

Review

A Survey of NOMA-Aided Cell-Free Massive MIMO Systems

Antonio Apiyo ^{1,*} and Jacek Izydorczyk ^{2,†} ¹ Mobile Networks, Nokia, 54-130 Wrocław, Poland² Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, 44-100 Gliwice, Poland

* Correspondence: antonioshem@gmail.com

† These authors contributed equally to this work.

Abstract: The Internet of Everything is leading to an increasingly connected intelligent digital world. Envisaged sixth-generation wireless networks require new solutions and technologies due to stringent network requirements. The benefits of cell-free massive MIMO (CF-mMIMO) and non-orthogonal multiple access (NOMA) have brought substantial attention to these approaches as potential technologies for future networks. In CF-mMIMO, numerous distributed access points are linked to a central processing unit, which allocates the same time-frequency resources to a smaller group of users. On the other hand, NOMA can support more users than its orthogonal counterparts by utilizing non-orthogonal resource allocation. This paper provides a comprehensive review and survey of NOMA-aided CF-mMIMO (CF-mMIMO-NOMA). Specifically, we present a comprehensive review of massive MIMO, CF-mMIMO, and NOMA. We then present a state-of-the-art research review of CF-mMIMO-NOMA. Finally, we discuss the challenges and potential of combining CF-mMIMO-NOMA with other enabling technologies to enhance performance.

Keywords: artificial intelligence; cell-free massive MIMO; mMIMO; NOMA; NOMA-aided CF-mMIMO; 6G



Citation: Apiyo, A.; Izydorczyk, J. A Survey of NOMA-Aided Cell-Free Massive MIMO Systems. *Electronics* **2024**, *13*, 231. <https://doi.org/10.3390/electronics13010231>

Academic Editor: David A. Sánchez-Hernández

Received: 1 December 2023

Revised: 26 December 2023

Accepted: 27 December 2023

Published: 4 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The envisaged sixth generation of mobile telecommunications technology (6G) is predicted to cause a significant increase in the number of online devices. Moreover, the rise of Internet of Everything (IoE) applications will produce huge demand for data traffic, accompanied by growing demands for diverse services, extensive coverage, and extremely high-speed, exceptionally reliable, and remarkably low-latency wireless communications. The use cases driving 6G networks may include holographic telepresence, digital twinning, autonomous robotics, vehicles, the Internet of Things, distributed and large-scale data, and the blockchain.

To support these use cases, 6G networks are expected to display improved network performance by increasing the peak data rate to 1 Tbps, increasing the data transfer speed for high mobility users to 1 Gbps, increasing connection density to 10^7 devices/km², decreasing air latency to 0.1 ms, and improving network reliability by supporting packet failure probability of 10^{-7} [1,2]. These requirements can only be met by using novel intelligent communication techniques. For instance, reconfigurable intelligent surfaces (RISs), extra-large MIMO, novel spectrum utilization, holographic radio communications, advanced multiple access schemes, and modulation techniques are all key methods to optimize data transmission rates [3]. Cell-free massive MIMO (CF-mMIMO) systems and the integration of terrestrial and non-terrestrial communications are both impactful methods for improving connectivity and providing comprehensive coverage. The CF-mMIMO solution is well suited for the future generation of indoor and outdoor scenarios. In cell-free (CF) systems, users are surrounded by access points (APs), which eliminates the concept of cell edges and the traditional problem of edge users suffering the worst

performance. Crucially, CF-mMIMO systems benefit from all the advantages of network MIMO solutions, and essential qualities of massive MIMO could be exploited to support scalable solutions. Additionally, system performance can further be bettered by combining CF-mMIMO systems with other technologies like NOMA, RISs, radio stripes, and machine learning (ML). These improvements include increased data rates, reliability, and security, which ultimately will help ensure that the target requirements of 6G networks are met.

NOMA is an emerging technology that offers promise for the forthcoming generation of wireless communications. The capacity of orthogonal multiple-access approaches to service users is often constrained by the availability of orthogonal resources. NOMA enables more users to be served than the available resources would usually allow but with greater complexity of the receivers as a trade-off. The benefits of NOMA include massive connectivity, low latency, improved spectral performance, and relaxed channel feedback [4–7].

Such benefits make the integration of NOMA with CF-mMIMO highly important to the future of wireless communication networks. Independently, NOMA and CF-mMIMO have attracted substantial research interest. However, little research concerns the integration of the two. The papers most closely related to this review are those that review cell-free massive MIMO technology. These include [8–15].

Motivated by the above discussion, we present a comprehensive study of NOMA with CF-mMIMO wireless communications systems. To our knowledge, this is the first paper to offer a full review of combining NOMA and cell-free massive MIMO.

For readability, Table 1 defines the acronyms used throughout this paper. Figure 1 describes the organization of work within the paper. In Section 2, the paper provides a brief introduction to the fundamentals of massive MIMO (mMIMO), CF-mMIMO, and NOMA. In Section 3, we present a focused literature review of NOMA with CF-mMIMO. We present a simple NOMA-aided CF-mMIMO system model that considers single-antenna APs and single-antenna user equipment (UE) with a time-division duplex (TDD) transmission protocol. We further analyze the various works that concern NOMA with CF-mMIMO, to highlight the challenges and objectives of different studies. Section 4 summarizes the challenges of NOMA with CF-mMIMO and its potential integration with other enabling technologies in the current area of research. Finally, Section 5 concludes the paper and discusses key findings and possible future research directions.

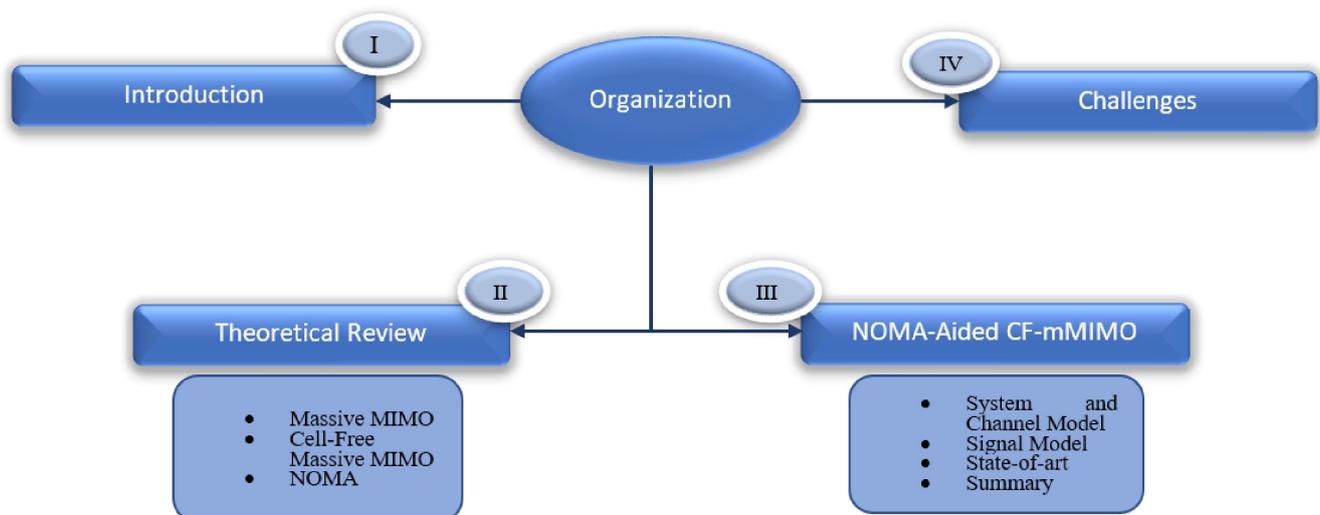


Figure 1. Organization of this paper.

Table 1. Main Acronyms Used.

Acronym	Definition
5G	Fifth generation of mobile telecommunications technology
6G	Envisaged sixth generation of mobile telecommunications technology
AI	Artificial intelligence
AP	Access points
B5G	Beyond fifth generation of mobile telecommunications technology
BE	Bandwidth efficiency
BS	Base station
CBSM	Correlation-based stochastic model
CDMA	Code division multiplexing access
CF	Cell-free
CF-mMIMO	Cell-free massive MIMO
CF-mMIMO-NOMA	NOMA-based Cell-free massive MIMO system
CF-mMIMO-OMA	OMA-based Cell-free massive MIMO system
CPU	Central processing unit
CSI	Channel state information
DL	Downlink
FDD	Frequency Division Duplex
FL	Federated learning
GBSM	Geometry-based stochastic channel model
GF	Grant free
GP	Geometric programming
i.i.d	Independent and identically distributed
IoE	Internet of Everything
IoT	Internet of Things
LDPC	Low-density parity check
LTE	Long term evolution
MEC	Multi-access edge computing
MIMO	Multiple input multiple output
mMIMO	Massive multiple input multiple output
ML	Machine learning
MMSE	Minimum mean square error
MSE	Mean-squared errors
MRT	Maximum ratio transmission
MUD	Multi-user detection
MU-MIMO	Multi-user multiple input multiple output
NOMA	Non-orthogonal multiple access
OFDMA	Orthogonal frequency division multiple access
OMA	Orthogonal multiple access
OTFS	Orthogonal time frequency and space

Table 1. Cont.

Acronym	Definition
PA	Power Allocation
PAPR	Peak-to-average power ratio
PD-NOMA	Power domain NOMA
RIS	Reconfigurable intelligent surface
rZF	Regularized zero-forcing
SC	Superposition coding
SCA	Sequential convex approximation
SDP	Semi-definite programming
SE	Spectral efficiency
SIC	Successive interference cancellation
SINR	Signal-to-interference noise ratio
SOCP	Second-order cone programming
TDD	Time division duplex
UC	User Clustering
UE	User equipment
UL	Upper link
ZF	Zero-forcing

2. Theoretical Review

To understand NOMA-aided CF-mMIMO, we revisit some basics of mMIMO, CF-mMIMO, and NOMA in this section. This will help the reader to better understand the potential benefits of combining these technologies in NOMA-aided CF-mMIMO systems.

2.1. mMIMO Fundamentals

MIMO technology has gained significant attention in wireless communications as it allows the utilization of many antennas at the transmitter, the receiver, or both, to boost link performance. MIMO offers increased data throughput and extended communication range without requiring more bandwidth. MIMO accomplishes this by exploiting spatial dimension encoding and decoding.

Multi-user MIMO (MU-MIMO) is a critical technique in 4G communication systems that is utilized to boost network capacity by categorizing users depending on their spatial location. This approach provides a stronger guarantee for spatial multiplexing, although it acquires challenging issues such as near–far power control and time and frequency synchronization from point-to-point systems [16].

A sizeable new antenna array technique massive MIMO (mMIMO), often referred to as large-scale MIMO, further scales the gains of MU-MIMO. These systems have many times more antennas at the base station (BS) than traditional MIMO systems, allowing them to serve dozens of UEs simultaneously [16,17].

As shown in [16–26], mMIMO provides significant improvements in spectral efficiency (SE), reducing energy consumption, and minimizing network interference via channel hardening, favorable propagation, and low complexity signal processing. Such improvements are crucial to addressing the demands of data-centric systems with the increasingly precious spectrum and energy resources.

Figure 2 shows the basic concept of mMIMO, where a BS with M antennas connects with K single-antenna UEs and $K \ll M$.

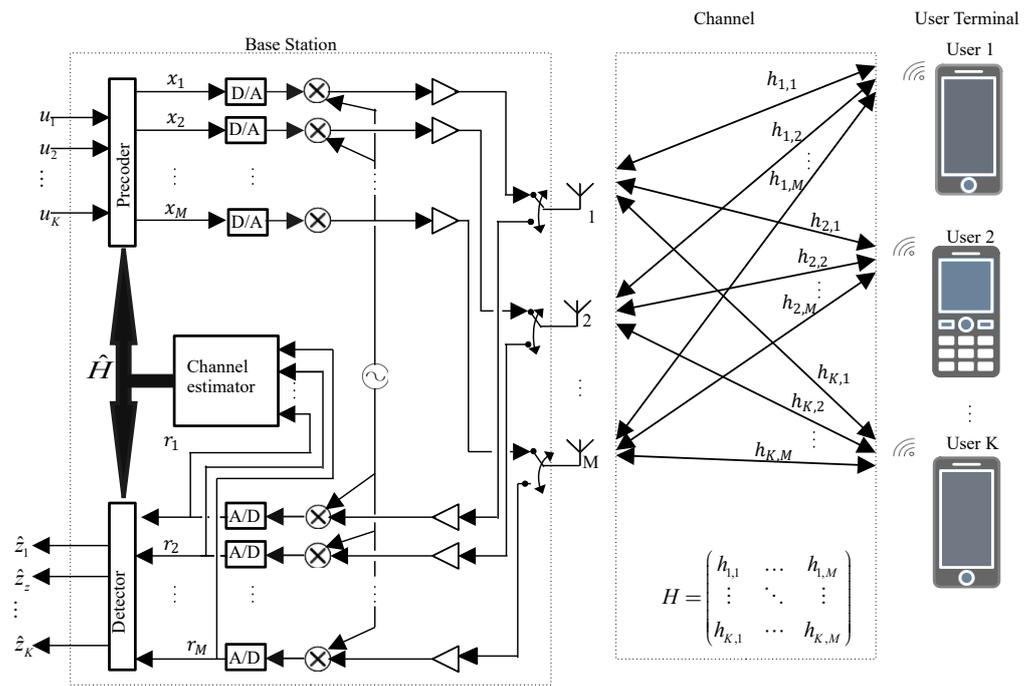


Figure 2. Simplified model of a multi-user massive MIMO system, using time-division duplexing and channel reciprocity.

To successfully multiplex uplinks and downlinks spatially, BS must first understand the propagation channel in both directions. This knowledge is essential for designing efficient downlink precoders and uplink decoders.

Because acquiring channel state information (CSI) in the downlink is challenging, mMIMO primarily relies on uplink channel estimation, channel reciprocity, and TDD. As such, Marzeta et al. [19] define an authorized mMIMO network as a multicarrier cellular network having L cells operating synchronously using TDD protocol. The j th BS has $M_j \gg 1$ antennas to provide channel hardening and communicates with K_j single-antenna UEs concurrently on every time-frequency sample, having an antenna-UE ratio of $M_j/K_j > 1$. Every BS works independently and utilizes linear receive combining and linear transmit precoding to handle signals.

Operation using the frequency division duplex (FDD) may be possible in specific scenarios [23]; the implementation of efficient FDD mMIMO poses substantial challenges which remain to be solved [18].

To provide channel estimation across both BSs and mobile stations with a massive number of channels, a sufficiently long channel coherence time is required to ensure efficient operation. The accuracy of channel estimation and the channel coherence time are fundamental limitations of mMIMO systems.

mMIMO systems offer several potential benefits, including [26]:

- **Reliability:** mMIMO systems can attain better reliability by using many antennas to provide diversity gain. This means that if one antenna is affected by fading or interference, the signal can still be received by another antenna.
- **Spectral efficiency:** mMIMO systems can attain better spectral efficiency by utilizing a vast number of antennas to create multiple beams. This permits multiple users to use the same frequency range without interfering with each other.
- **Energy efficiency:** mMIMO systems can transmit more data with less power thanks to coherent combining. Also, one can increase throughput by utilizing more antennas without raising the transmit power—hence, the energy efficiency improves and, similarly, so does the system stability.

- User tracking: user tracking becomes more accurate as mMIMO systems can use numerous antennas to create narrow beams allowing the system to focus on the desired user and ignore interference from other users.
- Cost efficiency: mMIMO systems do not require as many expensive components. For example, massive MIMO systems can use lower-cost amplifiers because they do not need to transmit as much power.
- Robustness: mMIMO systems can be very resilient to interference and jammings because of the vast number of antennas since they can still receive signals even if some of the antennas are affected by interference or jamming.
- Enhanced security: a large amount of antenna terminals results in many degrees of freedom ensuring security. Moreover, mMIMO systems are naturally robust to hacking and passive eavesdropping attempts because of the orthogonal channels of receivers and narrow beams.
- Simple signal processing: a huge amount of antennas decreases the interference effects, fast fading, uncorrelated noise, and thermal noise, simplifying the signal processing.

Even though the countless antennas benefit the communication system, mMIMO imposes novel signal processing challenges like pilot contamination, channel estimation, precoding, hardware efficiency, and data detection.

2.2. mMIMO Theory

An mMIMO wireless system can utilize a single-cell (SC) or multi-cell (MC) structure. A single-cell structure has one BS with multiple antennas serving multiple users, while a multi-cell system has multiple linked single-cell systems. Additionally, even if the BSs in various cells work jointly, the MC structure is classed as noncooperative or cooperative systems [27]. In MC systems, the capacity to accommodate more users is restricted considering the finite quantity of orthogonal pilot symbols available within a certain time frame and bandwidth. This constrains the total amount of users the system can support. To accommodate additional users, neighboring cells in the system employ non-orthogonal pilots. One of the simple schemes of slotting pilot sequences to users across various cells involves reusing a similar group of orthogonal pilot sequences across every co-channel cell. However, this approach can lead to pilot contamination as similar pilot sequences allocated to users in adjacent co-channel cells will interfere with one another [28].

