carried out as part of Release 19. Moreover, further use cases, for example Mobility enhancements, may be additionally investigated.

As indicated by the ITU-R WP 5D framework document for IMT-2030 [5], Integrated AI and communication has been identified as one of the usage scenarios for IMT-2030. As a result, it is expected that 6G will be AI/ML native.

While nobody knows precisely what that will imply from 3GPP perspective, the foundational work on AI/ML for NR air interface that is underway will for sure play a very important role on how native 6G will be specified [358] [359] [360] [361] [362].

The application of AI/ML techniques to MA may be investigated for 6G with the aim at figuring out the optimal MA scheme for the given application and use case. Indeed, channel access policies that fluently transition from contention- to schedule-based depending on the use case and environment are deemed desirable [360].

Similar to the Release 18 study where single-sided and two-sided (or cross-node) models are being considered depending on the use case, different AI/ML functions for 6G will require one or another approach. In particular, how the AI/ML engines running at the terminals and the network will figure out the optimal MA scheme is subject of research and will for sure open up new opportunities for improved system performance and user experience. Adaptability of the models through further offline or even online training will also play a key role in the end-to-end performance optimization.

As a result, the current investigations on joint training of two-sided models, as well as, their interoperability and performance testing will be very relevant for the future native AI work in 6G.

VIII. Conclusions

Wireless networks experience a paradigm shift with the integration of multi-functionality and the rise of AI and ML. 6G and beyond will be AI-native and multi-functionality-native, spurring the need to re-think multiple access techniques to make use of the available time, frequency, space, power, signal dimensions to serve users, devices, machines, services, training nodes in the most efficient way and cope with those new network advances. The paper highlighted that much work is left for researchers in this area. Aside many other schemes, RSMA has emerged in the past few years by providing a unified and conceptually simple understanding of SDMA, NOMA, physical layer multicasting, and uniquely shrinks the knowledge tree of MA schemes based on space, power, signal dimensions. Consequently RSMA has found numerous applications for 6G.

Throughout the rich literature on MA schemes, reported gains of RSMA over SDMA and NOMA range between a few % to several hundreds of %. They depend on many factors such as SNR, disparity of channel strengths, angle between user channel directions, objective function (e.g. sum-rate, weighted sum-rate, minimum rate), QoS rate constraints, number of antennas at the nodes, number of users, underloaded vs overloaded, CSI quality at the transmitters and the receivers, presence of impairments (phase noise, mobility and latency, subband feedback, etc), traffic types (unicast, multicast, mixed unicast/multicast), Gaussian signaling vs finite constellations, infinite vs finite block lengths, user deployment, and applications. Unsurprisingly, the scenario that leads to the smallest gains for RSMA over SDMA is sum-rate maximization in massively underloaded (many more antennas at the base station than the number of transmitted streams) deployments with accurate CSIT, Gaussian signaling, and infinite block lengths. As we depart from this setting and increase the number of users and streams/decrease the number of transmit antennas, increase the interference level, bring in more fairness in the objective function, impose a QoS rate constraint at each user, and/or account for CSIT imperfections/impairments/finite constellations/finite block lengths, gains of RSMA over SDMA increase.

What is nevertheless missing in the literature is a consistent performance evaluation methodology that enables a comprehensive performance comparison of MA schemes across all dimensions. This is crucial to obtain a holistic understanding of the performance benefits of MA schemes and the role played by the various MA dimensions for both uplink and downlink.

Moving next, research should move toward the direction of the design of UMA. UMA should further shrink the knowledge tree of MA schemes by unifying RSMA with all other dimensions, such as code domain MAs, and ultimately provide a unified and conceptually simple understanding of the current and future morass of MA schemes. Such UMA does not exist yet. It is hoped that the techniques and outlook presented in this article will help inspiring future research in this exciting and important area and pave the way for designing and implementing efficient MA techniques in future wireless systems.

IX. Acknowledgments

Bruno Clerckx would like to thank Rafael Cerna Loli, Longfei Yin, Sibo Zhang, and Xinze Lyu for providing RSMA simulation and experimental results.

References

[1] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Network*, vol. 34, no. 3, pp. 134–142, May 2020.

[2] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, “Fundamentals of wireless information and power transfer: From rf energy harvester models to signal and system designs,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 4–33, 2019.

[3] C. De Lima, D. Belot, R. Berkvens, A. Bourdoux, D. Dardari, M. Guillaud, M. Isomursu, E.-S. Lohan, Y. Miao, A. N. Barreto, M. R. K. Aziz, J. Saloranta, T. Sanguanpuak, H. Sarieddeen, G. Seco-Granados, J. Suutala, T. Svensson, M. Valkama, B. Van Liempd, and H. Wymeersch, “Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges,” *IEEE Access*, vol. 9, pp. 26 902–26 925, 2021.

[4] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, August 2019.

[5] W. P. 5D, “The ITU-R (framework) for IMT-2030,” International Telecommunication Union (ITU), Tech. Rep., July 2023.

[6] B. Clerckx, K. Huang, L. R. Varshney, S. Ulukus, and M.-S. Alouini, “Wireless power transfer for future networks: Signal processing, machine learning, computing, and sensing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 5, pp. 1060–1094, 2021.

[7] B. Clerckx, J. Kim, K. W. Choi, and D. I. Kim, “Foundations of wireless information and power transfer: Theory, prototypes, and experiments,” *Proceedings of the IEEE*, vol. 110, no. 1, pp. 8–30, 2022.

[8] 3GPP, “Study on non-orthogonal multiple access (NOMA) for NR,” 3rd Generation Partnership Project (3GPP), Technical Report TR 38.812 V16.0.0, Dec. 2018.

[9] ——, “Study on Downlink Multiuser Superposition Transmission (MUST) for LTE,” 3rd Generation Partnership Project (3GPP), Technical Report TR 36.859 V13.0.0, Dec. 2015.

[10] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, “A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1265–1308, 2023.

[11] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, “Is NOMA efficient in multi-antenna networks? a critical look at next generation multiple access techniques,” *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1310–1343, 2021.

[12] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, “Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, Mar. 2020.

[13] D. Gesbert, M. Kountouris, R. W. Heath, C.-b. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[14] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[15] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: fundamentals, survey, and future research trends," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 2073–2126, Jul. 2022.

[16] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2008.

[17] J. Park, B. Lee, J. Choi, H. Lee, N. Lee, S.-H. Park, K.-J. Lee, J. Choi, S. H. Chae, S.-W. Jeon, K. S. Kwak, B. Clerckx, and W. Shin, "Rate-splitting multiple access for 6G networks: Ten promising scenarios and applications," *IEEE Network*, pp. 1–1, 2023.

[18] A. Mishra, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G—part I: Principles, applications and future works," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2232–2236, 2022.

[19] L. Yin, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G—part II: Interplay with integrated sensing and communications," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2237–2241, 2022.

[20] H. Li, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G—part III: Interplay with reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2242–2246, 2022.

[21] Z. Ding, L. Lv, F. Fang, O. A. Dobre, G. K. Karagiannidis, N. Al-Dhahir, R. Schober, and H. V. Poor, "A state-of-the-art survey on reconfigurable intelligent surface-assisted non-orthogonal multiple access networks," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1358–1379, 2022.

[22] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, 2022.

[23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[24] "Second generation framing structure, channel coding and modulation systems for broadcasting interative services, news gathering and other broadband satellite applications; part 2: DVB-S2 extensions (DVBS2x), document ETSI En 302-307-2 v1.1.1, European Broadcasting Union (EBU), Grand-Saconnex, Switzerland, oct. 2014."

[25] L. Yin and B. Clerckx, "Rate-splitting multiple access for multigroup multicast and multibeam satellite systems," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 976–990, 2021.

[26] B. Clerckx and C. Oestges, *MIMO wireless networks: Channels, techniques and standards for multi-antenna, multi-user and multi-cell systems*. Academic Press, 2013.

[27] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[28] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[29] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[30] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[31] Y. Mao and B. Clerckx, "Beyond dirty paper coding for multi-antenna broadcast channel with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6775–6791, Nov. 2020.

[32] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (noma) for cellular future radio access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.

[33] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer, 2019, vol. 159.

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

[35] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[36] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.

[37] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, 2016.

[38] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 133, May 2018.

[39] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996.

[40] A. R. Flores, R. C. De Lamare, and B. Clerckx, "Tomlinson-Harashima precoded rate-splitting with stream combiners for MU-MIMO systems," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 3833–3845, 2021.

[41] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3232–3246, 2015.

[42] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, 2016.

[43] BBC, "Rws-230044 - RSMA for MU-MIMO – enhancement of unicast and joint unicast/multicast delivery," *3GPP Rel. 19 Workshop*, Taiwan, June 2023.