2.2.1. Precoding

The selection of the transmit signal in an mMIMO technique is crucial for ensuring that devices get the intended symbols while suppressing interference resulting from symbols meant for other devices. Precoding technology can mitigate the impacts of interference and fading, ultimately enhancing the system's throughput capacity.

Precoding strategies may be categorized as either linear or non-linear, based on considerations like the peak-to-average power ratio (PAPR) and maximum likelihood (ML) criteria [26].

Linear Precoding

Given that mMIMO presumes excessive BS antennas, it can be inferred that linear precoding methods are likely to perform effectively in scenarios characterized by favorable signal propagation conditions. Linear precoding can be applied in both SC and MC environments.

Figure 3 illustrates the general structure of communication systems incorporating precoding and decoding techniques. F is the feedforward matrix of linear precoding, B represents a feedback matrix of linear precoding, \mathbb{K} represents the feedforward matrix of non-linear decoding, and C represents the feedback matrix of non-linear precoding. These matrices define the precoding techniques employed, which may be linear, non-linear, or hybrid.

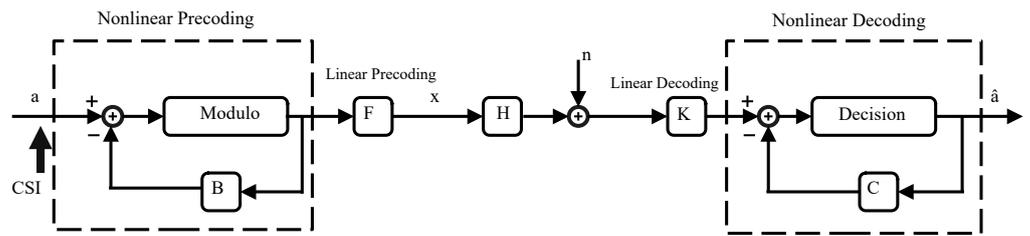


Figure 3. Generalized block diagram of a communication system with precoding and decoding techniques.

In Figures 2 and 3, the sequences relayed to the K users are represented as $u \triangleq (u_1, \dots, u_k)^T$, provided u_k corresponds to the symbols directed to the k th user. Following precoding, the resulting transmitted signals are represented as $x \triangleq (x_1, \dots, x_m)^T$, where x_m represents the relayed signal by the m th antenna. In downlink transmission, the signals acquired by the users are generally represented as $\hat{u} \triangleq (\hat{u}_1, \dots, \hat{u}_k)^T$, where \hat{u}_k denotes the signal received by k th user. The H matrix $K \times M$ dimensions describe the channel. The received signal could then be represented as follows:

$$\hat{U} = Hx + w. \tag{1}$$

where $w \sim \mathcal{CN}(0, I_k)$ is an i.i.d complex Gaussian-received noise vector with variance matrix I_k . The transmitted signals, x , are created via the precoding function:

$$x = f_{pre}(\hat{H}, u). \tag{2}$$

which depends on the estimation of the channel \hat{H} , the elements $\hat{u} \triangleq (u_1, \dots, u_k)^T$ that are sent to the users, and the available transmit power P . Mostly, linear precoding schemes are considered when the transmitted signal is expressed as $x = F(\hat{H})u$, with the matrix $F(\hat{H})$ denoting the linear precoding matrix derived from the channel estimate and other system parameters, like available transmit power.

In basic precoding techniques, the feedforward matrix F often involves a matrix inversion operation, which can result in large computational complexity [26]. Linear precoding methods can be categorized based on their approach to handling the matrix inversion process. These categories include basic linear precoding, a linear precoder based on matrix inversion approximation, a linear precoder based on fixed-point iterations, and a linear precoder based on matrix decomposition.

1 The basic linear precoding algorithms include

- The maximum-ratio transmission (MRT) precoding, also referred to as conjugate beamforming or a Matched Filter (MF). The feedforward matrix is given by

$$F_{MRT} = (\alpha_{MRT}H^H). \tag{3}$$

where α_{MRT} is a normalizing scalar that controls the transmit power or the received signal-to-noise ratio (SNR), among other variables. MRT maximizes the array gain of the transmission. However, disturbances from other users remain in the received signal due to the lack of active interference mitigation. In typical scenarios, MRT suppresses interference naturally with a boost in the amount of BS antennas and there is orthogonalization of user channels as the limit is approached [16].

- Zero-forcing (ZF) nullifies all inter-symbol and inter-user interference. The pseudo-inverse of H gives the precoding matrix:

$$F_{ZF} = (\alpha_{ZF}H^*(HH^H)^{-1}). \tag{4}$$

α_{ZF} is a normalizing scalar, and HH^H is the Gram matrix, G . The primary distinction between ZF and MRT lies in the matrix inversion step, which is responsible for achieving the desired interference suppression. However, such inverse computations often introduce considerable computational complexity. The properties of mMIMO channels enable a substantial decrease in complexity in comparison to the general inversion of matrices.

- The regularized zero-forcing (rZF) precoder shares properties of both MRT and ZF. The rZF matrix is given by

$$F_{RZF} = (\alpha_{RZF} H^* (HH^H + \beta_{reg} I_{kN})^{-1}). \quad (5)$$

The regularization constant, β_{reg} , may be set to balance the compromise between array gain and interference suppression, and I_{kN} represents an identity matrix of size kN . If β_{reg} is chosen to reduce the mean-squared error (MSE) $E\|u - \frac{1}{\sqrt{\rho}}\hat{u}\|^2$, where ρ is a scaling constant, we achieve the minimum MSE (MMSE) precoder.

- 2 The linear precoding based on the matrix inversion:
This is used to beat the inveterate noise boost that results when the number of M significantly exceeds K causing G to become diagonal dominant, with the non-diagonal components approaching zero. The diagonal elements tend to approach M . These algorithms include:
 - (a) The Truncated Polynomial Expansion (TPE) algorithm.
 - (b) The Neumann Series Approximation (NSA) algorithm.
 - (c) The Newton Iteration (NI) algorithm.
 - (d) The Chebyshev Iteration (CI) algorithm.
- 3 The fixed-fixed point iteration-based algorithms:
These methods iteratively solve linear precoding equation $Gx = a$ to determine the matrix G . They include:
 - (a) The Gauss–Seidel (GS) algorithm.
 - (b) The Successive Over-Relaxation (SOR) algorithm.
 - (c) The Conjugate Gradient (CG) algorithm.
 - (d) The Jacobi Iteration (JI) Algorithm.
- 4 Based on Matrix Decomposition:
In small-scale MIMO systems, the direct algorithms–matrix decomposition precoder is commonly employed as an alternative to explicit matrix inversion for the inversion process. It demonstrates superior numerical stability compared to basic linear precoder algorithms like the MRT, ZF, and MMSE algorithms. Additionally, it enables a modular design, allowing for the distribution of the inversion process across different components. However, in the subject of mMIMO systems, the use of direct algorithms–matrix decomposition has substantial computational complexity. The direct algorithms matrix decomposition involves decomposing the G matrix to a product of smaller matrices, similar to the QR algorithm and the Cholesky decomposition algorithm [26]. They include the following:
 - The QR Decomposition algorithm.
 - The Cholesky Decomposition (CD) algorithm.

Linear precoders generally offer the advantage of low complexity, making them computationally efficient. They provide good performance when the channel correlation is low. However, their performance deteriorates significantly in situations where there are high correlations among user channels [29].

Non-Linear Precoding

A comprehensive study has been carried out to determine the most effective precoding method for mMIMO systems to achieve high throughput performance while minimizing complexity. However, it is important to address the limitations of linear precoders, de-

spite their advantage of low complexity, as they may not provide sufficient precoding accuracy [26,29]. Non-linear precoding schemes exhibit greater robustness against channel correlation on UEs and can drastically boost the performance of mMIMO systems [29].

The non-linear precoding schemes include:

1. Dirty-paper coding (DPC.)
2. Tomlinson–Harashima (TH) precoding.
3. Vector Perturbation (VP) precoding.
4. Lattice Reduction-Aided (LR) precoding.

Peak-to-Average Power Ratio (PAPR) Precoding

Efficient non-linear amplifiers can be leveraged to realize the practical implementation of mMIMO, leading to reduced implementation costs. Hence, minimizing the PAPR becomes crucial to mitigate the impact of amplifier non-linearities. Several precoding algorithms have been developed to reduce PAPR, including:

1. Constant Envelope (CE) Precoding.
2. Approximate Message Passing (AMP) Precoding.
3. Quantized Precoding (QP).

There are other machine-learning precoding techniques in mMIMO. The authors of [30] discussed deep neural networks (DNN)-based downlink precoding. Specifically, DNN-based precoders reduce computational complexity with negligible performance degradation [31].

2.2.2. Uplink

The uplink transmission is utilized to send data and the pilot message from the UE to the BS. Assuming a reciprocal channel, during uplink, the user side features no precoding since they are not assumed to cooperate in reducing interference. Instead, users have control over their power levels. Using the same notation as shown in Figure 2, representing the power levels of the K users as a $K \times K$ diagonal matrix P_{UL} , and the transmitted user messages are collected as $z \triangleq (z_1, \dots, z_k)^T$, the information attained by the M antennas, justified as $r \triangleq (r'_1, \dots, r_M)^T$, can be expressed as:

$$r = H\sqrt{P_{UL}}z + w. \quad (6)$$

with $w \sim \mathcal{CN}(0, I_M)$ representing i.i.d zero-mean complex Gaussian noise. The received user symbols at the BS are analyzed from a detector function:

$$\hat{z} = g_{det}(r, \hat{H}). \quad (7)$$

which depends primarily on the channel estimate \hat{H} and the received signal r . It additionally considers inherent parameters such as the receiver SNR. There should be a class of detectors that use only linear processing to combine the received signals of all antennas:

$$\hat{z} = G(\hat{H})r. \quad (8)$$

Linear Detection Schemes

By a correct blending of the received signals from the M antennas, it becomes possible to amplify the required signals whilst rejecting unwanted signals. Since the downlink and uplink transmissions of TDD systems are broadcasted along an identical set of reciprocal channels, both processes typically achieve the same rate, uplink–downlink duality. Because of this, the precoding schemes mentioned earlier have direct analogs for uplink detection. In other words, the same principles and techniques used in downlink precoding can be applied to uplink detection to achieve desirable performance in terms of signal separation and interference rejection. By the premise that the channel estimation provides accurate information on true channels, we can consider $\hat{H} = H$.

The maximum ratio combiner (MRT) is the counterpart of the MRT precoder. It is also known as Matched Filtering. It has the following combining matrix:

$$G_{MRC}(H) \propto H^H. \quad (9)$$

Like MRT, MRC also combines the array gain. However, MRC does not actively reduce interference between signals from different users.

The zero-forcing combiner corresponds to the ZF precoder, and utilizes the following combining matrix:

$$G_{ZF}(H) \propto H^H(HH^H)^{-1}. \quad (10)$$

The full rank of the channel completely eliminates interference between user signals. However, array gain is sacrificed to accomplish this. The regularized zero-forcing combiner is analogous to the rZF precoder and is used to lessen the mean squared errors (MSE) of the approximated symbols. Therefore, it is also known as minimum mean-squared error (MMSE) detection or regularized ZF detection. Its output combining matrix is

$$G_{RZF} \propto H^H(HH^H + \beta_{reg}I_k)^{-1}. \quad (11)$$

where the regularization constant is set to balance between array gain and interference reduction. By selecting β_{reg} to restrict the MSE associated with the signals sent z and the signal analyzed and obtained, \hat{z} , denoted as $E\|u - \frac{1}{\sqrt{\rho}}\hat{u}\|^2$, where ρ is a scaling constant, we get the MMSE combiner.

2.2.3. Channel Estimation

Before any data are precoded or detected, the channel parameters should be estimated. mMIMO presents two significant concerns for channel estimation: the estimation of numerous channel coefficients and the requirement of multiple pilot signals. The former problem pertains to complexity, while the latter problem relates to the allocation of radio resources. The transmission of pilot signals reduces the portion of resources that are available for data transmission. This, along with the channel's rate of change, presents one of the fundamental limitations of mMIMO. However, we can ameliorate this problem by employing TDD, with the assumption that reciprocity and channel estimation performance is only for the uplink. Similar approaches using FDD also exist, which require channel estimates for both uplink and downlink. However, these approaches depend on particular channel properties [16].

In the case of TDD systems, the uplink and downlink transmissions use the same frequency spectrum with distinct time periods, hence channel reciprocity matters. The BS must have CSI to detect the signal received from the users; hence, in the uplink, the users send orthogonal pilot symbols to the BS. Then, the BS estimates the channels depending on the obtained pilot signals. In the downlink, like the uplink, the users send orthogonal pilot sequences to the BS, and the BS approximates CSI. Then, the BS forms pilot sequences and transmits them to the users. Users determine the optimal channel gain using the obtained pilot message [32].

In FDD systems, the uplink and downlink channels are not reciprocal since both transmissions use different frequency spectrums. In the downlink transmission, the BS requires CSI to precode the symbols prior to transferring them to users. The BS antenna transmits orthogonal pilot sequences to users. Subsequently, every user performs channel estimation using the obtained pilot signals and provides feedback to the BS via uplink. CSI is essential for the BS to accurately decode the message sent from the users in uplink transmission. Users send orthogonal pilot sequences to the BS, which allows it to approximate the channels depending on the pilot signals. This approach has limited application in mMIMO systems where the number of time-frequency resources required for the pilot transmission on the downlink increases proportionally with the number of antennas. Consequently, the number of channel responses that terminals should estimate is also boosted by the number of antennas [32].

However, due to the restriction of coherent properties of the channel, the amount of patterns is restricted. The reuse of orthogonal pilot patterns is necessary across different cells. These pilot sequences may be nonorthogonal, leading to pilot contamination problems [33]. The presence of pilot contamination negatively impacts the performance of the system. There are different ways of suppressing pilot contamination. However, this article will not discuss these because pilot contamination is not a primary constraint in CF-mMIMO.

2.3. mMIMO Channel Models

Certain specific phenomena must be considered when modeling the behavior of mMIMO radio channels. These include propagation effects that are relevant when using a BS having a physically numerous array consisting of many antenna elements and serving many users located in close proximity [16]. Several varieties of mMIMO channel models exist, including conventional MIMO models, the extensions of conventional MIMO models, and an mMIMO extension of the COST 2100 channel model [16]. For simplicity, we review the models presented in [24,25].

We provide a basic review of two types of channel models: correlation-based stochastic models (CBSMs) and geometry-based stochastic models (GBSMs). These models are often utilized to assess the performance of wireless communication systems.

CBSMs serve as conceptual models for the performance evaluation of mMIMO networks. There exist three types of CBSM:

- The i.i.d Rayleigh channel model makes no assumptions about correlation or mutual coupling between transmit and receive antennas. The fast-fading matrix elements are i.i.d random Gaussian variables. The i.i.d quality provides favorable propagation, offering both superior performance and simplified algorithm design for mMIMO systems. Moreover, the channel-hardening phenomena can lessen the influence of fast fading on the scheduling gain, reducing the complexity of scheduling schemes [25].
- The correlation channel model accounts for antenna correlation owing to both the antenna spacing and scattering properties of the surrounding environment. This model introduces the angle of the arrivals parameter, which can be initialized to differentiate between UE and enhance the channel estimation accuracy.
- The mutual coupling channel model accounts for the increased impact of mutual impedance as the number of antennas grows. Moreover, it considers load impedance and antenna impedance for increased realism. As such, this model is more practical for use with mMIMO systems. In addition, the model can help in the investigation of the effect of antenna space on mMIMO performance—a key factor in the design of the antenna array configuration.

GBSMs are cluster-based models which describe a propagation channel with several clusters having varying delays and power factors. Each cluster can be modeled in 2D or 3D.

The use of mMIMO for wireless communications is increasing rapidly thanks to its numerous advantages in 5G standardizations. With the introduction of beyond-5G (B5G) and 6G use cases, future systems will require high SE, making MIMO an ideal candidate. However, the literature acknowledges several challenges faced by mMIMO. These include, but are not limited to, pilot contamination, hardware impairments, uplink detection complexities, and issues surrounding channel estimation, precoding techniques, user scheduling, and energy efficiency. Duplex modes are a further area of concern, particularly full-duplex and FDD transmission modes [16–25].

2.4. CF-mMIMO Fundamentals

Mobile communications continue to evolve as human demand increases. The current focus of mobile communications is shifting increasingly toward 6G, while 5G deployments continue. CF-mMIMO has attracted substantial interest as a potential enabling technology for the envisioned 6G network [34,35]. In addition, the authors of [36,37] suggest possible 6G architectures that are CF and feature mesh connectivity. CF-mMIMO was first intro-

duced in [38,39]. It was proposed as an alternative to address the inter-cell interference in cellular networks [39].

A CF-mMIMO network consists of several widespread APs that are linked to a central processing unit (CPU). The APs collectively service all of the UE within the network simultaneously [8–10,38–45]. The APs are connected over a fronthaul link to a CPU that facilitates cooperation among the APs. The CF network may be partitioned into two sections: the edge and the core. The edge comprises APs, CPUs, and the fronthaul links, while the core network handles all the UE functions. The connectivity between the core and edge is established through backhaul links. CPUs communicate with the core network through such links. Figure 4 shows a simple CF-mMIMO network architecture.

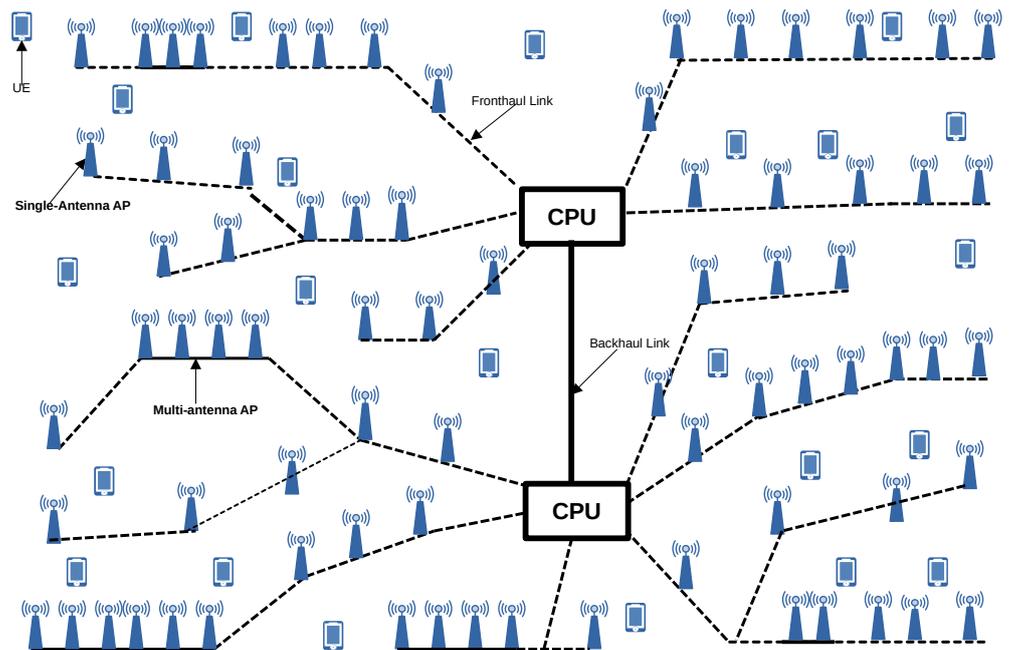


Figure 4. A simple CF-mMIMO network architecture.

The term “cell-free” indicates that the network features no cell boundaries from the perspective of the UE. This status arises because all APs that interact with a UE will participate in communication [45]. CF-mMIMO can be considered as an umbrella term for CF networks, that covers conventional mMIMO, conventional coordinated multipoint, and traditional ultra-dense networks as a case study.

The most distinguishing features of CF architecture are as follows. First, the utilization of TDD protocol to exploit channel reciprocity for the uplink and downlink communications. Second, estimates of the uplink channels are computed locally at every AP, using the pilot signals transmitted by the UEs. These estimates are then exploited locally, and so do not have to be sent via the backhaul link. Third, the beamformers utilized at the APs are generated locally, rather than at the CPU. Finally, the backhaul is utilized to transmit information symbols via the downlink and statistics via the uplink, which are sufficient for making centralized uplink information decisions.

TDD Communication Protocol

As for conventional mMIMO, TDD operation is preferred for CF-mMIMO because the channel estimation overhead is solely determined by the number of users and unaffected by the number of APs. This property gives CF-mMIMO substantial scalability: the addition of more APs will not impact channel estimation and will always provide an increased data rate. Using the reciprocity principle, the uplink channel estimates may also be employed for

the downlink channel estimates, serving both the uplink and downlink communications. Figure 5 depicts the TDD communication protocol.

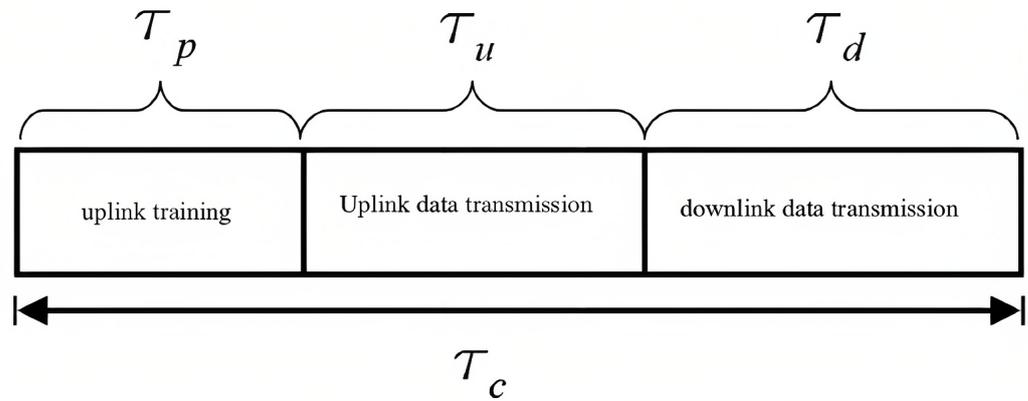


Figure 5. The TDD transmission protocol.

The coherence interval, which spans τ_c time units, contains distinctive three activities: uplink training (of length τ_p time units), uplink data transmission (of length τ_u time units), and downlink data transmission (of length τ_d time units).

During the uplink training phase, users first send their pilot symbols consisting of τ_p time units to the APs. Subsequently, the APs calculate their channels for every user for the pilot message received. As in [12,45], we represent the pilot sequence transmitted k th UE as $\phi_k \in \mathbb{C}^{N_{UE} \times \tau_p}$. We assume that the rows of ϕ_k are orthogonal, that is, $\phi_k \phi_k^D = I_s$. Given that a single pilot sequence can be shared among many UEs, we use the notation $t_k \in \{1, \dots, \tau_p\}$ to represent the pilot index dedicated to the k th UE. The signal received by l th AP during the pilot transmission is represented as $y_l^{PILOT} \in \mathbb{C}^{N \times \tau_p}$ and can be expressed as:

$$Y_l^{PILOT} = \sum_{i=1}^K \sqrt{\eta_i} h_{il} \phi_{it}^T + N_l. \tag{12}$$

where $N_l \in \mathbb{C}^{N \times \tau_p}$ represents the receiver noise, having i.i.d elements following $N_C(0, \sigma_{UL}^2)$. The uplink signal acquired serves as the lead for the l th AP to determine the channel. The determination can be performed either directly at the l th AP or handed over to the CPU. For the second scenario, the AP functions as a relay and transmits the acquired pilot signals to the CPU through fronthaul links.

Channel h_{kl} is typically estimated by either an AP or CPU, using the received pilot signal Y_l^{PILOT} . Interference from UEs can be removed by using orthogonal pilots and multiplying Y_l^{PILOT} by the normalized conjugate of the corresponding pilot ϕ_{tk} . It yields the following:

$$Y_{t_k l}^{PILOT} = \sqrt{\eta_i \tau_p} h_{kl} + \sum_{i \in P_k / \{K\}} \sqrt{\eta_i \tau_p} h_{il} + n_{t_k l}. \tag{13}$$

where the first term represents the desired part, the second term represents interference, and the third term represents the noise. The linear MMSE scheme is typically employed for channel estimation. This technique leverages channel statistics to achieve accurate estimates. We choose $\tau_p > K$ if the coherence interval is sufficiently larger than the number of users. The pilot sequence may then be allotted to K UEs in a pairwise orthogonal manner. Otherwise, non-orthogonal pilot symbols should be utilized across the setup. For this instance, the channel estimates of a particular UE will suffer interference from the pilot messages sent by different users, degrading the system performance whether numerous APs are used. This is known as pilot contamination.

During the uplink payload transmission, every user simultaneously transmits messages to the APs. The signal sent by the k th user will be denoted as $x_{u,k} \in \mathbb{C}$ and it satisfies

$\mathbb{E}\{|x_{u,k}|^2\} = 1$. These signals are mutually independent and have no correlation with noise and channel coefficients. The normalized transmit power is represented as ρ_u . The acquired signal at the m th AP may then be denoted as

$$Y_{u,m} = \sqrt{\rho_u} \sum_{k=1}^K g_{mk} \sqrt{\eta_k} x_{u,k} + n_{u,m}. \quad (14)$$

where $n_{u,m} \sim \mathcal{CN}(0, 1)$ represents the additive noise. Each AP employs its local channel estimates to examine the incoming signals and transmits the examined signals to the CPU. The m value of each AP is used to multiply $y_{u,m}$ by \hat{g}_{mk}^* , providing conjugate beamforming and matched filtering. The CPU then detects all messages sent by the M APs, with the k th user's signal detected as:

$$\begin{aligned} r_{u,m} &= \sum_{m=1}^M \hat{g}_{mk}^* y_{u,m} \sqrt{\rho_u} \\ &\times \sum_{k'=1}^K \sum_{m=1}^M \hat{g}_{mk'}^* g_{mk'} x_{u,k'} + \sum_{m=1}^M \hat{g}_{mk}^* n_{u,m}. \end{aligned} \quad (15)$$

Each AP should use simple linear processing, like the maximum ratio for signal processing. Signal processing may also be performed on CPUs. The APs must transmit their channel estimations and recorded information to the CPU for signal detection.

Finally, during the downlink payload transmission phase, the APs utilize their local channel estimations to pre-code the symbols meant for the K UEs and then broadcast the precoded symbols to all users. The information symbol for the k th user is $x_{d,k} \in \mathbb{C}$ with $\mathbb{E}\{|x_{d,k}|^2\} = 1$. The data signal sent by the m th AP for the K users is denoted as

$$Y_{dm} = \sqrt{\rho_d} \sum_{k=1}^K \hat{g}_{mk}^* x_{d,k}. \quad (16)$$

Users then individually extract the expected symbol from the received signals. Each AP can use either maximum ratio processing or conjugate beamforming. The received signal is given by

$$\begin{aligned} r_{dk} &= \sum_{m=1}^M g_{mk} y_{d,m} + n_{d,k} \\ &= \sqrt{\rho_d} \sum_{m=1}^M \sum_{k'=1}^K \hat{g}_{mk'}^* g_{mk} x_{d,k'} + n_{d,k}. \end{aligned} \quad (17)$$

CF-mMIMO is anticipated to deliver many significant advantages, such as substantial throughput, remarkably low latency, exceptional reliability, excellent energy usage and uniform coverage, flexible and economical utilization, channel hardening, optimal propagation conditions, and a consistent level of service [8,43].

Despite these benefits, CF-mMIMO still faces some challenges and limitations. First, practical implementations require a large number of backhaul connections, particularly if there are numerous APs. Hence, an appropriate transmission approach is required. A user-centric approach is one possible solution. However, the system performance when using such an approach will not be greater than that of the traditional CF-mMIMO in scenarios where there are more UEs than APs. This approach does allow the use of multiple CPUs. Secondly, it is essential to synchronize the system to ensure that users can coherently be served by the APS. Finally, CF-mMIMO is susceptible to pilot contamination, which significantly degrades the system's performance.

The incorporation of CF-mMIMO with other novel schemes offers the potential to further enhance system performance in various aspects, including increased achievable rates,

improved reliability, enhanced security, and higher connection density [9,11]. These technologies include physical layer security, RIS, the radio stripes system, federated learning, ML, unmanned aerial vehicles (UAVs), and NOMA—our focus.

CF-mMIMO builds upon the benefits and characteristics of mMIMO while providing additional features beyond those. Channel hardening and favorable propagation are inherited from mMIMO. CF-mMIMO also features macro-diversity and signal spatial sparsity, which is not present in mMIMO. We now define these features.

Channel hardening is the phenomenon where the impacts of small-scale fading are mitigated. This causes device channels to become similar to that of deterministic wired channels as the antennas increase toward infinity. Favorable propagation refers to the channels of various devices becoming orthogonal as the antennas approach limitlessness, making various equipment spatially differentiable. Macro diversity is a signal combination technique that finely integrates multiple copies of a signal into a single powerful signal. Macro diversity gain is increased due to the geographic distribution of the APs, with each AP being surrounded by several neighbors. This reduces the distance from any device to the nearest AP, in comparison to mMIMO. The signals a user sends to dissimilar APs undergo varying levels of large-scale fading. As such, neighboring APs typically capture stronger signal energy than more distant APs. This macro diversity results in non-negligible channel gains for neighboring APs, leading to signal spatial sparsity.

2.5. NOMA Fundamentals

Users can share a limited portion of the radio spectrum thanks to multiple access mechanisms. This allows for efficient utilization of the bandwidth by serving multiple users simultaneously, resulting in increased capacity. Such schemes should be implemented without degradation in the system performance.

Generally, multiple access schemes can be broadly categorized into orthogonal and non-orthogonal approaches. A variety of multiple access schemes are discussed in [6,7,46].

Orthogonal multiple access (OMA) systems utilize a resource allocation method where users are assigned orthogonal resources to prevent interference within the cells. The number of active users is constrained by the availability of orthogonal resources. For this reason, OMA is unable to fulfill the rising needs of throughput requirements and the system capacity for subsequent systems [47].

Non-orthogonal multiple access (NOMA) systems can support multiple users within a single resource, leading to enhanced throughput for both individual users and the overall system. However, this comes to the detriment of higher receiver complexity, necessary for separating the non-orthogonal signals [6]. Figure 6 shows a comparison between basic downlink NOMA and orthogonal frequency division multiple access (OFDMA).

As discussed in [4–7], possible benefits of NOMA include the following:

- Massive connectivity, allowing an unlimited user capacity.
- Low latency, allowing NOMA to support flexible scheduling and grant-free transmission.
- Improved SE—each NOMA user enjoys access to the entire bandwidth, and appropriately grouped users have improved data rates.
- Relaxed channel feedback—since perfect uplink CSI is not a must at the BS, only the received signal power must exist for channel feedback.

The components of the NOMA cellular system include multi-user grouping, resource allocation (of power, code, etc.), and successive interference cancellation (SIC) or multi-user detection (MUD) techniques for removing the controlled NOMA additions. We now discuss the defining properties of NOMA non-orthogonality.

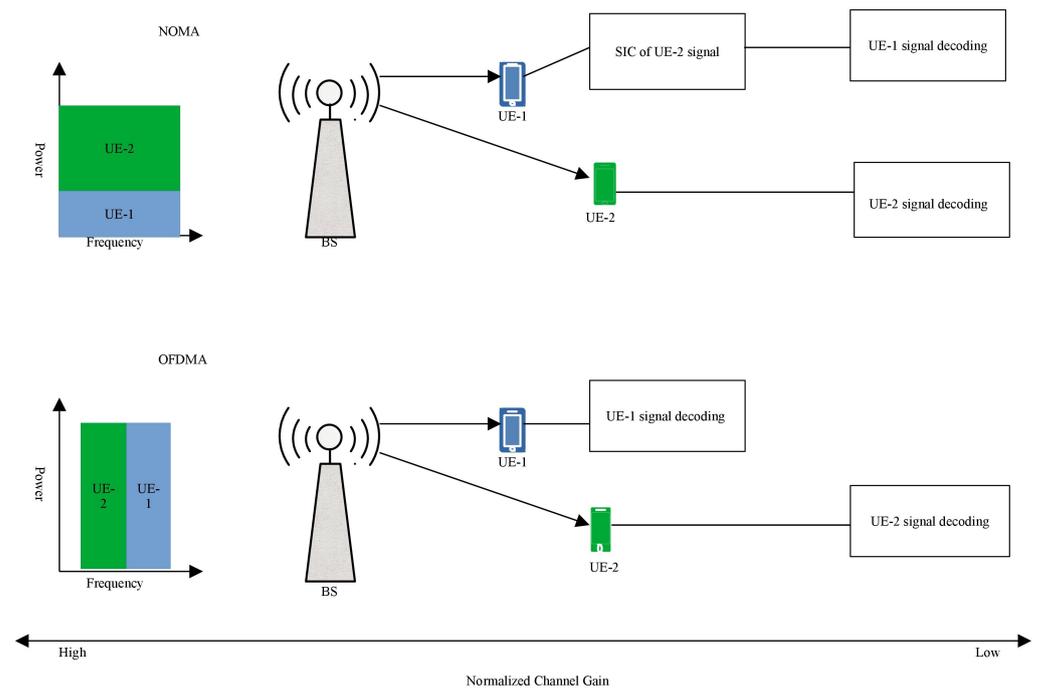


Figure 6. A comparison between basic downlink NOMA and OFDMA.

In the downlink NOMA, the transmitter uses superposition coding (SC) to combine multiple signals, while the receiver employs successive interference cancellation (SIC) for decoding and removing the interference from other signals. Such techniques allow utilization of the same spectrum by all users [48,49]. SIC is also involved in uplink transmission [49].