[44] "Open ran in high demand density environments technical guidance," *Department for Science and Innovation and Technology, UK*, Updated 6 April 2023.

[45] S. Zhang, B. Clerckx, D. Vargas, O. Haffenden, and A. Murphy, "Rate-splitting multiple access: Finite constellations, receiver design, and sic-free implementation," *arXiv:2305.17178*, 2023.

[46] X. Lyu, S. Aditya, J. Kim, and B. Clerckx, "Rate-splitting multiple access: The first prototype and experimental validation of its superiority over SDMA and NOMA," *arXiv 2305.07361*, 2023.

[47] H. Joudeh and B. Clerckx, "Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7276–7289, 2017.

[48] Y. Mao, E. Piovano, and B. Clerckx, "Rate-splitting multiple access for overloaded cellular internet of things," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4504–4519, 2021.

[49] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8754–8770, 2019.

[50] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6227–6242, 2016.

[51] C. Hao and B. Clerckx, "MISO networks with imperfect CSIT: A topological rate-splitting approach," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2164–2179, 2017.

[52] E. Piovano and B. Clerckx, "Optimal dof region of the k-user MISO BC with partial CSIT," *IEEE Communications Letters*, vol. 21, no. 11, pp. 2368–2371, 2017.

[53] C. Hao, B. Rassouli, and B. Clerckx, "Achievable dof regions of MIMO networks with imperfect CSIT," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6587–6606, 2017.

[54] S. Park, J. Choi, J. Park, W. Shin, and B. Clerckx, "Rate-splitting multiple access for quantized multiuser MIMO communications," *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 7696–7711, 2023.

[55] A. Mishra, Y. Mao, L. Sanguinetti, and B. Clerckx, "Rate-splitting assisted massive machine-type communications in cell-free massive MIMO," *IEEE Communications Letters*, vol. 26, no. 6, pp. 1358–1362, 2022.

[56] A. Mishra, Y. Mao, C. K. Thomas, L. Sanguinetti, and B. Clerckx, "Mitigating intra-cell pilot contamination in massive MIMO: A rate splitting approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3472–3487, 2023.

[57] O. Dizdar, Y. Mao, and B. Clerckx, "Rate-splitting multiple access to mitigate the curse of mobility in (massive) MIMO networks," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6765–6780, 2021.

[58] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO

systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8196–8211, 2017.

[59] A. R. Flores, R. C. de Lamare, and B. Clerckx, "Linear precoding and stream combining for rate splitting in multiuser MIMO systems," *IEEE Communications Letters*, vol. 24, no. 4, pp. 890–894, 2020.

[60] A. Mishra, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for downlink multiuser MIMO: Precoder optimization and phy-layer design," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 874–890, 2022.

[61] H. Chen, D. Mi, B. Clerckx, Z. Chu, J. Shi, and P. Xiao, "Joint power and subcarrier allocation optimization for multigroup multicast systems with rate splitting," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2306–2310, 2020.

[62] A. Z. Yalcin, M. Yuksel, and B. Clerckx, "Rate splitting for multi-group multicasting with a common message," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12 281–12 285, 2020.

[63] L. Yin and B. Clerckx, "Rate-splitting multiple access for satellite-terrestrial integrated networks: Benefits of coordination and cooperation," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 317–332, 2023.

[64] Z. W. Si, L. Yin, and B. Clerckx, "Rate-splitting multiple access for multigateway multibeam satellite systems with feeder link interference," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 2147–2162, 2022.

[65] Y. Liu, B. Clerckx, and P. Popovski, "Network slicing for eMBB, URLLC, and mMTC: An uplink rate-splitting multiple access approach," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[66] G. Zhou, Y. Mao, and B. Clerckx, "Rate-splitting multiple access for multi-antenna downlink communication systems: Spectral and energy efficiency tradeoff," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4816–4828, July 2022.

[67] J. Zhang, B. Clerckx, J. Ge, and Y. Mao, "Cooperative rate splitting for MISO broadcast channel with user relaying, and performance benefits over cooperative NOMA," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1678–1682, Nov. 2019.

[68] Y. Mao, B. Clerckx, J. Zhang, V. O. K. Li, and M. A. Arafah, "Max-min fairness of K-user cooperative rate-splitting in MISO broadcast channel with user relaying," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6362–6376, Oct. 2020.

[69] H. Cho, B. Ko, B. Clerckx, and J. Choi, "Coverage increase at thz frequencies: A cooperative rate-splitting approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9821–9834, 2023.

[70] Y. Xu, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access with finite blocklength for short-packet and low-latency downlink communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 333–12 337, 2022.

[71] ——, "Max-min fairness of rate-splitting multiple access with finite blocklength communications," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6816–6821, 2023.

[72] J. Xu, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for short-packet uplink communications: A finite blocklength analysis," *IEEE Commun. Lett.*, pp. 1–1, 2022.

[73] A. Salem, C. Masouros, and B. Clerckx, "Secure rate splitting multiple access: How much of the split signal to reveal?" *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 4173–4187, 2023.

[74] H. Xia, Y. Mao, X. Zhou, B. Clerckx, S. Han, and C. Li, "Weighted sum-rate maximization for rate-splitting multiple access based multi-antenna broadcast channel with confidential messages," *arXiv 2202.07328*, 2023.

[75] Z. Liu and L.-L. Yang, "Sparse or dense: A comparative study of code-domain NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4768–4780, 2021.

[76] B. M. Zaidel, O. Shental, and S. Shamai Shitz, "Regular sparse NOMA: Ultimate performance in closed form," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 5–20, Mar. 2022.

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

[78] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[79] D. Guo and C.-c. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 421–431, 2008.

[80] J. van de Beek and B. M. Popovic, "Multiple access with low-density signatures," in *IEEE Global Telecommunications Conference*, 2009, pp. 1–6.

[81] R. Razavi, R. Hoshyar, M. A. Imran, and Y. Wang, "Information theoretic analysis of LDS scheme," *IEEE Communications Letters*, vol. 15, no. 8, pp. 798–800, 2011.

[82] R. Hoshyar, R. Razavi, and M. Al-Imari, "LDS-OFDM an efficient multiple access technique," in *IEEE Vehicular Technology Conference*, 2010, pp. 1–5.

[83] M. AL-Imari, M. A. Imran, and R. Tafazolli, "Low density spreading for next generation multicarrier cellular systems," in *International Conference on Future Communication Networks (ICFCN)*, 2012, pp. 52–57.

[84] M. AL-Imari, M. A. Imran, R. Tafazolli, and D. Chen, "Performance evaluation of low density spreading multiple access," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2012, pp. 383–388.

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

[86] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both rayleigh fading and Gaussian channels," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 502–518, 1996.

[87] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *IEEE Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5.

[88] M. Vameghestahbanati, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Multidimensional constellations for uplink SCMA systems—a comparative study," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2169–2194, 2019.

[89] S. Chaturvedi, Z. Liu, V. A. Bohara, A. Srivastava, and P. Xiao, "A tutorial on decoding techniques of sparse code multiple access," *IEEE Access*, vol. 10, pp. 58 503–58 524, 2022.

[90] L. Yu, Z. Liu, M. Wen, D. Cai, S. Dang, Y. Wang, and P. Xiao, "Sparse code multiple access for 6G wireless communication networks: Recent advances and future directions," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 92–99, 2021.

[91] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *IEEE Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5.

[92] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[93] X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, and J. Xu, "Successive interference cancelation amenable multiple access (SAMA) for future wireless communications," in *IEEE International Conference on Communication Systems*, 2014, pp. 222–226.

[94] L. Ma, S. Tong, H. Zheng, B. Bai, and X. Dai, "Edgewise serial message passing detection of uplink SCMA systems for better user fairness and faster convergence rate," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1285–1288, 2019.

[95] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave-division multiple-access (IDMA) communications," in *Proc. 3rd International Symposium on Turbo Codes & Related Topics*, 2003, pp. 173–180.

[96] L. Ping, "Interleave-division multiple access and chip-by-chip iterative multi-user detection," *IEEE Communications Magazine*, vol. 43, no. 6, pp. S19–S23, 2005.

[97] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave division multiple-access," *IEEE Transactions on Wireless Communications*, vol. 5, no. 4, pp. 938–947, 2006.

[98] P. Wang, J. Xiao, and L. Ping, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 4–11, 2006.

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

[100] L. Ping, Q. Guo, and J. Tong, "The OFDM-IDMA approach to wireless communication systems," *IEEE Wireless Communications*, vol. 14, no. 3, pp. 18–24, 2007.