The goal of SC is to transmit two messages concurrently by combining them into one signal with two layers. The source node generates two varying messages: the base message and the superposed message. Assume a BS that is communicating with two end users in a downlink communication. The messages are broadcast to two receivers. The receiver with the strongest channel can decode both messages, while the receiver with the worse channel can decode only the base message [49,50].

SIC is a physical technique that enables the simultaneous decoding of messages at the receiver. It allows for the concurrent processing of multiple signals. Receivers implementing SIC decode stronger signals (one at a time), subtract (cancel) them from the combined signals, and extract the weaker signals from the remaining remnant [49,51].

The NOMA variant that employs both SC and SIC is known as the power domain NOMA (PD-NOMA). It was suggested as the long-term evolution of the third-generation partnership project [49].

Overloading allows NOMA to accommodate multiple transmissions simultaneously in the same time-frequency resource block by assigning different codes to different users and adopting a unique user-specific spreading sequence [6,7]. This concept is inspired by classic code division multiple access (CDMA) systems.

Code domain NOMA encompasses various NOMA schemes that have been developed using this approach, such as low-density spreading (LDS) CDMA, LDS orthogonal frequency division multiplexing, sparse code multiple access, and multi-user shared access [6].

Linear transform decoding defines NOMA that relies on the multi-user detection complexity. Within an OMA scheme, this approach is employed to segregate the signals of distinct users into orthogonal subspaces utilizing a linear transform. Schemes that do fit this description are categorized as NOMA [52].

From an information theoretical view, NOMA refers to any technique that permits simultaneous transmission over the same time, frequency, code, or space resources to

achieve a superior rate region in comparison to the orthogonalization of one or more of these resources. This includes SC and SIC, rate-splitting, and dirty paper coding [49].

The usage of the various techniques presented above provides many different variants of NOMA. Two principal categories of NOMA exist: PD-NOMA and code domain NOMA [6]. Figure 7 shows a simple overview of NOMA variants.

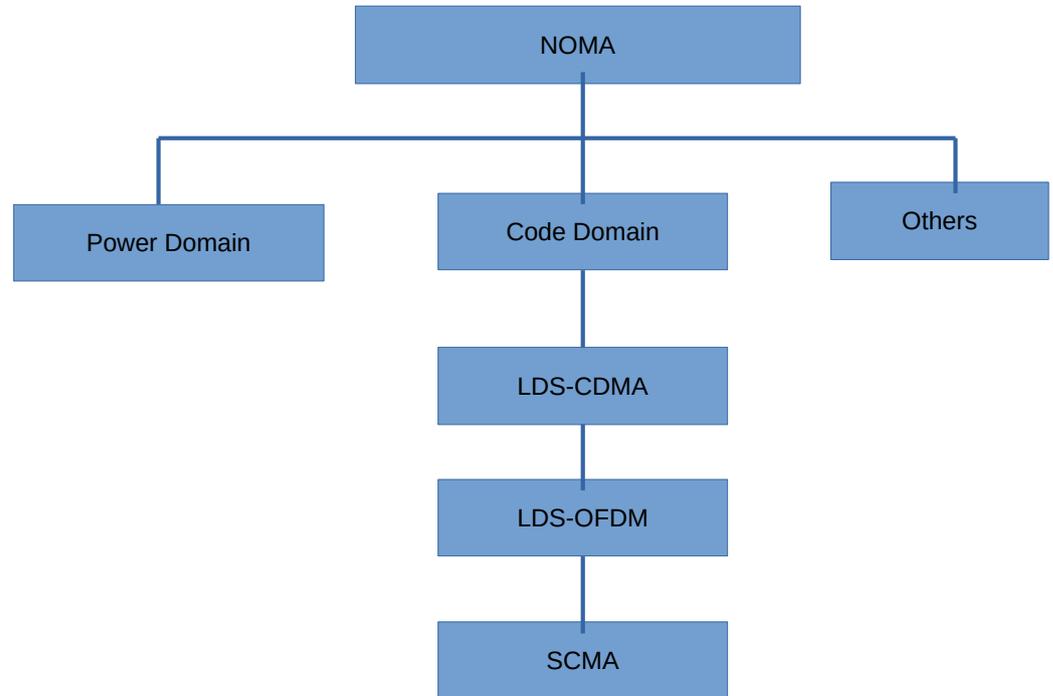


Figure 7. An overview of NOMA variants [6].

To understand the use of SIC with NOMA, consider Figure 8. It shows the basic NOMA scheme when utilizing SIC at the UE receivers in the cellular downlink. The total system transmission bandwidth is considered to be 1 Hz. The BS sends a signal intended for both Users 1 and 2, with $x_i (i = 1, 2)$ and $E[|x_i|^2] = 1$. The signal has transmission power P_i with a maximum total of P [6]. With the superposition coding of x_1 and x_2 , the transmitted signal is represented as

$$x = \sqrt{P_1}x_1 + \sqrt{P_2}x_2. \quad (18)$$

The signal received by UE- i can be represented as

$$y_i = h_i x + w_i. \quad (19)$$

where h_i represents the complex channel coefficient within UE- i and the BS, and w_i is the receiver Gaussian noise, which includes inter-cell interference. $N_{0,i}$ denotes the power density of w_i .

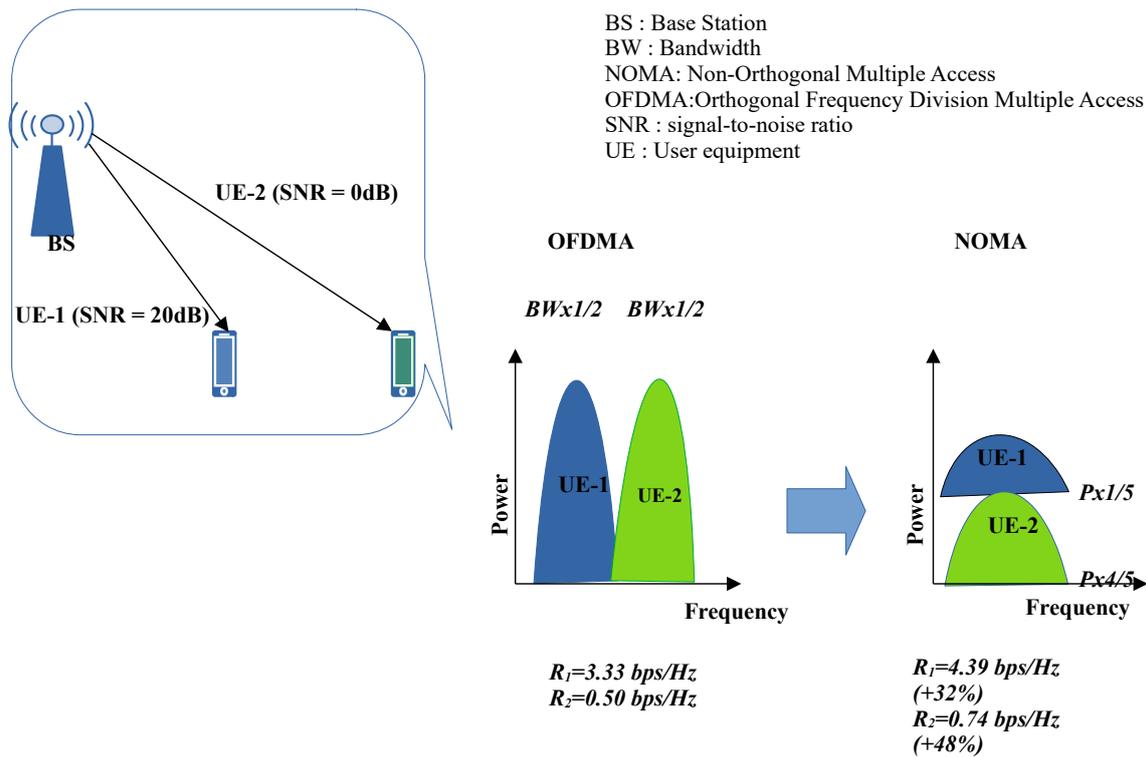


Figure 8. A simple example of OFDMA versus NOMA.

In the NOMA downlink, the SIC method is applied at the receiver of the UE. The optimal decoding order is given in order of increasing channel gain, normalized by the noise and the inter-cell interference power, represented as $\frac{|h_j^2|}{N_{0,i}}$. Following the pattern, a user may perfectly decode the signals of any other user that is earlier in the decoding order for interference cancellation. Then, UE-*i* eliminates the inter-user interference from the *j*th user, where $\frac{|h_j^2|}{N_{0,i}}$ is less than $\frac{|h_i^2|}{N_{0,i}}$. For the case of two UEs, UE-2 is not performing interference cancellation because it appears in front in the decoding sequence, $\frac{|h_1^2|}{N_{0,1}} > \frac{|h_2^2|}{N_{0,2}}$. UE-2 first decodes x_2 , and then cancels the x_2 component from the signal received, y_1 . Thereafter, UE-1 can decode x_1 without interference from x_2 . Assuming error-free detection of x_2 at UE-1, the throughput of UE-*i*, R_i , is expressed as

$$R_1 = \log_2 \left[1 + \frac{P_1|h_1|^2}{N_{0,1}} \right]. \tag{20}$$

$$R_2 = \log_2 \left[1 + \frac{P_2|h_2|^2}{P_1|h_2|^2 + N_{0,2}} \right]. \tag{21}$$

This can be compared with OFDMA, if OFDMA is assumed to use orthogonal user multiplexing, with a bandwidth of $\alpha(0 < \alpha < 1)$ Hz allocated to UE 1 and the rest $(1 - \alpha)$ Hz allocated to UE 2. R_i is represented as

$$R_1 = \alpha \log_2 \left[1 + \frac{P_1|h_1|^2}{\alpha N_{0,1}} \right]. \tag{22}$$

$$R_2 = (1 - \alpha) \log_2 \left[1 + \frac{P_2|h_2|^2}{(1 - \alpha)N_{0,2}} \right]. \tag{23}$$

When compared to OFDMA, the performance gain of NOMA becomes significantly greater when there is a disparity in channel gains. Figure 8 shows that this NOMA scheme can achieve higher rates than OFDMA [5]. The figure assumes a scenario with two UEs: one located at the cell interior and the other at the cell edge. In this case, the $P_1|h_1|^2/N_{0,1}$ and $P_2|h_2|^2/N_{0,2}$ are considered to be 20 and 0 dB, respectively. This choice of values demonstrates the potential performance of the NOMA scheme in leveraging the natural differences in user channel gains. By effectively utilizing the near–far effect, higher spectral efficiency can be achieved, leading to improvements in the overall system capacity and the rate of the cell-edge user [5].

Figure 9 shows the boundaries of the rate regions achievable with superposition coding and optimal orthogonal schemes for an asymmetric downlink additive white Gaussian noise (AWGN) channel for Equations (22) and (23), (with $SNR_1 = 0$ dB and $SNR_2 = 20$ db). We observe that the results of the superposition coding outperform that of the orthogonal scheme [53].

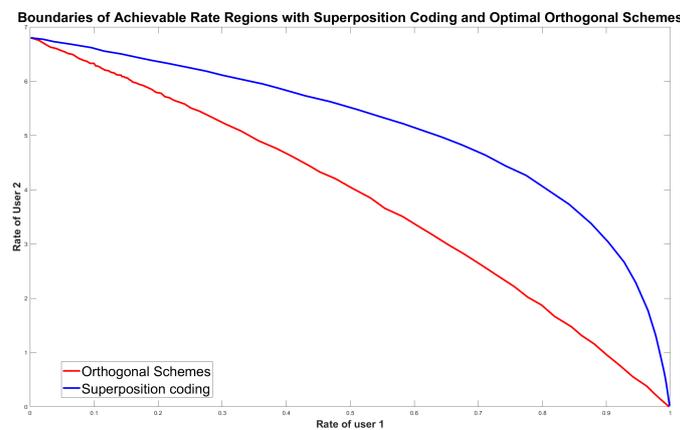


Figure 9. Boundaries of the rate regions achievable for two-user downlink asymmetric AWGN, using superposition coding (solid line) and orthogonal schemes.

To investigate the practical performance improvement of NOMA, the authors of [6] conduct a multi-cell system-level simulation and evaluate the performance gain of NOMA using wideband scheduling and power allocation (PA). The wideband case is chosen because the system performance is independent of the frequency-selective channel information. This is significant in practical wide-area deployments. To further illustrate this, Figure 10 shows a cumulative distribution function of the user throughput for OFDMA and NOMA with SIC. The results clearly demonstrate that the user throughput of NOMA is approximately 27% greater than that of OFDMA for both cell and cell-edge users.

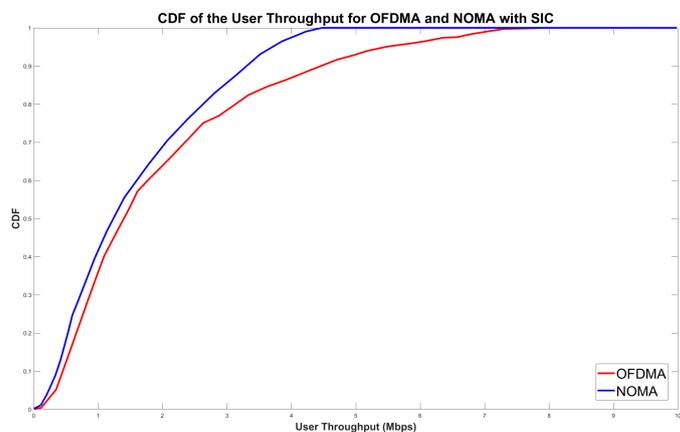


Figure 10. A system-level evaluation of OFDMA and NOMA when using wideband scheduling and power allocation.

2.6. Clustering and User Pairing in NOMA

The performance of NOMA systems depends on favorable SIC at user nodes which, however, is dependent upon user pairing or user clustering (UC) schemes [54]. The UC schemes target to lower SIC complexity and cross-interference to possible extents by classing the network into smaller NOMA clusters which lessen and guarantee the overall system's maximum sum-rate throughput [55]. In [56], the authors investigate how user pairing affects sum rate and individual user rates. The results show that the sum rate enlarges by scheduling two users whose channel conditions differ considerably. The critical issues for UC in NOMA is examined by the authors of [57]. In downlink NOMA, it is advantageous to disperse the users with better channel gains in a cell into distinct clusters as this will greatly enhance the overall sum throughput of the cluster. Pairing users with low channel gains with those having high channel gains is beneficial for increasing the overall throughput. High channel gain users may attain better rates even at lower power levels while allocating a significant portion of power to weaker users. Finally, the throughput of the remaining users in the NOMA cluster is predominantly influenced by the distribution of transmit power levels. For uplink NOMA, users experience various channel gains; hence, power control will result in sum throughput degradation. Diversity plays a crucial role in minimizing inter-user interference and thence maximizing the cluster throughput. Finally, including high channel gain users operating at their maximum power levels in each cluster is beneficial as they can make a significant contribution to the overall throughput.

Clustering in massive MIMO with NOMA usually is divided into two main approaches: Joint Resource Aware User Clustering and Learning-Assisted User Clustering [58].

i. Joint Resource Aware User Clustering

This entails the following:

- (a) Clustering based on the channel correlation coefficient considers interference-aware approaches for both cellular and device-to-device (D2D). The BS constructs a channel graph for mobile users by evaluating the channel correlation among them, while D2D users are partitioned into separate clusters using an interference graph. After completing the construction of channel and interference graphs, cluster matching is performed to find the optimum match of each D2D paired cluster and its corresponding mobile user cluster.
- (b) Energy-efficient user admission-based clustering focuses on maximizing energy efficiency in current and future systems. It involves clustering multiple users using efficient power allocation (PA) techniques. The goal is to admit users gleaned from their predefined Quality of Service (QoS) needs, with a focus on energy efficiency. Users are admitted in increasing order of the amount of power needed to meet respective QoS needs.
- (c) Joint user clustering and power allocation aims to elevate the sum rate by selecting the two best users in a cluster. It follows a constant transmit power allocation method; it derives a minimal distance metric to differentiate between the cell center and edge users. For the uplink scenario, a unique cluster creation and power regulation technique can be employed. In the uplink, the clustering strategy involves designing high-power and low-power clusters, dividing active mobile users based on composite channel gains instead of individual gains.
- (d) Spatial position-based clustering offers a reduced complexity compared to alternative schemes. The clusters are created based on the spatial positions of the users, which are determined through the use of a global position-tracking system. Users with closer spatial proximity are grouped, and a multi-antenna cluster head is chosen for their service. However, it is important to note that without employing additional interference mitigation techniques, this approach may suffer from significant performance deterioration regarding the achievable sum rate.

ii. Learning Assisted User Clustering

It entails the following:

- (a) K-means is an unsupervised machine learning algorithm with low complexity. It starts by randomly selecting initial cluster heads for predetermined clusters, and then each user is assigned to the nearest cluster head. The associated cluster members then change the head position by taking the average for all users.
- (b) K-means ++ is highly responsive to the choice of starting centroids, as any inconsistencies in the initial centroids or the presence of multiple centroids in the same cluster can lead to ineffective user grouping.
- (c) Fuzzy C-means is an unsupervised clustering algorithm utilized for feature analysis, enabling the classification of users into multiple clusters. First, the cluster members and fuzzy exponent are initialized. Then the membership function is allocated to every user to establish their proportional relationship to the group. The grouping is iteratively generated by revising the association factor of every data point, aiming to minimize the objective function below a given limit.
- (d) Clustering as a multi-level classification problem is a multi-level classification problem developed by adopting a gradient-boosted decision tree and a sorting network as base modules. Members are first divided into two distinct categories and then sorted so that successive members exhibit great spatial correlation. The objective is to assign members to distinct clusters to minimize the sum rate.