[101] C. Xu, Y. Hu, C. Liang, J. Ma, and L. Ping, "Massive MIMO, non-orthogonal multiple access and interleave division multiple access," *IEEE Access*, vol. 5, pp. 14 728–14 748, 2017.

[102] S. Kang, X. Dai, and B. Ren, "Pattern division multiple access for 5G," *Telecommun. Netw. Technol.*, vol. 5, no. 5, pp. 43–47, 2015.

[103] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1046–1061, 1999.

[104] 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 Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[105] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, "Pattern division multiple access: A new multiple access technology for 5G," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 54–60, 2018.

[106] X. Dai, T. Yan, Q. Li, H. Li, and X. Wang, "Pattern division random access (pdra) for m2m communications with massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 631–12 639, 2021.

[107] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[108] J. Zhu and M. Gastpar, "Gaussian multiple access via compute-and-forward," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2678–2695, 2017.

[109] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[110] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.

[111] E. Sula, J. Zhu, A. Pastore, S. H. Lim, and M. Gastpar, "Compute–forward multiple access (CFMA): Practical implementations," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1133–1147, 2019.

[112] M. Qiu, Y.-C. Huang, S.-L. Shieh, and J. Yuan, "A lattice-partition framework of downlink non-orthogonal multiple access without SIC," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2532–2546, Jun. 2018.

[113] M. Qiu, L. Yang, Y. Xie, and J. Yuan, "On the design of multi-dimensional irregular repeat-accumulate lattice codes," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 478–492, 2018.

[114] M. Qiu, Y.-C. Huang, J. Yuan, and C.-L. Wang, "Lattice-partition-based downlink non-orthogonal multiple access without SIC for slow fading channels," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1166–1181, Feb. 2019.

[115] M. Qiu, Y.-C. Huang, and J. Yuan, "Downlink non-orthogonal multiple access without SIC for block fading channels: An algebraic rotation approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3903–3918, Aug. 2019.

[116] ——, "Discrete signaling and treating interference as noise for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7253–7284, Nov. 2021.

[117] ——, "Downlink transmission with heterogeneous URLLC services: Discrete signaling with single-user decoding," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 7, pp. 2261–2277, 2023.

[118] A. Jimenez Felstrom and K. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2181–2191, 1999.

[119] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5274–5289, 2010.

[120] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 803–834, 2011.

[121] S. Moloudi, M. Lentmaier, and A. Graell i Amat, "Spatially coupled turbo-like codes," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6199–6215, 2017.

[122] M. Qiu, X. Wu, J. Yuan, and A. Graell i Amat, "Generalized spatially-coupled parallel concatenated codes with partial repetition," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 5771–5787, 2022.

[123] D. Truhachev, "Universal multiple access via spatially coupling data transmission," in *IEEE International Symposium on Information Theory*, 2013, pp. 1884–1888.

[124] D. Truhachev and C. Schlegel, "Coupling data transmission for multiple-access communications," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4550–4574, 2019.

[125] A. Karami and D. Truhachev, "Coupling information transmission with window decoding," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 560–563, 2019.

[126] ——, "Self-coupling data transmission for random multiple access communications," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1168–1172, 2020.

[127] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-division-multiplexing: Theory and practice," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216–232, 2016.

[128] L. Zhang, Y. Wu, W. Li, B. Rong, K. Salehian, S. Lafleche, X. Wang, S. I. Park, H. M. Kim, J.-y. Lee, N. Hur, P. Angueira, and J. Montalban, "Layered-division-multiplexing for high spectrum efficiency and service flexibility in next generation ATSC 3.0 broadcast system," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 116–123, 2019.

[129] L. Zhang, Y. Wu, W. Li, K. Salehian, S. Lafleche, X. Wang, S. I. Park, H. M. Kim, J.-y. Lee, N. Hur, P. Angueira, and J. Montalban, "Layered-division multiplexing: An enabling technology for multicast/broadcast service delivery in 5G," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 82–90, 2018.

[130] E. Basar, u. Aygulu, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[131] Y.-T. Lai, Y.-R. Ciou, and J.-M. Wu, "Index modulation multiple access," in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2018, pp. 1–5.

[132] J. Li, S. Dang, M. Wen, Q. Li, Y. Chen, Y. Huang, and W. Shang, "Index modulation multiple access for 6G communications: Principles, applications, and challenges," *IEEE Network*, vol. 37, no. 1, pp. 52–60, 2023.

[133] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[134] Z. Wei, W. Yuan, S. Li, J. Yuan, G. Bharatula, R. Hadani, and L. Hanzo, "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 136–144, 2021.

[135] Y. Hong, T. Thaj, and E. Viterbo, *Delay-Doppler Communicationss: Principles and Applications*. London, U.K.: Elsevier Academic Press, 2022.

[136] H. Lin and J. Yuan, "Orthogonal delay-doppler division multiplexing modulation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11 024–11 037, 2022.

[137] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4801–4815, 2022.

[138] Z. Wu and L. Dai, "Multiple access for near-field communications: Sdma or ldma?" *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 6, pp. 1918–1935, 2023.

[139] P. Zhang, X. Xu, C. Dong, K. Niu, H. Liang, Z. Liang, X. Qin, M. Sun, H. Chen, N. Ma *et al.*, "Model division multiple access for semantic communications," *Frontiers of Information Technology & Electronic Engineering*, pp. 1–12, 2023.

[140] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 491–506, 2020.

[141] L. Liu, E. G. Larsson, W. Yu, P. Popovsk, C. Stefanovic, and E. D. Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[142] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 148–156, 2020.

[143] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 615–637, 2021.

[144] L. Liu and W. Yu, "Massive connectivity with massive MIMO-Part I: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[145] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2925–2951, 2021.

[146] E. Björnson, E. de Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random access protocol for pilot allocation in crowded massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2220–2234, Apr. 2017.

[147] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance-based massive random access with massive MIMO," *IEEE Transactions on Information Theory*, vol. 68, no. 3, pp. 1696–1715, 2021.

[148] Y. Polyanskiy, "A perspective on massive random-access," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2523–2527.

[149] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *IEEE Transactions on Information Theory*, vol. 67, no. 9, pp. 5853–5884, 2021.

[150] J. Gao, Y. Wu, S. Shao, W. Yang, and H. V. Poor, "Energy efficiency of massive random access in MIMO quasi-static rayleigh fading channels with finite blocklength," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1618–1657, 2023.

[151] A. Glebov, P. Rybin, K. Andreev, and A. Frolov, "Energy efficiency of unsourced random access over the binary-input Gaussian channel," *IEEE Communications Letters*, vol. 27, no. 9, pp. 2313–2317, 2023.

[152] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "LDPC codes with soft interference cancellation for uncoordinated unsourced multiple access," in *IEEE International Conference on Communications*, 2021, pp. 1–6.

[153] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6509–6533, 2020.

[154] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Transactions on Information Theory*, vol. 67, no. 10, pp. 6894–6915, 2021.

[155] A. Vem, K. R. Narayanan, J.-F. Chamberland, and J. Cheng, "A user-independent successive interference cancellation based coding scheme for the unsourced random access Gaussian channel," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8258–8272, 2019.

[156] E. Marshakov, G. Balitskiy, K. Andreev, and A. Frolov, "A polar code based unsourced random access for the Gaussian MAC," in *IEEE Vehicular Technology Conference (VTC-Fall)*, 2019, pp. 1–5.

[157] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[158] M. J. Ahmadi, "Novel unsourced random access algorithms over Gaussian and fading channels," Ph.D. dissertation, Bilkent University, 2024.

[159] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.

[160] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO unsourced random access," 2019.

[161] M. Gkagkos, K. R. Narayanan, J.-F. Chamberland, and C. N. Georghiades, "FASURA: A scheme for quasi-static fading unsourced random access channels," *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6391–6401, 2023.

[162] M. M. Şahin, O. Dizdar, B. Clerckx, and H. Arslan, "Multicarrier rate-splitting multiple access: Superiority of OFDM-RSMA over OFDMA and OFDM-NOMA," *IEEE Communications Letters*, vol. 27, no. 11, pp. 3088–3092, 2023.

[163] M. M. Sahin, O. Dizdar, B. Clerckx, and H. Arslan, "Ofdm-rsma: Robust transmission under inter-carrier interference," *arXiv 2310.01686*, 2023.

[164] C. Xu, B. Clerckx, S. Chen, Y. Mao, and J. Zhang, "Rate-splitting multiple access for multi-antenna joint radar and communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1332–1347, 2021.

[165] M. Moltafet, N. Mokari, M. R. Javan, H. Saeedi, and H. Pishro-Nik, "A new multiple access technique for 5G: Power domain sparse code multiple access (PSMA)," *IEEE Access*, vol. 6, pp. 747–759, 2018.

[166] N. M. Balasubramanya, A. Gupta, and M. Sellathurai, "Combining code-domain and power-domain NOMA for supporting higher number of users," in *IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[167] S. Sharma, K. Deka, V. Bhatia, and A. Gupta, "Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond," *IEEE Communications Letters*, vol. 23, no. 6, pp. 971–974, 2019.

[168] M. T. P. Le, L. Sanguinetti, E. Björnson, and M.-G. D. Benedetto, "Code-domain NOMA in massive MIMO: When is it needed?" *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4709–4723, 2021.

[169] M. Jia, X. Zhang, J. Sun, X. Gu, and Q. Guo, "Intelligent resource management for satellite and terrestrial spectrum shared networking toward b5g," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 54–61, 2020.

[170] M. Hmila, M. Fernández-Veiga, M. Rodríguez-Pérez, and S. Herrería-Alonso, "Energy efficient power and channel allocation in underlay device to multi device communications," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5817–5832, 2019.

[171] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, "Unsupervised machine learning-based user clustering in millimeter-wave-noma systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7425–7440, 2018.

[172] P. M. Santos, B. R. Manoj, M. Sadeghi, and E. G. Larsson, "Universal adversarial attacks on neural networks for power allocation in a massive MIMO system," *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 67–71, 2022.

[173] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438–5453, 2018.

[174] J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, "Deep learning methods for universal MISO beamforming," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1894–1898, 2020.

[175] W. Lee, M. Kim, and D.-H. Cho, "Deep power control: Transmit power control scheme based on convolutional neural network," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1276–1279, 2018.

[176] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1065–1069, 2020.

[177] A. B. M. Adam, L. Lei, S. Chatzinotas, and N. U. R. Junejo, "Deep convolutional self-attention network for energy-efficient power control in NOMA networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5540–5545, 2022.

[178] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5621–5635, 2018.

[179] K. He, X. Chen, Q. Wu, S. Yu, and Z. Zhou, "Graph attention spatial-temporal network with collaborative global-local learning for citywide mobile traffic prediction," *IEEE Transactions on Mobile Computing*, vol. 21, no. 4, pp. 1244–1256, 2022.

[180] R. Li, C. Wang, Z. Zhao, R. Guo, and H. Zhang, "The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility," *IEEE Communications Letters*, vol. 24, no. 9, pp. 2005–2009, 2020.

[181] J.-Y. Xia, S. Li, J.-J. Huang, Z. Yang, I. M. Jaimoukha, and D. Gündüz, "Metalearning-based alternating minimization algorithm for nonconvex optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 5366–5380, 2023.

[182] Y. Lu, L. Liu, J. Panneerselvam, B. Yuan, J. Gu, and N. Antonopoulos, "A GRU-based prediction framework for intelligent resource management at cloud data centres in the age of 5G," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 486–498, 2020.

[183] M. Lee, G. Yu, and G. Y. Li, "Graph embedding-based wireless link scheduling with few training samples," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2282–2294, 2021.

[184] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 101–115, 2021.

[185] M. Lee, S. Hosseinalipour, C. G. Brinton, G. Yu, and H. Dai, "A fast graph neural network-based method for winner determination in multi-unit combinatorial auctions," *IEEE Transactions on Cloud Computing*, vol. 10, no. 4, pp. 2264–2280, 2022.

[186] M. Lee, G. Yu, H. Dai, and G. Y. Li, "Graph neural networks meet wireless communications: Motivation, applications, and future directions," *IEEE Wireless Communications*, vol. 29, no. 5, pp. 12–19, 2022.

[187] Z. Gao, L. Yang, and Y. Dai, "Fast adaptive task offloading and resource allocation via multiagent reinforcement learning in heterogeneous vehicular fog computing," *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 6818–6835, 2023.

[188] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an lstm auto-encoder," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 370–382, 2022.

[189] Y. Yuan, G. Zheng, K.-K. Wong, B. Ottersten, and Z.-Q. Luo, "Transfer learning and meta learning-based fast downlink beamforming adaptation," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1742–1755, 2021.

[190] R. Cerna Loli and B. Clerckx, "A meta-learning-based precoder optimization framework for rate-splitting multiple access," *IEEE Wireless Communications Letters*, vol. 13, no. 2, pp. 347–351, 2024.

[191] ——, "Meta-learning based optimization for large scale wireless networks: Framework and applications," *to be submitted*.

[192] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, "Iterative algorithm induced deep-unfolding neural networks: Precoding design for multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1394–1410, 2021.

[193] C. Xu, Y. Jia, S. He, Y. Huang, and D. Niyato, "Joint user scheduling, base station clustering and beamforming design based on deep unfolding technique," *IEEE Transactions on Communications*, pp. 1–1, 2023.

[194] M. Lee, G. Yu, and G. Y. Li, "Learning to branch: Accelerating resource allocation in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 958–970, 2020.

[195] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, and L.-C. Wang, "Deep reinforcement learning for mobile 5g and beyond: Fundamentals, applications, and challenges," *IEEE Vehicular Technology Magazine*, vol. 14, no. 2, pp. 44–52, 2019.

[196] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1581–1592, 2020.

[197] C. He, Y. Hu, Y. Chen, and B. Zeng, "Joint power allocation and channel assignment for NOMA with deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2200–2210, 2019.

[198] N. Q. Hieu, D. T. Hoang, D. Niyato, and D. I. Kim, "Optimal power allocation for rate splitting communications with deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2820–2823, 2021.

[199] J. Ji, L. Cai, K. Zhu, and D. Niyato, "Decoupled association with rate splitting multiple access in UAV-assisted cellular networks using multi-agent deep reinforcement learning," *IEEE Transactions on Mobile Computing*, pp. 1–15, 2023.

[200] X. Wang, G. Sun, Y. Xin, T. Liu, and Y. Xu, "Deep transfer reinforcement learning for beamforming and resource allocation in multi-cell MISO-OFDMA systems," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 815–829, 2022.

[201] T. Zhang, S. Wang, Y. Zhuang, C. You, M. Wen, and Y.-C. Wu, "Reconfigurable intelligent surface assisted ofdm relaying: Subcarrier matching with balanced SNR," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 2216–2230, 2023.

[202] J. Zhou, S. Dang, B. Shihada, and M.-S. Alouini, "Power allocation for relayed OFDM with index modulation assisted by artificial neural network," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 373–377, 2021.

[203] J. Feng, S. Han, H. Chen, and C. Yang, "Uplink scheduling for MIMO-OFDMA systems with rate constraints by deep learning," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

[204] Y. Hu, M. Chen, Z. Yang, M. Chen, and G. Jia, "Optimization of resource allocation in multi-cell OFDM systems: A distributed reinforcement learning approach," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–6.

[205] Y. Zhou, Z. M. Fadlullah, B. Mao, and N. Kato, "A deep-learning-based radio resource assignment technique for 5G ultra dense networks," *IEEE Network*, vol. 32, no. 6, pp. 28–34, 2018.

[206] S. Zhang, Z. Ni, L. Kuang, C. Jiang, and X. Zhao, "Load-aware distributed resource allocation for MF-TDMA ad hoc networks: A multi-agent DRL approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 6, pp. 4426–4443, 2022.

[207] M. Chen, W. Saad, and C. Yin, "Echo state networks for self-organizing resource allocation in lte-u with uplink–downlink decoupling," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 3–16, 2017.

[208] J.-B. Wang, J. Wang, Y. Wu, J.-Y. Wang, H. Zhu, M. Lin, and J. Wang, "A machine learning framework for resource allocation assisted by cloud computing," *IEEE Network*, vol. 32, no. 2, pp. 144–151, 2018.

[209] W. V. Mauricio, D. C. Araujo, F. Hugo Neto, F. Rafael Lima, and T. F. Maciel, "A low complexity solution for resource allocation and SDMA grouping in massive MIMO systems," in *15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–6.

[210] A. M. Elbir, "A deep learning framework for hybrid beamforming without instantaneous csi feedback," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 743–11 755, 2020.

[211] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, "A deep learning framework for optimization of MISO downlink beamforming," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1866–1880, 2020.

[212] Z. Zhang, M. Tao, and Y.-F. Liu, "Learning to beamform in joint multicast and unicast transmission with imperfect CSI," *IEEE Transactions on Communications*, vol. 71, no. 5, pp. 2711–2723, 2023.

[213] J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, "A bipartite graph neural network approach for scalable beamforming optimization," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 333–347, 2023.

[214] S. Liu, J. Guo, and C. Yang, "Multidimensional graph neural networks for wireless communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[215] X. Xu, Y. Liu, Q. Chen, X. Mu, and Z. Ding, "Distributed auto-learning gnn for multi-cell cluster-free NOMA communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 4, pp. 1243–1258, 2023.