While these techniques have their challenges, the purpose of UC is to mitigate multi-cluster and intra-cluster interference to facilitate the transmission of the desired signal. This involves determining the optimal amount of clusters and, consequently, the users within each cluster.

As long as the users in a cluster are determined, the process of user grouping is carried out. It is inadvisable to ask NOMA users to conduct NOMA jointly as NOMA is limited by interference. Therefore, users might be allocated into various categories where NOMA will be deployed. Several user-pairing algorithms are based on desired performance, deployment environment, and implementation complexity. User pairing should provide high throughput with minimum computational complexity while ensuring user fairness. With an increasing number of UEs, UE pairing becomes challenging due to significant CSI acquisition along with feedback overhead, in addition to the need to execute complicated algorithms [57,59].

3. NOMA-Aided CF-mMIMO

Despite its advantages and potential, the literature features little work on NOMA in CF-mMIMO (CF-mMIMO-NOMA) systems. In the following part, we will go over the system, channel, and signal models and finally survey the state-of-the-art literature that is available for CF-mMIMO.

3.1. System, Channel, and Signal Models

In the system and channel model, we focus on a downlink communication of a CF-mMIMO-NOMA system where M single antenna APs serve KN single antenna UEs. The users are spatially dispersed in N groups using the TDD protocol at the same time-frequency resource block. Every cluster comprises K users. The APs communicate to the CPU through a fronthaul network with unlimited capacity, ensuring perfect and error-free communication. Figure 11 shows a CF-mMIMO-NOMA network with two users per cluster.

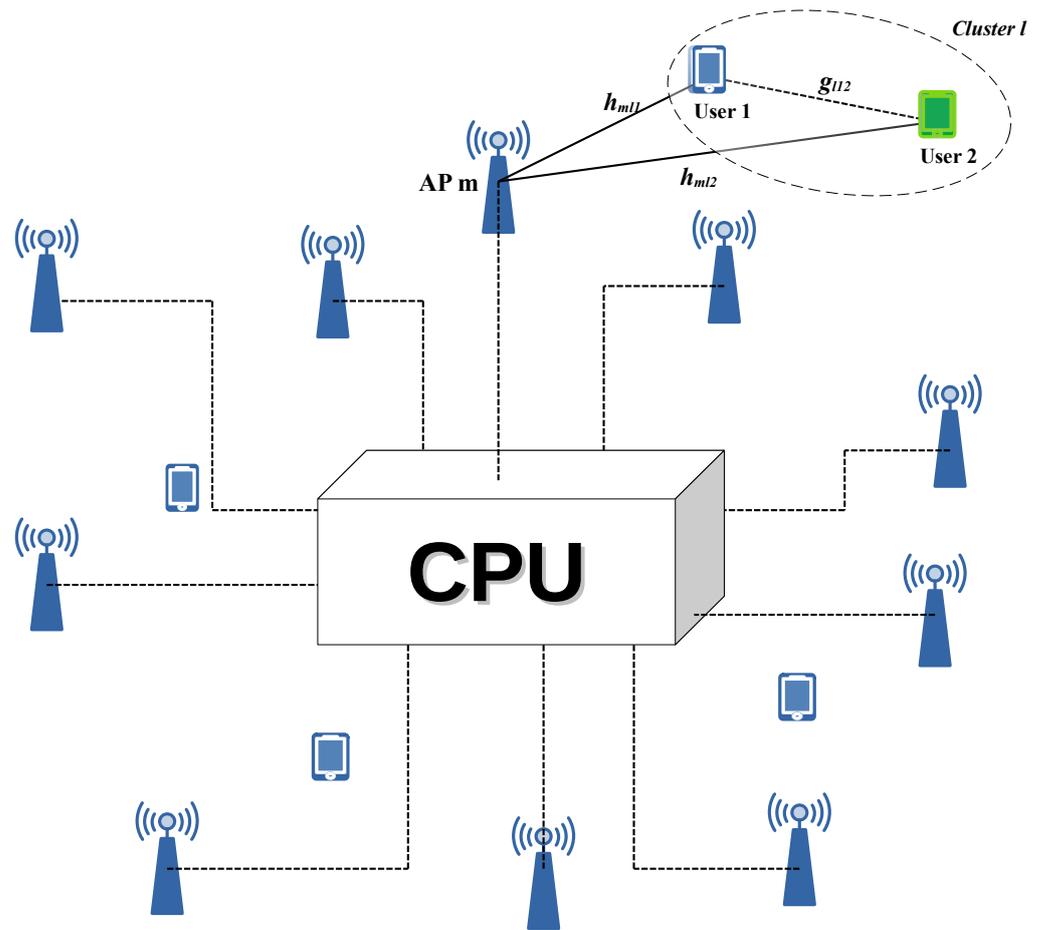


Figure 11. A CF-mMIMO-NOMA network with two users per cluster.

The downlink channel between the m th AP and the k th user in the n th cluster, where $m \in \{1, \dots, M\}$, $k \in \{1, \dots, K\}$, and $n \in \{1, \dots, N\}$, is presented as

$$h_{mnk} = \zeta_{mnk}^{1/2} \tilde{h}_{mnk}. \tag{24}$$

here, ζ_{mnk} represents large-scale fading and $\tilde{h}_{mnk} \sim \mathcal{CN}(0, \tilde{h}_{mnk})$ is circularly symmetric Gaussian distributed with zero mean and unit variance. The former is considered as known in advance because it varies very gradually over time. Therefore, it should be approximated after every 40 coherence time intervals. The latter captures the effect of quasi-static Rayleigh fading [60]. Thus, $h_{mnk} \sim \mathcal{CN}(0, \tilde{h}_{mnk})$ is a complex normal random distribution variable with zero mean and covariance ζ_{mnk} [61].

Moreover, users belonging to a particular NOMA clustering are organized based on their respective channel environments:

$$\sum_{m=1}^M |h_{mn1}|^2 \geq \sum_{m=1}^M |h_{mn2}|^2 : 1 \leq n \leq N. \tag{25}$$

Notably, the channel reciprocity assumption of the TDD protocol allows the downlink channel to be approximated by uplink pilots. As a result, the initial section of every coherence block is dedicated to uplink pilots, while the remainder is dedicated to data transmission. The authors of [62] show that the assumption of channel reciprocity does not consider hardware mismatches between the AP and UE. Such mismatches act against the reciprocity assumption, and impact system performance.

3.1.1. CSI Acquisition and Uplink Pilot Training

The APs conduct channel estimation for the uplink channels utilizing user-transmitted pilot signals. For the downlink channel estimation, TDD channel reciprocity is utilized. To reduce the channel estimation overhead, users in a particular cluster are assigned a similar pilot sequence, which has a period τ symbols. The N pilot sequences dedicated to the N clusters are mutually orthogonal, hence, $\tau \geq N$. The pilot sequence dedicated to the K users in the n th cluster is denoted as $\phi_n \in \mathbb{C}^{\tau \times 1}$ and satisfies $\|\phi_n\|^2 = 1$ and $\phi_n^H \phi_l = 0$ for $n \neq l$. During the uplink channel estimation, the m th AP obtains a pilot signal given by

$$Y_m^P = \sqrt{\tau p_p} \sum_{n=1}^N \sum_{k=1}^K h_{mnk} \phi_n + n_m. \tag{26}$$

where p_p represents the pilot transmit power (uplink normalized SNR) and $n_m \sim \mathcal{CN}_{\tau \times 1}(O_{\tau \times 1}, I_i)$ denotes the AWGN vector at the m th AP.

For the reason of channel estimation h_{mnk} , the pilot signal received by the m th AP y_m^P is extended onto ϕ_n as

$$\tilde{Y}_{mn}^P = \phi_n^H y_m^P = \sqrt{\tau p_p} \sum_{k=1}^K h_{mnk} + \phi_n^H n_m. \tag{27}$$

If the pilot sequences happen to be similar or orthogonal, \tilde{y}_{mn}^P serves as an adequate statistic. In this case, the MMSE estimate of h_{mnk} provided by \tilde{y}_{mn}^P is

$$\tilde{h}_{mnk} = \frac{\mathbb{E}[\tilde{y}_{mn}^{P*} h_{mnk}]}{\mathbb{E}[|\tilde{y}_{mn}^P|^2]} \tilde{y}_{mn}^P = \frac{\sqrt{\tau p_p} \zeta_{mnk}}{1 + \tau p_p \sum_{k=1}^K \zeta_{mnk}} \tilde{y}_{mn}^P. \tag{28}$$

By using the fact that \tilde{y}_{mn}^P is a Gaussian distribution, \tilde{h}_{mnk} may be denoted as

$$\tilde{h}_{mnk} = \sqrt{\eta_{mnk}} v_{mn}. \tag{29}$$

where $v_{mn} \sim \mathcal{CN}(O, 1)$, and η_{mnk} is defined as

$$\eta_{mnk} = \mathbb{E}[|\tilde{h}_{mnk}|^2] = \frac{\tau p_p \zeta_{mnk}^2}{1 + \tau p_p \sum_{k=1}^K \zeta_{mnk}}. \tag{30}$$

The channel estimation error is defined as $\epsilon_{mnk} = h_{mnk} - \tilde{h}_{mnk}$, where ϵ_{mnk} and \tilde{h}_{mnk} are statistically independent. Supplementarily, $\mathbb{E}[|\epsilon_{mnk}|^2] = \zeta_{mnk} - \eta_{mnk}$. Each AP serves each user in its cluster using power control coefficient η_{mnk} . The CPU computes these coefficients, which are then transmitted to the APs via the fronthaul networks.

3.1.2. Signal Model

The APs utilize a conjugate beamformer for downlink data transmission. The beamformer is designed using the locally determined CSI, which is obtained from uplink MMSE channel estimation, using the principle of channel reciprocity.

The K users' data signal in the n th cluster is superposition coded as

$$x_n = \sum_{k=1}^K \sqrt{P_{nk}} x_{nk}. \tag{31}$$

where x_{nk} and P_{nk} are the data signal and transmit power, respectively, assigned to the k th user in the n th cluster for $n \in \{1, \dots, N\}$ and $k \in \{1, \dots, K\}$. In addition, for x_{nk} and x_{ml} , $m, n \in \{1, \dots, N\}$ and $k, l \in \{1, \dots, K\}$, x_{nk} and x_{ml} satisfy

$$\mathbb{E}\{x_{nk} x_{ml}\} = \begin{cases} 1, & \text{if } n = m \text{ and } k = l \\ 0, & \text{Otherwise.} \end{cases} \tag{32}$$

Hence, the expected value of the squared magnitude of $x_n(\mathbb{E}[|x_n|^2])$ is equal to the summation of the transmit powers given to the users in the n th cluster ($\sum_{k=1}^K P_{nk} = P_n$). P_n signifies the overall signal power assigned to the n th cluster. The signal sent at the m th AP is expressed as

$$t_n = \sum_{n=1}^N \frac{\tilde{h}_{mnk}^*}{\tilde{h}_{mnk}} x_n = \sum_{n=1}^N \frac{v_{mn}^*}{v_{mn}} x_n. \tag{33}$$

where the conjugate beamformer is designed using the short-term power constraint. The mean transmit power at the m th AP considering all N clusters is denoted as P_{tm} and can be calculated as the sum of the individual signal powers allocated to each cluster P_n . In the CF-mMIMO-NOMA system, all KN users belonging to the N clusters get services simultaneously from the M APs. As a result, the signal acquired by the k th user in the n th cluster may be expressed as follows:

$$\begin{aligned} y_{nk} &= \sum_{m=1}^M h_{mnk} + n_{nk} \\ &= \underbrace{\sqrt{P_{nk}} c_{nk} x_{nk}}_{\text{Desired Signal}} + \underbrace{c_{nk} \sum_{k'=1, k' \neq k}^K \sqrt{P_{nk'}} x_{nk'}}_{\text{Intra-cluster Interference before SIC}} \\ &\quad + \underbrace{\sum_{n'=1, n' \neq n}^N c_{n'k} x_{n'}}_{\text{Inter-cluster interference}} + \underbrace{n_{nk}}_{\text{AWGN}}. \end{aligned} \tag{34}$$

where $c_{nk} = \sum_{m=1}^M h_{mnk} \frac{v_{mn}^*}{|v_{mn}|}$, $c_{n'k} = \sum_{m=1}^M h_{mn'k} \frac{v_{mn}^*}{|v_{mn}|}$, and $n_{nk} \sim \mathcal{CN}(0, 1)$.

In PD-NOMA, the power allocation strategies assign higher powers to the users with weaker channel strengths, resulting in a power ordering of $P_{n1} \leq \dots \leq P_{nk} \leq \dots \leq P_{nK}$ within the n th cluster. This power allocation scheme ensures that within each cluster, the k th user is capable of effectively decoding the message set for the l th user given that the k th is able to decode its own message. Consequently, the k th user can progressively eliminate the intra-cluster interference originating from the l th user where l is greater than k , before decoding its signal. The k th user considers signals from users with indices $l \geq k$ as interference. For brevity, we leave optimal PA and optimal UC as open problems.

It is important to note that for TDD CF-mMIMO, the user terminals cannot use real-time CSI. Nonetheless, with an increasing quantity of APs, the channel conditions become more stable. Therefore, the expected value of c_{nk} can be employed as an approximation of the efficient channel gain for decoding x_n at the k th user in the n th cluster. In practice, acquiring the expected value of c_{nk} is not challenging, given that it is dependent solely on the statistical features of the channels, and stays constant over numerous coherence intervals. Perfect successive interference cancellation (SIC) is not feasible as a result of intra-cluster pilot contamination, channel estimation errors, and limited statistical CSI information of the users. Following imperfect SIC, a post-processed message in the k th user terminal in the n th cluster may be described as

$$\begin{aligned} \tilde{y}_{nk} &= \sqrt{P_{nk}} c_{nk} x_{nk} \\ &\quad + \underbrace{c_{nk} \sum_{k'=1}^{k-1} \sqrt{P_{nk'}} x_{nk'}}_{\text{Inherent intra-cluster interference after SIC}} + \sum_{n'=1, n' \neq n}^N c_{n'k} x_{n'} \\ &\quad + \underbrace{\sum_{k''=k+1}^K \sqrt{P_{nk''}} [c_{nk} x_{nk''} - \mathbb{E}[C_{nk}] \hat{x}_{nk''}]}_{\text{error propagation or residual interference due to imperfect SIC}} + n_{nk}. \end{aligned} \tag{35}$$

In this equation, \hat{x}_{nk} represents an estimate of x_{nk} for all n, k , and x_{nk} is supposed to be taken from a Gaussian distribution with zero-mean and unit variance. Thus, \hat{x}_{nk} and x_{nk} are mutually Gaussian distributed with a normalized correlation coefficient ρ_{nk} :

$$x_{nk} = \rho_{nk}\hat{x}_{nk} + \eta_{nk} \tag{36}$$

where $x_{nk} \sim \mathcal{CN}(0, 1)$, $\eta_{nk} \sim \mathcal{CN}\left(0, \frac{\sigma_{\eta_{nk}}^2}{[1+\sigma_{\eta_{nk}}^2]}\right)$, and $\rho_{nk} = \frac{1}{\sqrt{1+\sigma_{\eta_{nk}}^2}}$. Furthermore, x_{nk} and e_{nk} are statistically independent. From [60], the upper bound ergodic sum rate is given by

$$\tilde{R} < \bar{R} = \phi \sum_{n=1}^N \sum_{k=1}^K \mathbb{E}[\log(1 + \gamma_{nk})] \tag{37}$$

where \tilde{R} represents the achievable sum rate of CF-mMIMO-NOMA, and expressed as

$$\tilde{R} = \sum_{n=1}^N \sum_{k=1}^K \tilde{R}_{n,k} \tag{38}$$

Figure 12 shows a plot of the achievable sum rate of various NOMA systems when compared to an OMA counterpart. The plot shows that NOMA offers better performance for CF-mMIMO, particularly with a greater amount of users.