[216] X. Gao, Y. Liu, X. Liu, and L. Song, "Machine learning empowered resource allocation in IRS aided MISO-NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3478–3492, 2022.

[217] J.-M. Kang, I.-M. Kim, and C.-J. Chun, "Deep learning-based MIMO-NOMA with imperfect SIC decoding," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020.

[218] M. Wu, Z. Wan, Y. Wang, S. Liu, and Z. Gao, "Deep learning-based rate-splitting multiple access for massive MIMO-OFDM systems with imperfect CSIT," in *2022 International Symposium on Wireless Communication Systems (ISWCS)*, 2022, pp. 1–6.

[219] M. R. Camana, C. E. Garcia, and I. Koo, "Deep learning-assisted power minimization in underlay MISO-SWIPT systems based on rate-splitting multiple access," *IEEE Access*, vol. 10, pp. 62 137–62 156, 2022.

[220] N. Q. Hieu, D. T. Hoang, D. Niyato, and D. I. Kim, "Optimal power allocation for rate splitting communications with deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2820–2823, 2021.

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

[222] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.

[223] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[224] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2184–2199, 2019.

[225] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2018.

[226] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[227] M. S. Safari, V. Pourahmadi, and S. Sodagari, "Deep UL2DL: Data-driven channel knowledge transfer from uplink to downlink," *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 29–44, 2020.

[228] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3133–3143, 2020.

[229] B. Ozpoyraz, A. T. Dogukan, Y. Gevez, U. Altun, and E. Basar, "Deep learning-aided 6G wireless networks: A comprehensive survey of revolutionary PHY architectures," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1749–1809, 2022.

[230] E. Bjornson, J. Hoydis, and L. Sanguinetti. Foundations and Trends in Signal Processing, 2017, vol. 11, no. 3-4.

[231] C.-J. Chun, J.-M. Kang, and I.-M. Kim, “Deep learning-based joint pilot design and channel estimation for multiuser MIMO channels,” *IEEE Communications Letters*, vol. 23, no. 11, pp. 1999–2003, 2019.

[232] G. Gui, H. Huang, Y. Song, and H. Sari, “Deep learning for an effective nonorthogonal multiple access scheme,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[233] X. Rao and V. K. N. Lau, “Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, 2014.

[234] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, “Compressed channel feedback for correlated massive MIMO systems,” *Journal of Communications and Networks*, vol. 18, no. 1, pp. 95–104, 2016.

[235] J. Ballé, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image compression,” in *5th International Conference on Learning Representations (ICLR)*, 2017.

[236] S. Lombardo, J. Han, C. Schroers, and S. Mandt, “Deep generative video compression,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[237] C.-K. Wen, W.-T. Shih, and S. Jin, “Deep learning for massive MIMO CSI feedback,” *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

[238] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, “Deep learning-based CSI feedback approach for time-varying massive MIMO channels,” *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 416–419, 2019.

[239] M. B. Mashhadi, Q. Yang, and D. Gündüz, “Distributed deep convolutional compression for massive MIMO CSI feedback,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2621–2633, 2021.

[240] X. Ma, Z. Gao, F. Gao, and M. Di Renzo, “Model-driven deep learning based channel estimation and feedback for millimeter-wave massive hybrid MIMO systems,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2388–2406, 2021.

[241] Z. Yin, W. Xu, R. Xie, S. Zhang, D. W. K. Ng, and X. You, “Deep CSI compression for massive MIMO: A self-information model-driven neural network,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 10, pp. 8872–8886, 2022.

[242] X. Liang, H. Chang, H. Li, X. Gu, and L. Zhang, “Changeable rate and novel quantization for CSI feedback based on deep learning,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10 100–10 114, 2022.

[243] S. Zhang, W. Xu, S. Jin, X. You, D. W. K. Ng, and L.-C. Wang, “Dual-propagation-feature fusion enhanced neural CSI compression for massive MIMO,” *IEEE Transactions on Communications*, vol. 71, no. 9, pp. 5182–5198, 2023.

[244] Y. Yapici, I. Guvenc, and H. Dai, “Low-resolution limited-feedback NOMA for mmWave communications,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5433–5446, 2020.

[245] D. B. Amor, M. Joham, and W. Utschick, “Rate splitting in FDD massive MIMO systems based on the second order statistics of transmission channels,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1351–1365, 2023.

[246] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, “Model-based deep learning: Key approaches and design guidelines,” in *2021 IEEE Data Science and Learning Workshop (DSLW)*, 2021, pp. 1–6.

[247] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, “Deep learning for physical-layer 5g wireless techniques: Opportunities, challenges and solutions,” *IEEE Wireless Communications*, vol. 27, no. 1, pp. 214–222, 2020.

[248] M. A. Aref and S. K. Jayaweera, “Deep learning-aided successive interference cancellation for MIMO-NOMA,” in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–5.

[249] J.-M. Kang, I.-M. Kim, and C.-J. Chun, “Deep learning-based MIMO-NOMA with imperfect SIC decoding,” *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020.

[250] R. C. Loli, O. Dizdar, B. Clerckx, and C. Ling, “Model-based deep learning receiver design for rate-splitting multiple access,” *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[251] Z. Chen and W. Yu, “Sparse activity detection for massive connectivity,” *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[252] J. Wang, Z. Zhang, and L. Hanzo, “Joint active user detection and channel estimation in massive access systems exploiting Reed-Muller sequences,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 739–752, Jun. 2019.

[253] Y. Cui, S. Li, and W. Zhang, “Jointly sparse signal recovery and support recovery via deep learning with applications in MIMO-based grantfree random access,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 788–803, May 2021.

[254] S. Li, W. Zhang, Y. Cui, H. V. Cheng, and W. Yu, “Joint design of measurement matrix and sparse support recovery method via deep autoencoder,” *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1778–1782, Dec. 2019.

[255] X. Wu, S. Zhang, and J. Yan, “A CNN architecture for learning device activity from MMV,” *IEEE Communications Letters*, vol. 25, no. 9, pp. 2933–2937, Sep. 2021.

[256] Y. Cui, S. Li, and W. Zhang, “Jointly sparse signal recovery and support recovery via deep learning with applications in MIMO-based grant-free random access,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 788–803, Mar. 2021.

[257] Y. Bai, W. Chen, B. Ai, Z. Zhong, and I. J. Wassell, “Prior information aided deep learning method for grant-free NOMA in mMTC,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 112–126, Jan. 2022.

[258] J. Johnston and X. Wang, “Model-based deep learning for joint activity detection and channel estimation in massive and sporadic connectivity,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9806–9817, Nov. 2022.

[259] W. Zhu, M. Tao, X. Yuan, and Y. Guan, “DeeplLearned approximate message passing for asynchronous massive connectivity,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5434–5448, Aug. 2021.

[260] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, “Communication efficient federated learning,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 17, Apr. 2021.

[261] J. Chen, X. Pan, R. Monga, S. Bengio, and R. Jozefowicz, “Revisiting distributed synchronous SGD,” *[Online] Available: https://arxiv.org/abs/1604.00981*, 2017.

[262] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, “A joint learning and communications framework for federated learning over wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[263] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, “Energy efficient federated learning over wireless communication networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021.

[264] M. Chen, D. Gunduz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, “Distributed learning in wireless networks: Recent progress and future challenges,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.

[265] R. Balakrishnan, M. Akdeniz, S. Dhakal, and N. Himayat, “Resource management and fairness for federated learning over wireless edge networks,” in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Atlanta, GA, USA, May 2020.

[266] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, “Scheduling policies for federated learning in wireless networks,” *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 317–333, Jan. 2020.

[267] W. Xia, T. Q. S. Quek, K. Guo, W. Wen, H. H. Yang, and H. Zhu, “Multi-armed bandit-based client scheduling for federated learning,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7108–7123, Nov. 2020.

[268] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, “Federated learning over wireless networks: Optimization model design and analysis,” in *Proc. IEEE Conference on Computer Communications*, Paris, France, May 2019.

[269] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, “Adaptive federated learning in resource constrained edge computing systems,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205–1221, June 2019.

[270] C. T. Dinh, N. H. Tran, M. N. H. Nguyen, C. S. Hong, W. Bao, A. Y. Zomaya, and V. Gramoli, “Federated learning over wireless networks: Convergence analysis and resource allocation,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 1, pp. 398–409, Feb. 2021.

[271] Z. Wang, Y. Zhou, Y. Zou, Q. An, Y. Shi, and M. Bennis, “A graph neural network learning approach to optimize ris-assisted federated learning,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 9, pp. 6092–6106, 2023.

[272] J. Mao and A. Yener, “Roar-fed: Ris-assisted over-the-air adaptive resource allocation for federated learning,” *arXiv preprint arXiv:2309.09883*, 2023.

[273] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and ris-assisted massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4255–4268, 2022.

[274] Y. Oh, Y.-S. Jeon, M. Chen, and W. Saad, "Fedvqcs: Federated learning via vector quantized compressed sensing," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[275] H. Sun, X. Ma, and R. Q. Hu, "Adaptive federated learning with gradient compression in uplink NOMA," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16 325–16 329, Dec. 2020.

[276] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 170–185, Jan. 2021.

[277] M. S. Al-Abiad, M. Z. Hassan, and M. J. Hossain, "Energy-efficient resource allocation for federated learning in NOMA-enabled and relay-assisted Internet of Things networks," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 24 736–24 753, Dec. 2022.

[278] N. Führling, K. Ando, H. S. Rou, and G. T. F. D. Abreu, "A rate splitting multiple access interface for clustered wireless federated learning," *IEEE Access*, vol. 11, pp. 82 652–82 664, August 2023.

[279] S.-H. Park and H. Lee, "Completion time minimization of fog-ran-assisted federated learning with rate-splitting transmission," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 10 209–10 214, 2022.

[280] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2755–2770, 2018.

[281] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3929–3944, 2020.

[282] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, J. Morais, U. Demirhan, and N. Srinivas, "Deepsense 6G: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 122–128, 2023.

[283] U. Demirhan and A. Alkhateeb, "Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 2655–2660.

[284] G. Charan, T. Osman, A. Hredzak, N. Thawdar, and A. Alkhateeb, "Vision-position multi-modal beam prediction using real millimeter wave datasets," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 2727–2731.

[285] S. Jiang, G. Charan, and A. Alkhateeb, "Lidar aided future beam prediction in real-world millimeter wave v2i communications," *IEEE Wireless Communications Letters*, vol. 12, no. 2, pp. 212–216, 2022.

[286] G. Charan, A. Hredzak, C. Stoddard, B. Berrey, M. Seth, H. Nunez, and A. Alkhateeb, "Towards real-world 6G drone communication: Position and aided beam prediction," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 2951–2956.

[287] J. Morais, A. Behboodi, H. Pezeshki, and A. Alkhateeb, "Position aided beam prediction in the real world: How useful gps locations actually are?" *arXiv preprint arXiv:2205.09054*, 2022.

[288] M. Zecchin, M. B. Mashhadi, M. Jankowski, D. Gündüz, M. Kountouris, and D. Gesbert, "Lidar and position-aided mmwave beam selection with non-local cnns and curriculum training," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 2979–2990, 2022.

[289] S. Rezaie, C. N. Manchón, and E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[290] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30 507–30 526, 2019.

[291] S. Jiang and A. Alkhateeb, "Computer vision aided beam tracking in a real-world millimeter wave deployment," in *2022 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2022, pp. 142–147.

[292] H. Luo, U. Demirhan, and A. Alkhateeb, "Millimeter wave v2v beam tracking using radar: Algorithms and real-world demonstration," *arXiv preprint arXiv:2308.01558*, 2023.

[293] S. Wu, M. Alrabeiah, C. Chakrabarti, and A. Alkhateeb, "Blockage prediction using wireless signatures: Deep learning enables real-world demonstration," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 776–796, 2022.

[294] S. Wu, C. Chakrabarti, and A. Alkhateeb, "Proactively predicting dynamic 6G link blockages using lidar and in-band signatures," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 392–412, 2023.

[295] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided dynamic blockage prediction for 6G wireless communication networks," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.

[296] G. Charan and A. Alkhateeb, "Computer vision aided blockage prediction in real-world millimeter wave deployments," in *2022 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2022, pp. 1711–1716.

[297] U. Demirhan and A. Alkhateeb, "Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 2655–2660.

[298] ——, "Radar aided proactive blockage prediction in real-world millimeter wave systems," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4547–4552.

[299] M. Nerini and B. Clerckx, "Overhead-free blockage detection and precoding through physics-based graph neural networks: Lidar data meets ray tracing," *IEEE Wireless Communications Letters*, vol. 12, no. 3, pp. 565–569, 2023.

[300] W. Xu, F. Gao, J. Zhang, X. Tao, and A. Alkhateeb, "Deep learning based channel covariance matrix estimation with user location and scene images," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8145–8158, 2021.

[301] J. Chen, F. Gao, X. Tao, G. Liu, C. Pan, and A. Alkhateeb, "Computer vision aided codebook design for MIMO communications systems," *IEEE Transactions on Wireless Communications*, 2022.

[302] G. Charan and A. Alkhateeb, "User identification: The key enabler for multi-user vision-aided wireless communications," *arXiv preprint arXiv:2210.15652*, 2022.

[303] G. Charan, M. Alrabeiah, T. Osman, and A. Alkhateeb, "Camera based mmwave beam prediction: Towards multi-candidate real-world scenarios," *arXiv preprint arXiv:2308.06868*, 2023.

[304] A. Alkhateeb, S. Jiang, and G. Charan, "Real-time digital twins: Vision and research directions for 6G and beyond," *IEEE Communications Magazine*, pp. 1–7, 2023.

[305] C. Ruah, O. Simeone, and B. Al-Hashimi, "Digital twin-based multiple access optimization and monitoring via model-driven bayesian learning," 2023.

[306] S. Dwivedi, R. Shreevastav, F. Munier, J. Nygren, I. Siomina, Y. Lyazidi, D. Shrestha, G. Lindmark, P. Ernstrom, E. Stare, S. M. Razavi, S. Muruganathan, G. Masini, A. Busin, and F. Gunnarsson, "Positioning in 5G networks," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 38–44, 2021.

[307] J. A. del Peral-Rosado, R. Raulefs, J. A. Lopez-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[308] P. Gao, L. Lian, and J. Yu, "Cooperative ISAC with direct localization and rate-splitting multiple access communication: A pareto optimization framework," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1496–1515, 2023.

[309] G. Kwon, A. Conti, H. Park, and M. Z. Win, "Joint communication and localization in millimeter wave networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1439–1454, 2021.

[310] U. Demirhan and A. Alkhateeb, "Cell-free ISAC MIMO systems: Joint sensing and communication beamforming," 2023.

[311] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Transactions on Information Theory*, pp. 1–1, 2022.

[312] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021.

[313] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*. IEEE, 2011, pp. 110–117.

[314] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[315] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.

[316] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for b5g networks with distributed signal processing: Semantic communication, edge computing, and wireless

sensing," *IEEE journal of selected topics in signal processing*, vol. 17, no. 1, pp. 9–39, 2023.

[317] Z. Yang, M. Chen, Z. Zhang, and C. Huang, "Energy efficient semantic communication over wireless networks with rate splitting," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1484–1495, 2023.

[318] Z. Zhao, Z. Yang, Y. Hu, L. Lin, and Z. Zhang, "Semantic information extraction for text data with probability graph," in *2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. IEEE, 2023, pp. 1–6.

[319] Z. Zhao, Z. Yang, Q.-V. Pham, Q. Yang, and Z. Zhang, "Semantic communication with probability graph: A joint communication and computation design," *arXiv preprint arXiv:2310.00015*, 2023.

[320] L. Zhao, Y. Cui, S. Yang, and S. S. Shitz, "An optimization framework for general rate splitting for general multicast," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1573–1587, Mar. 2022.

[321] R. Huang, V. W.S. Wong, and R. Schober, "Rate-splitting for intelligent reflecting surface-aided multiuser VR streaming," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1516–1535, May 2023.

[322] L. Zhao, Y. Cui, Z. Liu, Y. Zhang, and S. Yang, "Adaptive streaming of 360 videos with perfect, imperfect, and unknown FoV viewing probabilities in wireless networks," *IEEE Trans. Image Process.*, vol. 30, pp. 7744–7759, Sept. 2021.

[323] N. Q. Hieu, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "When virtual reality meets rate splitting multiple access: A joint communication and computation approach," *IEEE J. Sel. Areas in Commun.*, vol. 41, no. 5, pp. 1536–1548, May 2023.

[324] M. Setayesh and V. W.S. Wong, "Viewport prediction, bitrate selection, and beamforming design for THz-enabled 360-degree video streaming," *arXiv preprint arXiv:2401.13114*, Jan. 2024.

[325] P. Yang, T. Q. Quek, J. Chen, C. You, and X. Cao, "Feeling of presence maximization: mmWave-enabled virtual reality meets deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 10 005–10 019, Nov. 2022.

[326] R. Zhang, J. Liu, F. Liu, T. Huang, Q. Tang, S. Wang, and F. R. Yu, "Buffer-aware virtual reality video streaming with personalized and private viewport prediction," *IEEE J. Sel. Areas in Commun.*, vol. 40, no. 2, pp. 694–709, Feb. 2022.