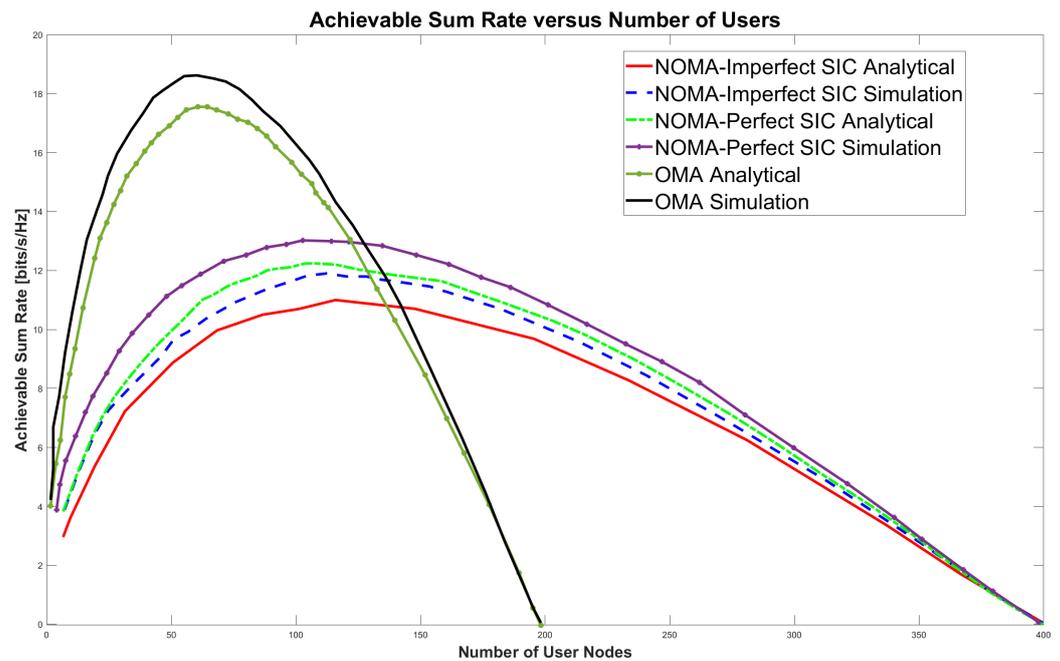


Figure 12. The achievable sum rate of various NOMA and OMA systems versus the number of users.

3.2. State-of-the-Art

CF-mMIMO-NOMA was first introduced in [60]. An achievable sum rate of PD-NOMA is obtained considering the effects of intra-cluster pilot contamination, inter-cluster interference, and imperfections in SIC. The system model entails a downlink communication with single antenna APs and users, operating using TDD protocol. The APs employ conjugate beamforming. Numerical evaluations demonstrated better results of NOMA in comparison to OMA. The authors of [63] were the first to investigate the CF-mMIMO-NOMA uplink. They derived a comprehensive analytical expression for the SE of conjugate beamforming that acknowledges the effect of intra-cluster pilot contamination, inter-cluster interference, imperfect SIC, channel estimation errors, and power optimization. By employing an iterative geometric programming (GP) algorithm based on sequential convex

approximation (SCA), the derived expression aims to improve spectral efficiency (SE). Through simulation results, it was revealed that CF-mMIMO-NOMA outdoes cell-free massive MIMO with OMA (CF-mMIMO-OMA) in terms of efficient spectrum utilization despite its limited availability. Also, with power optimization, the uplink spectral efficiency is maximized. Identical to [64,65], the authors suggest an optimal backhaul combining (OBC) that improves the uplink signal-to-interference noise ratio (SINR). The findings indicate that by ensuring that the total number of users allocated to every pilot symbol is no greater than the total number of BSs, the correlated interference can be successfully reduced. Moreover, by adopting OBC CF-mMIMO-NOMA overperforms CF-mMIMO-OMA in terms of both maximizing the minimum QoS and enhancing connectivity.

From the perspective of communication protocol, all recent works have used TDD; none of them have considered FDD as a transmission protocol. Although the majority of the literature focuses on single-antenna APs, Refs. [66–69] consider multi-antenna APs. Most of the literature assumes a Rayleigh fading channel. That ignores the presence of spatial correlation, while channels often exhibit a combination of small-scale fading due to NLoS propagations and static Line-of-Sight (LoS). These characteristics can be effectively captured by modeling the channel as Rician fading. The authors of [69–71] model the channel using Rician fading. Specifically, Ref. [69] considers a system with multiple-antenna APs communicating with single-antenna UEs across spatially correlated Rician fading channels.

While most works use conjugate beamforming precoders at the APs, the authors of [64,67] employ MRT precoders, the authors of [72] employ full-pilot zero-forcing (fpZF) precoders, and the authors of [68] compare the performance of all three practical linear decoders. Specifically, the authors of [68] thoroughly analyze the system performance using MRT, fpZF, and modified regularized ZF (mRZF). Furthermore, the authors obtain a closed-form expression for the sum rate given a Rayleigh fading channel and accounting for the impact of intra-cluster pilot contamination, inter-cluster interference, and imperfect SIC. The analytical results show that MRT with perfect SIC is outperformed by mRZF and fpZF, despite possessing the same front-hauling overhead. The highest rates were achieved by mRZF because it attempts to strike a balance between mitigating inter-cluster interference and enhancing intra-cluster power. However, in regions with higher user numbers, MRT surpasses fpZF. Furthermore, the authors of [73] derive a closed-form SINR using both conjugate beamforming and normalized conjugate beamforming to maximize bandwidth efficiency. The findings demonstrate the superior performance of conjugate beamforming compared to normalized conjugate beamforming. Finally, Ref. [71] compares the performance of MR precoding to L-MMSE precoding in different channel estimators. The results show that L-MMSE offers better performance than MR precoding while, among the three estimators (MMSE, EWMMSSE, and LS), MMSE estimation is the best because it utilizes comprehensive statistical channel information. In contrast, EW-MMSE only leverages partial statistical information, and the LS lacks any prior knowledge of channel statistics.

Most authors consider static APs and users. However, the authors of [74,75] consider random APs and users. Specifically, they investigate how the achievable rates of CF-mMIMO-NOMA systems may be improved under random AP or user deployments by using stochastic geometry-based modeling to accurately characterize performance network-wide. The authors evaluate a Poisson point process (PPP) of APs and users using Rayleigh fading and log-distance path loss. Primary observations include a reduced rate gain of NOMA as the density of APs decreases and as path loss exponents increase and the provision of reduced latency by NOMA at the expense of a reduced overall rate.

Two major challenges faced by NOMA are user clustering and user ordering. UC is a key technique that enables the deployment of NOMA for numerous users by lowering the complexity of SIC [73]. The majority of NOMA works from the literature study UC schemes that group two users per cluster with random pairing [73,76]. In [73], three pairing schemes are implemented: far, near, and random pairing. The authors of [67] use the Jaccard index to compute the correspondence between the large-scale fading profile of each user and a

predetermined centroid. Those users with high similarities are assigned to different groups. This is UC via a low complexity optimal method. In [64], the authors introduce an iterative algorithm for user localization, which involves minimizing the correlation coefficients and the large-scale fading profiles of two users in a group. The authors of [77] propose a clustering algorithm for which each AP only requires knowledge of each user's location. The CPU sorts the pairs of closest users into common clusters, thereby significantly reducing the path loss between them. The authors of [78] study a novel UC algorithm that employs cooperative links between users without requiring complex optimization. The algorithm ensures that there are reliable channels between users belonging to a NOMA cluster. The authors of [79,80] study a dynamic user pairing technique where two users who meet the iterative algorithm's criteria are paired regardless of geographical distance. This technique is shown to outperform random pairing, far pairing, and close-pairing strategies. However, the above clustering schemes do not consider any learning features. Moreover, random UC yields a suboptimal solution, whereas exhaustive search methods entail high computational complexity. The authors of [72] propose the use of unsupervised ML for UC algorithms. Specifically, K-means++ and improved K-means++ are suggested as effective approaches to create disjointed clusters of users. The provided numerical results confirm the effectiveness of these algorithms over far, near, and random pairing schemes, and the Jaccard-based UC scheme.

Even though grouping users in clusters in CF-mMIMO-NOMA can significantly serve more users than conventional OMA, it results in a reduced sum rate as a result of inter-cluster interference. To achieve the highest possible rate while ensuring equal service for all users, a hybrid NOMA/OMA mode selection technique based on the channel's statistics is proposed [61,73,76]. This approach, when combined with SIC, demonstrates improved efficiency in comparison to single-mode NOMA or OMA systems.

Different power optimization techniques are used for CF-mMIMO NOMA. The authors of [63] aim to maximize the uplink SE using an iterative GP algorithm relying on the principles of SCA. Likewise, in [66], SCA is used in maximizing the sum SE considering the power constraints of every AP and the SIC. In addition to the max–min power method presented in [64,65], the authors of [81] consider a max–min algorithm with adaptive SIC, while the authors of [82] propose a quasi-concave max–min transmit power control problem, which is solved using the bisection method to obtain the optimal solution. In [72], the sum SE is maximized using an iterative algorithm based on inner approximation. The algorithm accomplishes this by optimizing the normalized transmit power while considering constraints such as the power budget at the APs, the SIC conditions, and all lowest-needed SE at the UEs. In [73,76], the authors formulate a max–min bandwidth efficiency (BE) optimization problem with per-AP power constraints. The authors of [76] conceive a bisection scheme to explore the best approach to this issue. Additionally, the power reduction issues associated with conjugate beamforming are addressed by utilizing second-order cone programming (SOCP) while standard semidefinite programming (SDP) is employed for normalized conjugate beamforming. To further enhance performance, the authors propose a mode-switching scheme dependent on the average BE. The authors of [69] consider a dynamic intra-cluster PA method that produces dynamic power coefficients for different APs. The authors of [61] investigate an adaptive algorithm that dynamically switches between different modes, namely OMA, non-cooperative NOMA, and cooperative NOMA modes. The goal is to enhance the system's achievable sum rate and energy efficiency. Initially, the system operates in OMA mode, before switching to different modes to utilize their benefits as required by the scenario. The system presented in [83] utilizes Dinkelbach's method-based algorithm to solve its non-convex optimization and obtain an optimal solution for maximizing energy efficiency. The algorithm has two layers, with the bottom layer solving the power control optimization. The solution is obtained using the difference of convex functions programming approach. The authors further detail the complexity of the algorithms. The authors of [78,84] consider a power optimization technique for user-centric CF-mMIMO-NOMA. Specifically, the authors of [84] use a particle swarm algorithm to distribute transmit power for the APs. The authors of [78] adopt a joint power optimization technique for CF-mMIMO-NOMA based on AP

selection, in which APs within a single group conducting SC to obtain their superimposed signals employ a common PA. This both enables the CPU to optimize the PA strategy for each AP group individually and reduces the calculations required for joint power optimization.

Finally, while most works compare NOMA to OMA in CF-mMIMO networks, the authors of [85] study sum rate data of a clustered adaptive relay coordinated transmission scheme based on NOMA, and compare it to direct transmission, relay AF transmission, and relay DF transmission schemes. Importantly, while their proposed schemes offer better sum rates, the limited capacity, and energy of relay nodes still present drawbacks to network performance. Also, the authors of [86] consider intelligent reflecting surfaces (IRS) with CF-mMIMO-NOMA and compare the result to CF-mMIMO-NOMA with no IRS. However, from the results, the ergodic performance is reduced as the number of IRSs and APs is increased while at lower numbers of IRSs and APs, the performance outperforms the traditional CF-mMIMO-NOMA with no IRS.

A summary of research papers on CF-mMIMO-NOMA is shown in Table 2.

3.3. Summary

In this section, we have studied the system, channel, and signal model and provided a review of the available literature on CF-mMIMO-NOMA. In general, NOMA with CF-mMIMO has been shown to outperform OMA-based CF-mMIMO in several works in regards to achievable throughput, sum rate, spectral efficiency, and supporting more users. In contrast, when the number of users is low, OMA is superior to NOMA because of factors such as intra-cluster pilot contamination from shared pilots within the clusters, and residual ICI caused by imperfect SIC. To maximize the minimum downlink bandwidth efficiency, NOMA/OMA mode selection is proposed in [73,76]; however, it is dependent both on the length of the channel's coherence time and on the total number of users.

UC in NOMA remains a design challenge compared to OMA since an efficient beam to cover the users in a NOMA cluster must be formed. Several solutions are proposed including deep learning techniques such as K-means ++ providing better performance in comparison to other techniques. Nevertheless, the impact of virtual and actual cluster center users on the possible sum rate is yet to be distinguished and thoroughly examined. In order to address this problem, joint optimization schemes can be implemented to maximize the achieved sum rate [87]. Furthermore, deep learning approaches may be employed to boost the effectiveness of data/model-driven transmitter/receiver solutions for CF-mMIMO-NOMA. These techniques provide valuable tools for optimizing the fundamental compromises involving system variables, such as channel estimation, power allocation, and SIC decoding, through deep learning frameworks.

Table 2. Summary of papers on NOMA-Aided CF-mMIMO.

Ref.	Topic	System Model	Design Objective	Optimization Method	Key Findings	Limitations
[60]	Achievable sum-rate; power domain; NOMA.	Single antenna APs and users with a Rayleigh fading channel operating in TDD mode.	Maximize achievable sum-rate and capacity	Matched filter-based precoding (conjugate beamforming precoders)	When numerous users are simultaneously served, NOMA demonstrates superior sum-rate gain compared to OMA, achieved by grouping users into clusters and utilizing the proposed pilot assignment technique.	Complex to guarantee high reliability with user fairness. Performance degradation by intra-cluster pilot contamination and the residual interference of imperfect SIC for a few users.

Table 2. Cont.

Ref.	Topic	System Model	Design Objective	Optimization Method	Key Findings	Limitations
[61]	User grouping; achievable sum-rate; power efficiency; PD-NOMA.	Downlink system with single antenna APs and users. Homogenous PPPs to model node locations.	Maximize PE and achievable sum rate.	Matched filter-based precoding; adaptive switching algorithm; UC algorithm.	Proposed cooperative CF-mMIMO-NOMA has the most superior performance in comparison to conventional OMA and CF-mMIMO-NOMA systems. The UC procedure offers much stronger performance than previous methods.	The cluster size is directly proportional to the large-scale fading in users and APs, thus reducing performance for large clusters. It also leads to reduced data rates in cooperative links.
[62]	Imperfect SIC; non-reciprocities; hardware mismatches.	Single antenna APs and users with a Rayleigh fading channel operating in TDD mode.	Maximize achievable sum rate.	A use-and-then-forget bound to obtain an achievable downlink rate.	NOMA provides improved user fairness. AP phase mismatch and imperfect SIC have a similar effect on the achievable sum rate.	Considers fixed user distribution.
[63]	Maximize sum-rate; SE; uplink.	APs equipped with antennas. Users are randomly grouped into clusters.	Maximize SE.	Conjugate beamforming receiver; an SCA-based GP algorithm.	For a few users, OMA performs better than NOMA, but for many users, NOMA is more efficient in utilizing spectrum bands. The proposed algorithm greatly enhances the system.	NOMA's performance is degraded for a small number of users due to pilot contamination and residual intra-cluster interference caused by imperfect SIC.
[64,65,81]	SE; optimal combining; power control. SIC	Uplink transmission with a Rayleigh fading channel.	Maximize the uplink SINR.	The maximum ratio combining technique; an OBC method; zero-forcing backhaul combining (ZFBC) [64]. An adaptive SIC method [81].	OBC outperforms both equal gain combining and ZFBC. With the use of OBC, CF-mMIMO-NOMA is superior to CF-mMIMO-OMA in both max-min QoS and connectivity. The proposed method outperforms conventional SIC methods, particularly for higher QoS values [81].	With an increasing number of antennas on every BS, the impact of correlated interference becomes more significant, leading to SINR degradation in both OBC and ZFBC.
[66]	Power optimization, Sum SE	Downlink multi-antenna APs and single antenna users with a Rayleigh fading channel.	Maximize SE.	Conjugate beamforming; SCA-based sum-SE approximation algorithm.	NOMA exploits the limited spectrum bands better than OMA. The proposed sum-SE maximization algorithm boosts sum-SE by exploiting NOMA.	In the training phase, SIC operation is not performed. QOS is not guaranteed.

Table 2. Cont.