[327] M. Setayesh and V. W.S. Wong, "A content-based viewport prediction framework for 360° video using personalized federated learning and fusion techniques," in *Proc. of Int'l Conf. on Multimedia and Expo (ICME)*, Brisbane, Australia, Jul. 2023.

[328] E. Björnson, O. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020.

[329] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4522–4535, 2020.

[330] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 990–1002, 2020.

[331] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2604–2620, 2020.

[332] B. Zheng, C. You, and R. Zhang, "Fast channel estimation for IRS-assisted OFDM," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 580–584, 2021.

[333] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, 2021.

[334] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 883–917, 2022.

[335] M. Nerini, S. Shen, H. Li, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization," *arXiv preprint arXiv:2305.05013*, 2023.

[336] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2311–2324, 2023.

[337] T. Fang and Y. Mao, "A low-complexity beamforming design for beyond-diagonal RIS aided multi-user networks, (early access)," *IEEE Communications Letters*, 2023.

[338] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1035–1071, 2022.

[339] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Communications Letters*, vol. 24, no. 4, pp. 753–757, 2020.

[340] A. S. de Sena, P. H. J. Nardelli, D. B. da Costa, P. Popovski, and C. B. Papadias, "Rate-splitting multiple access and its interplay with intelligent reflecting surfaces," *IEEE Communications Magazine*, vol. 60, no. 7, pp. 52–57, 2022.

[341] H. Li, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G—part III: Interplay with reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2242–2246, 2022.

[342] M. Soleymani, I. Santamaria, and E. A. Jorswieck, "Rate splitting in MIMO RIS-assisted systems with hardware impairments and improper signaling," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4580–4597, 2023.

[343] R. Huang, V. W. Wong, and R. Schober, "Rate-splitting for intelligent reflecting surface-aided multiuser VR streaming," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1516–1535, 2023.

[344] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Rate-splitting multiple access and dynamic user clustering for sum-rate maximization in multiple RISs-aided uplink mmwave system," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7365–7383, 2022.

[345] ——, "Rate splitting multiple access for sum-rate maximization in irs aided uplink communications," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2246–2261, 2023.

[346] T. Zhang and S. Mao, "Joint beamforming design in reconfigurable intelligent surface-assisted rate splitting networks," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[347] T. Fang, Y. Mao, S. Shen, Z. Zhu, and B. Clerckx, "Fully connected reconfigurable intelligent surface aided rate-splitting multiple access for multi-user multi-antenna transmission," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022, pp. 675–680.

[348] M. Soleymani, I. Santamaria, E. Jorswieck, and B. Clerckx, "Optimization of rate-splitting multiple access in beyond diagonal RIS-assisted URLLC systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[349] H. Li, S. Shen, and B. Clerckx, "Synergizing beyond diagonal reconfigurable intelligent surface and rate-splitting multiple access," *arXiv preprint arXiv:2303.06912*, 2023.

[350] H. Li, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for 6G–Part III: Interplay with reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2242–2246, 2022.

[351] 3GPP, "Feasibility study for evolved universal terrestrial radio access (UTRA) and universal terrestrial radio access network (UTRAN)," 3rd Generation Partnership Project (3GPP), Technical Report TR 35.912 V8.0.0, Dec. 2008.

[352] ZTE, "Study on non-orthogonal multiple access (NOMA) for NR," 3rd Generation Partnership Project (3GPP), document RP-181403, 3GPP TSG RAN Meeting #80, June 2018.

[353] 3GPP, "Study on new radio (NR) access technology," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.912 V15.0.0, June 2016.

[354] ZTE, "Revised wid on 2-step rach for NR," 3rd Generation Partnership Project (3GPP), document RP-200085, 3GPP TSG RAN Meeting #87e, March 2020.

[355] Viavi, "Rws-230414 - high-level release 19 proposal," *3GPP Rel. 19 Workshop*, Taiwan, June 2023.

[356] Qualcomm, "Study on artificial intelligence (AI)/machine learning (ML) for NR air interface," 3rd Generation Partnership Project (3GPP), document RP-221348, 3GPP TSG RAN Meeting #96, June 2022.

[357] 3GPP, "Study on artificial intelligence (AI)/machine learning (ML) for NR air interface," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.843 V1.0.0, Sept. 2023.

[358] Qualcomm, "Vision, market drivers, and research directions on the path to 6G," Tech. Rep., Dec. 2022.

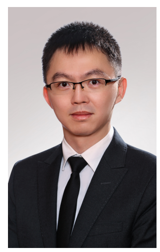
[359] Ericsson, "6G – Connecting a cyber-physical world," Tech. Rep., Feb. 2022.

[360] Nokia, "Toward a 6G AI-Native Air Interface," Tech. Rep., March 2021.
[361] Huawei, "6G: The Next Horizon," Tech. Rep., Nov. 2021.
[362] Samsung, "6G: The Next Hyper-Connected Experience for All," Tech. Rep., July 2020.

Bruno Clerckx is a (Full) Professor, the Head of the Communications and Signal Processing Group, and the Head of the Wireless Communications and Signal Processing Lab, within the Electrical and Electronic Engineering Department, Imperial College London, London, U.K. He received the MSc and Ph.D. degrees in Electrical Engineering from Université Catholique de Louvain, Belgium, and the Doctor of Science (DSc) degree from Imperial College London, U.K. He spent many years in industry with Silicon Austria Labs (SAL), Austria, where he was the Chief Technology Officer (CTO) responsible for all research areas of Austria's top research center for electronic based systems and with Samsung Electronics, South Korea, where he actively contributed to 4G (3GPP LTE/LTE-A and IEEE 802.16m). He has authored two books on "MIMO Wireless Communications" and "MIMO Wireless Networks", 300 peer-reviewed international research papers, and 150 standards contributions, and is the inventor of 80 issued or pending patents among which several have been adopted in the specifications of 4G standards and are used by billions of devices worldwide. His research spans the general area of wireless communications and signal processing for wireless networks. He received the prestigious Blondel Medal 2021 from France for exceptional work contributing to the progress of Science and Electrical and Electronic Industries, the 2021 Adolphe Wetrems Prize in mathematical and physical sciences from Royal Academy of Belgium, multiple awards from Samsung, IEEE best student paper award, and the EURASIP (European Association for Signal Processing) best paper award 2022. He is a Fellow of the IEEE and the IET.

Yijie Mao is currently an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. She received the B.Eng. degree from the Beijing University of Posts and Telecommunications, and the B.Eng. (Hons.) degree from the Queen Mary University of London (London, United Kingdom) in 2014. She received the Ph.D. degree in the Electrical and Electronic Engineering Department from the University of Hong Kong (Hong Kong, China) in 2018. She was a Postdoctoral Research Fellow at the University of Hong Kong (Hong Kong, China) from 2018 to 2019 and a postdoctoral research associate with the Communications and Signal Processing Group (CSP), Department of the Electrical and Electronic Engineering at the Imperial College London (London, United Kingdom) from 2019 to 2021. Her research interests include the design of future wireless communications and artificial intelligence-empowered wireless networks. Dr. Mao receives the Best Paper Awards of EURASIP Journal on Wireless Communications and Networking (JWCN) 2022 and IEEE International Mediterranean Conference on Communications and Networking (MeditCom) 2023, the Exemplary Reviewers for IEEE Transactions on Communications 2021 and IEEE Communications Letters 2022, 2023. She is currently serving as an associate editor for IEEE Communications Surveys & Tutorials and IEEE Communications Letters, and the lead guest editor for one special issue of IEEE Transactions on Green Communications and Networking. She was a guest editor for two special issues of IEEE Journal on Selected Areas in Communications and IEEE Open Journal of the Communications Society. She has been a workshop co-chair for 2020-2024 IEEE ICC, 2021-2023 IEEE WCNC, and 2020-2022 IEEE PIMRC, and she has been a Technical Program Committee (TPC) member of many symposia on wireless communication for several leading international IEEE conferences. She has been recognized as the World's Top 2% Scientists by Stanford University in 2023.