Ref.	Topic	System Model	Design Objective	Optimization Method	Key Findings	Limitations
[67]	Achievable rate; PD-NOMA.	Multi-antenna APs and single-antenna users in a Rayleigh fading channel.	Maximize sum rate.	MRT beamforming. UC algorithm based on the Jaccard distance coefficient.	Cognitive CF-mMIMO with NOMA supports more secondary users, improving the sum rate via channel gain difference.	NOMA strikes a balance in performance and complexity. It adds hardware complexity as a result of SIC processing and error propagation.
[68]	AP precoding; sum-rate.	Downlink system with multi-antenna APs, clusters, and users, with a Rayleigh fading channel.	Maximize sum rate.	MRT; fpZF; mRZF.	mRZF and fpZF achieve better performance than MRT when perfect SIC is available. CF-mMIMO-NOMA using fpZF or mRZF outdoes OMA with MRT.	System performance is degraded by intra-cluster pilot contamination and imperfect SIC, especially with a small number of users.
[69]	Achievable sum-rate; power allocation; channel estimation.	Multi-antenna APs, single antenna UEs, with a Rician fading channel.	Maximize achievable rates.	MMSE estimation; maximum ratio precoding.	More AP antennas and Rician factor boost sum rate; MMSE outperforms EMMSE for correlated Rayleigh fading.	Ignores the impact of self-interference between cluster head antennas.
[70]	Finite block-length coding; statistical delay bounded QoS provisioning	Rician fading channel with randomly located APs and mobile users grouped into clusters.	Maximize achievable data rates.	The Mellin transform; finite blocklength coding.	Proposed CF-mMIMO-NOMA scheme excels in statistical delay and error-rate bounded QoS provisioning in the finite block-length regime.	CSI estimation is insufficient when the number of APs is small.
[71]	Spectral efficiency; IoT; power control; user pairing	Spatially correlated Rician fading channels, randomly distributed multi-antenna APs and single-antenna UEs	Max-min power control	MMSE, EWMSE, and LS estimations with MRT precoding, large-scale-based user pairing, and low-complexity SCA method.	MMSE estimation outperforms EW-MMSE and LS estimations. L-MMSE precoders outperform MR precoding. For many users, NOMA outperforms OMA.	Imperfect SIC and pilot contamination within clusters degrade NOMA performance for a few users.
[72]	User clustering algorithm	APs have multiple antennas, UEs have single antennas.	Maximize sum SE.	fpZF; unsupervised ML-based UC algorithms: K-means++ and Improved K-means++.	MML-based UC algorithm outperforms baseline schemes, with the proffered PA algorithm having faster convergence and better performance than CF-mMIMO-NOMA.	SIC affects the performance of CF-mMIMO-NOMA

Table 2. Cont.

Ref.	Topic	System Model	Design Objective	Optimization Method	Key Findings	Limitations
[73,76]	BE	Downlink NOMA-Aided CF-mMIMO involves the use of APs, clusters, and users.	Maximize BE.	A bisection search method based on conjugate and normalized conjugate beamforming.; use of SOCP and SDP in the iteration of the bisection search.	The switching point between NOMA and OMA modes depends on channel coherence time and the number of users. NOMA with conjugate beamforming outperforms OMA.	The optimal mode switching to maximize the minimum max–min downlink BE depends on the length of the channel coherence time.
[74,75]	Achievable rate; PD-NOMA; user-fairness.	Rayleigh channel with log-distance path loss.	Maximize overall throughput [74]. Maximize achievable rate [75].	Matched filter-based precoding, homogenous PPP to model node locations.	NOMA improves rate performance over OMA in CF-mMIMO with low path loss and high AP density, enhancing user fairness.	NOMA rate gain decreases with lower AP density and higher user PA coefficients, sacrificing the overall rate for lower latency.
[77]	Simultaneous wireless information and power transfer (SWIPT).	A single antenna APs with two users per cluster.	Maximize ergodic sum rate.	Conjugate beamforming.	SWIPT-NOMA outperforms conventional CF-mMIMO-NOMA systems for many users.	Ignores the impacts of self-interference between cluster head antennas. An excessive number of antennas are required for the user.
[78]	AP selection; AP tiered user-grouping.	A single CPU, single antenna APs, and single antenna users.	Maximize achievable sum rate.	Distributed power optimization aided by AP tiering.	The distributed power optimization aided by AP tier sum rate is used in practical deployment to achieve good performance.	Performance depends on AP density.
[79,80]	Channel estimation, user pairing	Considers single antenna users, served with single antenna APs	Max-min downlink rate optimization	Dynamic user pairing, conjugate beamforming.	Dynamic user pairing excels in user rate with acceptable complexity compared to baseline strategies.	Complex to guarantee high reliability with user fairness. Does not consider random AP and user assignments.
[82]	Achievable rate; max–min fairness.	Primary and secondary systems use single-antenna APs and users.	Maximize the achievable sum rate.	Conjugate beamforming; the bisection method; max–min fairness.	NOMA and underlay spectrum boost concurrent user capacity.	Pilot contamination, imperfect SIC, and partial CSI significantly hinder performance.
[83]	Power control.	A network consisting of APs and single-antenna UEs.	Maximize achievable rate.	A two-layer algorithm, based on Dinkelbach's method.	The proposed algorithm outperforms fractional power control and OMA-based CF schemes in terms of energy efficiency.	Intra-cluster pilot contamination and imperfect SIC degrade the system performance of NOMA in the low-user regime.

Table 2. Cont.

Ref.	Topic	System Model	Design Objective	Optimization Method	Key Findings	Limitations
[84]	User-centric, big data; 6G.	APs and mobile users served via a TDD system.	Maximize achievable rate and energy efficiency.	Multiple matching algorithms; particle swarm algorithm; bipartite graph matching	Proposed NOMA-based CF-mMIMO outperforms others in energy efficiency, data rate, and interference.	User QoS is not guaranteed.
[85]	Power optimization; clustering NOMA; multi-access edge computing (MEC).	A cell-free edge network architecture with small BSs.	Maximize sum data rate.	Adaptive relay coordinated transmission scheme; power optimization algorithm based on continuous convex.	The sum rate data of the clustered NOMA-based adaptive relay coordinated transmission scheme is superior to that of the direct, relay AF, and relay DF transmission schemes.	Limited capacity and energy of relay nodes. QoS is not guaranteed.
[86]	Intelligent reflecting surfaces (IRS); SE; User clustering	Single antenna APs and users.	Improve weighted ergodic rate	Distance-aware user clustering, conjugate beamforming, and IRS	Ergodic rate is proportional to the number of IRSs and vice versa.	Interference from APS, IRSs, and IRSs phase shifts degrades performance.
[88]	SE, achievable sum rate.	Single antenna APs and users with a Rayleigh fading channel.	Maximize SE and achievable sum rate.	Conjugate beamforming; SCA-based sum-SE approximation algorithm.	In a fair comparison, NOMA outperforms OMA in both normal and stressed scenarios.	The effect of SIC imperfection on performance.
[89]	SWIPT; layered division multiplexing; energy efficiency.	Downlink backhaul-constrained CF-mMIMO system with SWIPT, which consists of multi-antenna APs, and single-antenna UEs.	Maximize energy efficiency.	A first-order algorithm to find both an initial feasible point and a nearly optimal solution for maximizing energy efficiency; SCA technique; Dinkelbach's algorithm.	Proposed algorithms achieve a similar energy efficiency to second-order approaches with lower computational complexity. The proffered first-order algorithms are better for massive access deployments because of their fast convergence speed and reduced computational complexity.	Doesn't consider the joint optimization of backhaul compression and transmit beamforming for best energy utilization, joint AP clustering and UE scheduling, and beamforming methods, to meet the next huge connectivity in practical applications.
[90]	Compressed Sensing.	Considers uplink transmission with single-antenna users and APs.	Improve system capacity.	Extended approximate message-passing algorithm (EAMP).	EAMP algorithm improves system capacity despite predetermined iteration times and step size, outperforming other algorithms.	Assumes AP receives signals from all users, compromising QoS. Signal recovery is not precise and quick, for a few users.

While most works presented here exploit the spatial domain in designing user grouping and related signal processing techniques, the usage of the angle domain remains a future research direction. Furthermore, the authors of [91,92] have investigated angle-based

processing techniques for FDD-based CF-mMIMO systems, and results show better performance. Consequently, angular models of mMIMO with NOMA channels may be used in designing angle information-aided pilot allocation, channel estimation, beamforming, and PA and interference reduction by array signal processing [93].

Lastly, it is important to acknowledge that none of the implementations of CF-mMIMO-NOMA have been carried out in practical scenarios. Therefore, further contributions and advancements are necessary to connect theoretical developments and practical scenarios in this field.

4. Challenges and Future Research Opportunities

Unresolved issues persist in fully harnessing the integration of CF-mMIMO-NOMA, despite the proposed solutions discussed earlier. In addition, CF-mMIMO-NOMA must be integrated with alternate enabling technologies to adapt to future communication requirements. Therefore, this section addresses several key challenges and open research directions for future work.

4.1. Channel Estimation

Coherent detection requires that the receiver evaluates the channel. Such estimation can be performed for the uplink and then used for the downlink via the reciprocity principle. Many channel estimation algorithms exist. Some techniques use least square approaches: MMSE and linear MMSE. Future networks with numerous users and antennas will pose challenges for channel estimation. Existing high-performance algorithms may lead to excessive signaling overhead and complexity. Hence, new channel estimation techniques are required to strike to balance complexity and performance well. Again, flexible schemes and dynamic protocols should be designed in such a way that they can adapt to varied environments and provide seamless connectivity and network interoperability. Compressive sensing (CS)-based channel estimation can be efficient for massive MIMO [94]. Thus, CS for massive MIMO-NOMA is a potential technique and should be redesigned to embody NOMA features such as the near–far user effect. Deep learning is an approach that can be used for channel estimation. Deep convolutional neural networks can reduce the cost and complexity of channel estimation [95,96]. Estimation techniques based on artificial intelligence (AI) take one of four types: deep convolution neural networks, deep recurrent neural networks, super-resolution technology, and compression sensing technology. ML-based estimators are proposed in [10] and show better performance than conventional methods. Further options for channel estimation exist, depending on their specific application within CF-mMIMO systems [10,11].

4.2. Non-Linear Effects on Signals

As seen in the state-of-the-art, preprocessing such as precoding techniques in cell-free massive MIMO and combining signals with different power levels NOMA can lead to non-linear effects. When these non-linear characteristics are known an appropriate receiver can be designed that takes into account the non-linear effects introduced by the transmitter. To avoid problems resulting from combining signals with different power levels, precoding of the superposition-coded signals to multiple user clusters is necessary. Most of the precoding techniques used in available literature are linear. However, it is worth more to investigate and find more efficient and high-performance non-linear precoders that could be distortion-aware with comparable complexity with linear precoders. For example in [97], proposes a suboptimal yet more practical non-linear precoding scheme that requires a multidimensional integer-lattice least square optimization, which can be found by several approaches. Also, precoders with scalability in terms of complexity could be investigated. Specifically, the use of divide-and-conquer approaches and methods based on sensor-array signal processing could be vital in mitigating the dimensionality of the transmitting-processing problem. There are plenty of opportunities to use AI-based technologies such

as machine learning and deep neural networks to design high-performance and low-complexity precoders [26].

In terms of implementations, most of the discussed precoders above operate at sub-6 GHz bands, which cannot be used in the state-of-the-art architectures with multiple antenna elements and operating at higher frequency bands, millimeter wave [26]. In cell-free massive MIMO, the number of antennas at the transmitter is larger compared to the served users. This makes the number of radio frequency chains (signal mixer and analog-to-digital-converter (ADC)) large hence requiring hundreds of low-cost amplifiers with low-output power. To keep the cost and circuit power consumption low, cost-effective and power-efficient hardware components are employed. This results in hardware impairments that may affect the system's performance. However, some of these impairments can be taken into account in the precoding and signal design, providing schemes with promising spectral efficiency performance.

With 5G and 6G technologies using frequencies above 6 GHz for communications, current transceivers may not be realistic and cost-effective at these frequencies. New transceivers to address these challenges have been proposed. They require joint optimization of precoding weights in digital and analog domains, hybrid precoding [98]. Nevertheless, research is required to ensure the design of sub-optimal, efficient, and practical architectures that improve the joint performance of hybrid decoders.

4.3. Signal Detection

In all the studied works, none focus on signal detection and receive processing. Cost-effective detection algorithms that can perform dimensionality reduction and schemes based on a receiving matched filter with non-linear interference-cancellation capabilities could be investigated. In addition, smarter algorithms with low complexity could be investigated. More importantly, the issue of performance at the cost of complexity is worth investigating. Finally, receiver design with smart signal-processing algorithms that will ensure SIC stability for NOMA signals is important.

4.4. Backhaul/Fronthaul Capacity

CF-mMIMO requires additional overhead to exchange information between APs and the CPU. Unfortunately, practical applications often face limitations in the availability or capacity of the fronthaul. The transmitted signals and structured lattice codes can be quantized to reduce the fronthaul load. Further investigation is needed to tackle the challenge of designing signal processing algorithms that can effectively operate within the limitations of limited fronthaul capacity [11].

The fronthaul can be implemented via wired or wireless connections. Wired connections provide high-capacity links but with high implementation costs and limited scalability [9]. The radio stripes approach has been used to bypass the latter problem for indoor systems. Nevertheless, deploying a wireless millimeter-wave fronthaul network presents the most feasible solution in terms of cost and scalability. Implementable wireless technologies include a microwave-based fronthaul network for backhaul and fronthaul network data transfers, as well as a hybrid millimeter wave or free space optics fronthaul network to meet the high-capacity demands. The authors of [15] also discuss deployment structures for CF-mMIMO and propose a ring deployment structure as a possible solution to offer an alternative transmission path, resulting in more dependable networks at a cheap cost. Nonetheless, empirical studies in this area need to be more mature, and their application scenarios need investigation.

4.5. Hardware Complexity

SIC processing and error propagation in NOMA systems introduces additional hardware complexity. As such, NOMA provides an equal mix of performance and complexity [67]. The resolution of hardware complexity due to SIC detection requires fair PA and

effective UC techniques. A reduction in error propagation requires an improvement in the quality of channel estimation.

Although adding more antennas in CF-mMIMO increases power consumption and hardware complexity, NOMA offers increased performance with more antennas at the BS. In [74,75], the rate gain of NOMA decreases with the density of APs. Additionally, the authors of [64] also consider the influence of adding more antennas on each BS. As the number of antennas increments, correlated interference dominates, prompting a degradation in SINR. In [70], CSI estimation is shown to provide poor results for a small number of APs.

Due to these issues, high priority is placed on research into the effects of hardware components and methods to mitigate the complexities caused by SIC and error propagation. It is important to evaluate the balance between performance and complexity that NOMA offers to show that it is beneficial and can improve overall system performance. Efficient and practical alternatives for SIC applications in transmissions may be discovered utilizing appropriate approaches.

4.6. Multi-Antenna UEs and APs

Most works concerning CF-mMIMO-NOMA consider only single antenna UEs and APs. However, most current and practical APs and UEs feature three or more antennas, and it is expected that future devices will have even more antennas. Generally, multiple antennas can be utilized to achieve either spatial multiplexing, enabling the transmission of multiple data streams to one UE, or improved precoding and combining to mitigate interference [10,99]. Within the literature, only [77] considers a two-antenna UE. As such, the question of resource allocation for multi-antenna UEs remains open.

4.7. Performance in Regions with a Low Number of Users

The performance of NOMA is adversely affected by inter-cluster pilot contamination and imperfect successive interference cancellation (SIC), especially when the number of users is small [63,65,68,72,82,83]. Therefore, methods to address this problem should be investigated. The authors of [73,76] propose an algorithm that switches modes between NOMA and OMA to maximize BE. The success of the suggested solution depends on two factors: the coherence time of the channel and the total number of users. Another approach called Pattern Division Multiple Access (PDMA) hinged on code patterns has been recommended to solve the challenges of propagation error in SIC [94]. However, this technique may increase the error probability of strong users. Given the envisioned intelligent physical layer in 6G, AI-enabled multiple access or similar approaches could be used to mitigate the influence of imperfect SIC and inter-cluster pilot contamination.

4.8. Synchronization

System synchronization poses a significant challenge in CF-mMIMO-NOMA systems due to a large number of APs and users, as well as their random distribution. Synchronization is crucial to ensure coordinated service to users across the entire system. Therefore, it is essential to explore simple, cost-effective, and innovative methods for achieving system synchronization.

4.9. Resource Allocation and Optimization

The Internet of Everything is envisioned to represent the future of the Internet. Radio resources may be limited and require proper resource management to ensure effective utilization. Resource allocation assigns radio resources to devices to boost throughput, data rate, energy efficiency, and user fairness. Key study fields include pilot allocation, power control, and user scheduling of CF-mMIMO systems in mobility scenarios. NOMA efficiently utilizes resources by serving multiple users at distinct power levels. However, resource allocation in NOMA is challenging due to interference from both co-channel and cross-channel transmissions [100].

Moreover, user pairing and optimum PA among NOMA users require a sophisticated algorithm to optimize performance while minimizing resource usage [101]. Different power control algorithms, such as game theory and ML-based PA approaches, can be utilized to address utility optimization problems and achieve better performance. Another important direction for research is the implementation of distributed resource allocation schemes, as they allow easy deployments and computational complexity on distributed units and control units. It also reduces the load on the fronthaul as resource allocation information is not sent to a single network node [14].

It is worth considering techniques for multi-objective optimization of resources such as maximizing capacity, ensuring fairness, and minimizing power consumption. One of the most commonly used multi-objective optimization problems is the min–max method [71,79,80]. Nevertheless, there are instances where conventional optimization models may not be adequate in addressing multi-objective problems, hence the use of other techniques such as game theory [102].