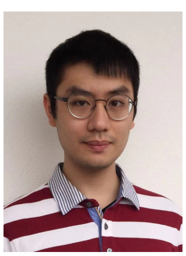
Zhaohui Yang (Member, IEEE) is currently a ZJU Young Professor with the Zhejiang Key Laboratory of Information Processing Communication and Networking, College of Information Science and Electronic Engineering, Zhejiang University. He received the Ph.D. degree from Southeast University, Nanjing, China, in 2018. From 2018 to 2020, he was a Post-Doctoral Research Associate at the Center for Telecommunications Research, Department of Informatics, King's College London, U.K. From 2020 to 2022, he was a Research Fellow at the Department of Electronic and Electrical Engineering, University College London, U.K. His research interests include joint communication, sensing, and computation, federated learning, and semantic communication. He received IEEE Communications Society Leonard G. Abraham Prize in 2024, IEEE Marconi Prize Paper Award in 2023, IEEE Katherine Johnson Young Author Paper Award in 2023. He currently serves as an Associate Editor of IEEE TGCN, IEEE CL, IEEE TMLCN. He has served as a Guest Editor for several journals including IEEE Journal on Selected Areas in Communications.



Mingzhe Chen is currently an Assistant Professor with the Department of Electrical and Computer Engineering and Frost Institute of Data Science and Computing at University of Miami. His research interests include federated learning, reinforcement learning, virtual reality, unmanned aerial vehicles, and Internet of Things. He has received four IEEE Communication Society journal paper awards including the IEEE Marconi Prize Paper Award in Wireless Communications in 2023, the Young Author Best Paper Award in 2021 and 2023, and the Fred W. Ellersick Prize Award in 2022, and four conference best paper awards at ICCCN in 2023, IEEE WCNC in 2021, IEEE ICC in 2020, and IEEE GLOBECOM in 2020. He currently serves as an Associate Editor of IEEE Transactions on Mobile Computing, IEEE Transactions on Communications, IEEE Wireless Communications Letters, IEEE Transactions on Green Communications and Networking, and IEEE Transactions on Machine Learning in Communications and Networking.

Ahmed Alkhateeb received his B.S. degree (distinction with honor) and M.S. degree in Electrical Engineering from Cairo University, Egypt, in 2008 and 2012, and his Ph.D. degree in Electrical Engineering from The University of Texas at Austin, USA, in August 2016. Between Sept. 2016 and Dec. 2017, he was a Wireless Communications Researcher at the Connectivity Lab, Facebook, in Menlo Park, CA. He joined Arizona State University (ASU) in spring 2018, where he is currently an Assistant Professor in the School of Electrical, Computer and Energy Engineering. His research interests are in the broad areas of wireless communications, communication theory, signal processing, machine learning, and applied math. Dr. Alkhateeb is the recipient of the 2012 MCD Fellowship from The University of Texas at Austin, the 2016 IEEE Signal Processing Society Young Author Best Paper Award for his work on hybrid precoding and channel estimation in millimeter wave communication systems, and the 2021 NSF CAREER Award.

Liang Liu is an Assistant Professor in the Department of Electrical and Electronic Engineering at The Hong Kong Polytechnic University (PolyU). He will be promoted to tenured Associate Professor in July 2024. His research interests include wireless communications and networking, advanced signal processing and optimization techniques, and Internet-of-Things (IoT). Dr. Liang LIU is the recipient of the 2021 IEEE Signal Processing Society Best Paper Award, the 2017 IEEE Signal Processing Society Young Author Best Paper Award, the Best Student Paper Award for 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), and the Best Paper Award for 2011 International Conference on Wireless Communications and Signal Processing (WCSP). He was listed in Highly Cited Researchers, also known as World's Most Influential Scientific Minds, by Clarivate Analytics (Thomson Reuters) in 2018. He is an editor for IEEE Transactions on Wireless Communications, and was a leading guest editor for IEEE Wireless Communications special issue on ”Massive Machine-Type Communications for IoT”. He is a co-author of the book ”Next Generation Multiple Access” published at Wiley-IEEE Press.



Min Qiu received his B.E. degree (Hons.) in Electronics and Telecommunications Engineering from The Australian National University, Canberra, ACT, Australia, in 2012. He worked in the electrical switchboard industry as an electrical engineer from 2013-2014. He received his Ph.D. degree in Electrical Engineering from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2019. He is currently a Postdoctoral Research Associate with UNSW. He received the Australian Government Research Training Program Scholarship for the duration of his Ph.D. degree, the Australia Awards-Endeavour Research Fellowship in 2018, and the Chinese Government Award for Outstanding Self-Financed Students Abroad in 2019. Dr. Qiu is a co-inventor of one patent on spatially coupled product-like codes. He was honored as the Exemplary Reviewer of the IEEE Transactions on Communications in 2018, 2019, 2021, and 2022, and the IEEE Communications Letters in 2021-2023. He has been a Technical Program Committee (TPC) member of leading international conferences such as IEEE ICC and IEEE VTC. His research interests include channel coding, information theory, and their applications for building reliable communication systems.

Jinhong Yuan received the B.E. and Ph.D. degrees in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia, where he is currently a Professor and acting Head of School and Head of Telecommunication Group with the School. He has published two books, five book chapters, over 300 papers in telecommunications journals and conference proceedings, and 50 industrial reports. He is a co-inventor of one patent on MIMO systems and four patents on low-density-parity-check codes. He has co-authored four Best Paper Awards and one Best Poster Award, including the Best Paper Award from the IEEE International Conference on Communications, Kansas City, USA, in 2018, the Best Paper Award from IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011, and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He is an IEEE Fellow and currently serving as an Associate Editor for the IEEE Transactions on Wireless Communications and IEEE Transactions on Communications. He served as the IEEE NSW Chapter Chair of Joint Communications/Signal Processions/Ocean Engineering Chapter during 2011-2014 and served as an Associate Editor for the IEEE Transactions on Communications during 2012-2017. His current research interests include error control coding and information theory, communication theory, and wireless and underwater communications.

Vincent W.S. Wong (S'94, M'00, SM'07, F'16) received the B.Sc. degree from the University of Manitoba, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. (now Microchip Technology Inc.). He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently a Professor. His research areas include protocol design, optimization, and resource management of communication networks, with applications to 5G/6G wireless networks, Internet of things, mobile edge computing, smart grid, and energy systems. Dr. Wong is the Editor-in-Chief of *IEEE Transactions on Wireless Communications*. He has served as an Area Editor of *IEEE Transactions on Communications* and *IEEE Open Journal of the Communications Society*, an Associate Editor of *IEEE Transactions on Mobile Computing* and *IEEE Transactions on Vehicular Technology*, and a Guest Editor of *IEEE Journal on Selected Areas in Communications*, *IEEE Internet of Things Journal*, and *IEEE Wireless Communications*. Dr. Wong is a General Co-chair of *IEEE INFOCOM* 2024. He was a Tutorial Co-Chair of *IEEE GLOBECOM*'18, a Technical Program Co-chair of *IEEE VTC*2020-*Fall* and *IEEE SmartGridComm*'14, and a Symposium Co-chair of *IEEE ICC*'18, *IEEE SmartGridComm* ('13, '17) and *IEEE GLOBECOM*'13. He received the 2022 Best Paper Award from *IEEE Transactions on Mobile Computing* and Best Paper Awards at the *IEEE ICC* 2022 and *IEEE GLOBECOM* 2020. He has served as the Chair of the IEEE Vancouver Joint Communications Chapter and IEEE Communications Society Emerging Technical Sub-Committee on Smart Grid Communications. He is an IEEE Vehicular Technology Society Distinguished Lecturer (2023−2025) and was an IEEE Communications Society Distinguished Lecturer (2019−2020). Dr. Wong is a Fellow of the IEEE and the Engineering Institute of Canada (EIC).

Juan Montojo is VP of Technical Standards at the Qualcomm's Corporate Standards group. He has Telecommunications Engineering degrees from the Technical University of Catalonia, Barcelona, Spain, and Institut Eurecom, Sophia Antipolis, France, as well as, MS and Ph.D. degrees in Electrical Engineering from the University of Southern California (USC) and University of California San Diego (UCSD), respectively. Dr. Montojo joined Qualcomm in San Diego in January 1997 and has worked in the system design and standardization of various communication systems including Globalstar, 3G, 4G, 5G, and WiFi as part of the corporate R&D and Standards groups. He led the 3GPP PHY layer standardization of WCDMA Rel-7 and of LTE Rel-8, 9, 10, and part of 11. He led the Qualcomm Research Germany group from Nov '11 to July '15 working on Gbps cable modem design and IEEE 802.3 standardization, as well as, 802.11ah modem design and implementation. Upon his return to San Diego, Dr. Montojo led the system design of 5G NR unlicensed and spectrum sharing until he moved back to the 3GPP team in January '17 where he helped with the execution of the 5G NR acceleration. As of late, Dr. Montojo has been focusing on AI/ML for NR air interface for which he was the rapporteur of the 3GPP Rel-18 study item and continues serving as rapporteur for the on-going Rel-19 work item. He is also researching 6G areas focusing on identifying the benefits that 6G could offer over 5G/5G-Adavanced. Dr. Montojo is Qualcomm's global lead for Cellular Standards activities and is ETSI Board member from November '23.