Also, meta-heuristics processes that ensure user QoS and can be tailored to multi-objective optimization are worth investigating. Metaheuristics can be used to solve many types of problems. They include algorithms Ant Colony Optimization, Genetic Algorithms, Iterated Local Search, Simulated Annealing, and Tabu search. They can be categorized into different categories. Among the categories, in CF-mMIMO-NOMA, bio-inspired metaheuristics have been implemented and the authors of [84] investigate particle-swarm to transmit power to different APs. While these metaheuristics are good with large, practical, and/or computationally demanding problems in large spaces, the solutions are not transferable and can not be analyzed numerically [103]. One solution to this would be hybridization. Hybridization of metaheuristics would help exploit the complementary characters of different strategies. This would help in obtaining top performance in solving many hard optimization problems [104].

Finally, while these algorithms and proposals look optimal, the algorithms with low complexities are worth investigating while ensuring good system performance.

4.10. AP-UE Optimization

Many performance optimization strategies have been investigated to improve QoS, the sum rate, and the max–min rate for CF-mMIMO-NOMA. However, these solutions cannot satisfy the network requirements, including diverse QoS requirements for users. To address this, joint optimization of AP and users would be a possible solution. In user-centric CF-mMIMO, a practical and lightweight AP-user association is essential in boosting the system's performance and enhancing spatial reuse. With multiple APs and users, two fundamental problems arise: assigning AP to users and allocating share transmission resources among multiple AP-user pairs. The authors of [87] show that QoS may be assured with comparatively modest cluster sizes that need reduced fronthaul capacity. However, combining AP-centric and user-centric approaches provides better data rates at the expense of higher fronthaul link data traffic [105]. It is worth investigating dynamic and intelligent approaches to AP selection and dynamic clustering that consider changing network demand factors and realistic fronthaul restrictions.

4.11. Interference Management

Interference management directly impacts system performance in mobile communications. The network's performance is limited by inter-user interference, inter-antenna interference, and radio frequency interference caused by numerous users and antennas. The interference model becomes complex due to the presence of more reflecting and scattering paths in the signals. Interference mitigation and management technique solutions are limited to a fixed amount of terminals and antennas. Meanwhile, the CF-mMIMO-NOMA system incorporates a larger number of terminals and antennas. In addition, a proper trade-off is needed between increased resource utility and decreased interference. Therefore, it

is necessary to establish new interference models, analyze common wireless transmission scenarios, and develop improved interference control techniques.

4.12. FDD

In the case of FDD, channel reciprocity cannot be applied as the uplink and downlink channels operate in separate frequency bands. This adds extra overhead for acquiring and providing CSI feedback. However, if the carrier frequencies of the uplink and downlink are within a few GHz of each other and the angle coherence time significantly exceeds the conventional coherence time, the system may still benefit from angle reciprocity, where the channel angle information remains relatively unchanged. All works concerning CF-mMIMO-NOMA assume the use of TDD. A small number of works concerning CF-mMIMO assume the use of FDD. Contemporary mobile communication systems are dominated by FDD. Moreover, FDD may be appropriate for millimeter wave bands because of the reciprocal nature of angular parameters across a wide bandwidth. As such, FDD should be considered for the implementation of CF-mMIMO systems. Currently, few works consider FDD-based CF-mMIMO systems [91,92,106–109].

4.13. Full-Duplex NOMA

Full duplex enables concurrent transmission of both downlink and uplink signals using the same time and frequency resources. This theoretically doubles SE. In addition to improved SE, a full duplex can also yield benefits in the medium access control layer [110]. By leveraging full duplex NOMA in CF-mMIMO systems, it becomes possible to support multiple uplink and downlink users concurrently using the same frequency, resulting in improved SE. Furthermore, this approach simplifies the design of mMIMO base stations by utilizing the abundance of antennas and capitalizing on the degrees of freedom offered by full-duplex transmission [111]. The integration of FD-NOMA and CF-mMIMO systems has challenges like complexity and energy consumption of signal processing for reducing self-interference and interference between multiple users. As the number of antennas increments, one of the challenges faced is the growing hardware complexity. However, it is critical to evaluate the influence of hardware impairment on the resolution of analog-to-digital converters in such systems [111].

4.14. Grant-Free Access

Also referred to as contention-based transmission, grant-free (GF) transmission takes place over preconfigured or semi-statistically configured resources for uplink or downlink users. The sharing of resources among multiple users in GF systems can lead to transmission collision. As opposed to grant-based transmission, the GF transmission scheme is characterized by an “arrive and go” approach that is suitable for services and applications that demand low latency, like ultra-reliable low-latency communications [112]. The GF transmission scheme can also significantly save power and reduce signaling overhead for uplink transmission because it halts the need to forward a scheduling request to the BS upon traffic arrival and can avoid the detection process when receiving control information. GF transmission may incur certain collisions due to its contention-based nature, with multiple users possibly sharing time and frequency resources. This can lead to unavoidable retransmissions and concerns about reliability. The combination of NOMA and GF transmission can ensure reliable, fast, and efficient data transmission, given that the NOMA receiver can separate overlapped signals with high reliability. Additionally, integrated protocols like random back-off approaches help reduce non-orthogonal collisions and packet-dropping rates. NOMA eliminates the need for the base station to access the grant procedure, but compressed sensing algorithms can address this issue by leveraging user activity sparsity [113]. With large spatial diversity and multiplexing gains, acceptable transmission reliability and SE can be realized. As studied previously, features of CF-mMIMO adopted from mMIMO include channel hardening and favorable propagation. Features distinct to CF-mMIMO include macro-diversity and signal sparsity [114]. In combination with GF

transmission, these features would open many new research avenues. Resource allocation schemes remain a challenge when there are numerous users or very low latency is required. This includes intelligent user pairing, dynamic precoding, and channel estimation, among others. Furthermore, there is currently no existing generalized framework for resource allocation in CF-mMIMO that effectively balances scalability, reliability, and latency trade-offs [114]. Finally, the application of GF transmission to CF-mMIMO-NOMA could be used to address collision issues for both data and pilots.

4.15. New Waveforms and High Mobility Scenarios

Modulation techniques significantly contribute to enhancing data rates and minimizing energy consumption in multiple access schemes. The requirements of 6G waveforms and modulation schemes include but are not limited to very high frequency, satellite communications, short-range communications, low-cost devices and hardware, high mobility scenarios, and ultra-reliable low-latency communications [112]. Such varied requirements lead to different potential research directions, such as low PAPR, low complexity to enable power saving, time-frequency localization to improve frequency utilization and SE, high mobility where the Doppler effect leads to time selectivity in the wireless channels, and robustness to radio frequency distortion.

High mobility communication systems require high robustness to the Doppler effect and compatibility with MIMO technologies. The channels use time-and-frequency-selective or doubly selective fading [112,115], due to the Doppler effect. Using these approaches, advanced schemes capable of meeting overhead, complexity, or other requirements should be considered [116] provides several schemes that can be used to address novel trials of 5G systems. These include the following: modulation based on pulse shaping, which includes generalized frequency division multiplexing and filter bank multicarrier; modulations based on sub-band filtering, which includes filtered orthogonal frequency division multiplexing and universal filtered multicarrier. Other approaches include spectral-precoded OFDM, guard interval discrete Fourier transform spread OFDM, and OTFS modulation. Interestingly, a combination of OTFS and NOMA [117] has attracted attention for application in environments that suffer from harsh channel conditions, such as terrestrial communications [115] and heterogeneous mobility scenarios [118]. OTFS modulation uses the delay-Doppler plane, wherein user signals are arranged orthogonally. The advantage of utilizing the time-invariant channel gains in the delay-Doppler plane is that it deciphers channel estimation and signal detection in high-mobility situations. Doppler fading channels using OTFS modulation are resilient with better SE and energy performance in MIMO [119]. Specifically, [118] proposes OTFS-NOMA uplink and downlink transmission schemes, in which users with varying mobility characteristics are clustered for the deployment of NOMA. One major conclusion is that OTFS-NOMA boosts SE and minimizes latency. The authors of [120] investigate the implementation of OTFS modulation in CF-mMIMO to improve SE when considering the effects of channel estimation. In high-mobility situations, the throughput of CF-mMIMO systems is improved when using OTFS compared to OFDM, as indicated by the results.

4.16. Low-Density Parity-Check with NOMA

With the introduction of 5G, new channel coding schemes have been implemented to meet the demanding requirements. For data channels, low-density parity-check (LDPC) codes are used due to their superior performance across various coding rates, supporting high peak rates and offering low latency and reliability at low coding rates. Contrastingly, polar codes are considered one of the initial types of codes that approach the Shannon capacity for shorter blocks.

LDPC codes are found to offer better performance when applied to NOMA [121]. The application of LDPC to CF-mMIMO systems can boost reliability. In addition, NOMA and LDPC methods can be employed to address transmission errors in data packets and to enhance the efficiency of data forwarding [122]. The authors of [123] observe that the

joining of NOMA with m-MIMO whereas applying LDPC codes offers strong performance. Furthermore, the authors of [124] find that NOMA schemes are strong for 5G LDPC codes. Therefore, further research should consider a combination of NOMA with LDPC in CF-mMIMO for research purposes. This should also include variants of LDPC codes, as in [125].

Finally, other coding techniques should also be investigated. For example, polar codes that operate near the Shannon limit have been found to offer high SE with MIMO and large system capacity with NOMA [126].

4.17. RIS-Aided NOMA

RISs, also referred to as large intelligent surfaces or intelligent reflecting surfaces, are planar surfaces composed of an array of passive reflecting elements. Every such element may individually inflict the expected phase shift on the incoming signal. Gleaned from the material from which the reflecting elements are constructed, RISs can be classed as antenna-array-based or meta surface-based [127]. RISs are anticipated to lower energy usage and boost the SE of wireless networks by artificially reconfiguring the propagation environment of electromagnetic waves. RISs boast the following advantages [128]:

- Cost-efficient manufacture and deployment.
- The ability to control and customize favorable radio environments.
- The ability to give better-accuracy contact and contactless sensing.

Due to their control of the propagation environment, RISs have proved their potential in various applications. Here are some examples of applications in the field of wireless technology:

- Enhancement of SE.
- Coverage extension.
- Enhancement of energy efficiency.

The individual advantages of NOMA, RISs, and CF-mMIMO render a combination of the three a promising technology for application to future networks. Very few works have studied RIS-NOMA or RIS-CF-mMIMO [9,113,117]. In [86], IRS with CF-mMIMO-NOMA is studied. The results show that at a lower number of RIS, the performance is better compared to traditional CF-mMIMO-NOMA, however, as the number of RIS enlarges, the users' ergodic rate is reduced due to interference from the RIS. RIS can be used with NOMA to beneficially handle the wireless channel vectors of every user, aligning them with each other to enhance performance, particularly when the channel vectors are orthogonal. RIS can also be used with NOMA to improve reliability at a very low SNR, accommodate more users, and enable higher modulation [112]. The authors of [129] investigate the outage performance of wireless systems incorporating RIS with NOMA. Specifically, positive results are generated when RIS is deployed to guarantee fairness among NOMA users in wireless systems.

CF-mMIMO systems that require extensive use of BSs can suffer from unsatisfactory energy efficiency as a result of the substantial expenses associated with hardware and power sources. RIS can be deployed to solve this issue. RIS requires no additional hardware implementation, and thus greatly reduces energy consumption and complexity for signal processing [128,130]. S. Elhoushy et al. propose the use of RIS to provide physical layer security [9]. The deployment of RIS is recommended for CF-mMIMO systems to achieve high secrecy rates, because of the flexibility with which RISs can enhance or suppress signal beams to different users. S. Elhoushy et al. further suggest the possibility of adjusting RIS-based beamforming coefficients for bettering the received signal at the expected user whilst canceling the received signal at the intruder, thereby increasing the secrecy of the system. Analysis of secrecy rates is an interesting research area. In addition, the authors of [85] suggest the application of RIS to replace relay nodes as a transmission medium and assist users with the worst channel quality in completing their message transmission, where

diversity gain may be exploited to optimize user QoS. Nonetheless, there is an insufficient study of RIS-aided CF-mMIMO systems and their applications with NOMA.

For the best exploitation of these technologies, challenges such as channel estimation, resource allocation, and system optimization must be addressed [13,131]. For example, the authors of [13] suggest that resource allocation is a challenging issue due to the size and complexity of the involved systems.

4.18. Machine Learning (ML)/Artificial Intelligence (AI)

AI plays a vital role in wireless networks, bringing intelligence and automation by replicating human cognitive processes and intelligent behaviors. It has emerged as a highly valued technology in future mobile communication systems. The use of AI in future networks will be required to increase robustness, performance, and efficiency. Thus, AI technologies are anticipated to have a vital role in networks; particularly those that are dynamic, and too complex scenarios that are challenging for human analysis. Considering their advantages, a combination of AI and CF-mMIMO-NOMA could yield very strong performances. ML methods are the most applied forms of AI. As shown in Figure 13, common ML techniques encompass deep learning, supervised learning, unsupervised learning, and deep reinforcement learning.

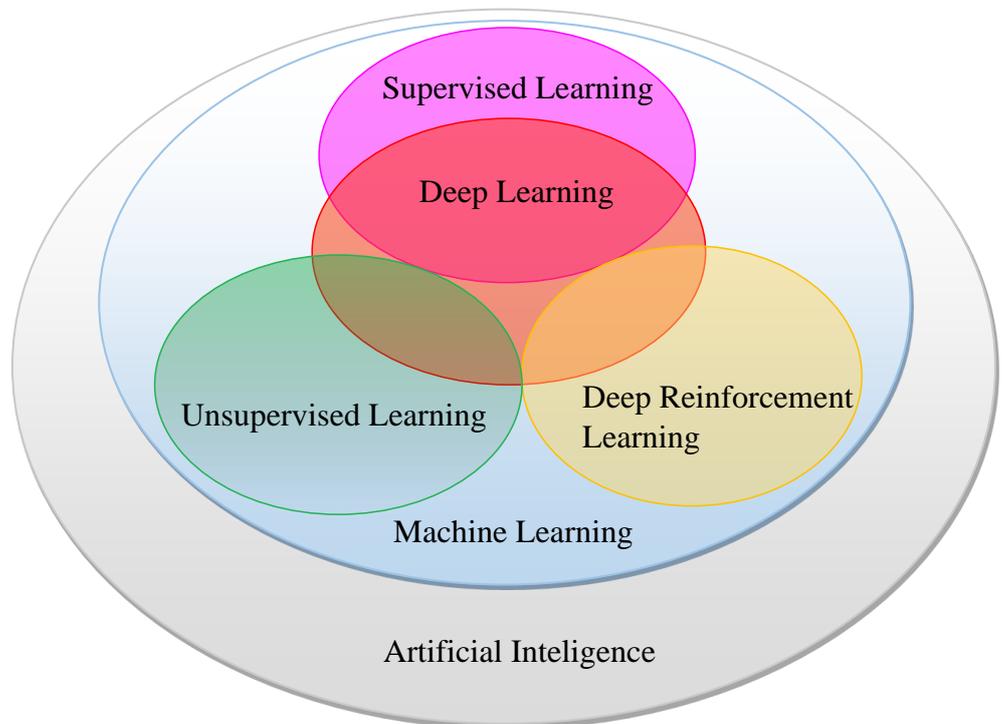


Figure 13. The relationship between AI, ML, and deep learning [132].

ML techniques have been successfully deployed in a variety of areas of wireless communication systems, including the physical layer and medium access control. These approaches have been applied for tasks such as CSI feedback, channel estimation and prediction, MIMO detection, channel coding and decoding, UC, and power control [11,117,132–134]. The utilization of ML to boost the performance of CF-mMIMO-NOMA is therefore highly attractive. When used in CSI feedback, for example, if supplied with abundant sensing data, AI can extract more useful features from angle and distance information to assist CSI recovery or prediction. The authors of [72] use unsupervised ML for UC. The authors of [10] suggest the idea of federated learning, which enables collaborative learning among distributed devices while preserving their local training and control privacy. They suggest that applying federated learning can enhance channel estimation

and resource allocation in CF-mMIMO systems. The distributed processing approach of federated learning is a natural fit for CF-mMIMO.

Both the transmitters and receivers of multiple access schemes can utilize AI techniques [112]. AI assistance could allow the design of a multiple access transmission scheme with low cost, low PAPR, low latency, high reliability, and massive connectivity properties. For example, data-driven neural networks can be used to design NOMA signatures together with many other modules, such as waveforms and MIMO precoding. To improve detection and reduce complexity, an AI-assisted receiver could facilitate multi-user detection for NOMA. Specifically, the authors of [117] give a detailed overview of the ML tools utilized in NOMA communications.

To achieve truly intelligent systems in 6G, AI-enabled systems must be more reliable, efficient, and easier to implement. Therefore, AI should be implemented to facilitate training and improve learning efficiency, while data- and model-driven AI should be implemented to ensure more timely and accurate learning [112].

4.19. CF-mMIMO-NOMA with MEC

Edge computing represents a progression from cloud computing, enabling the hosting of applications to be moved closer to end-users and the data produced by those applications, residing at the network edge. It is regarded as one of the critical pillars for attaining the stringent core indicators of 5G, particularly low latency and bandwidth efficiency. ETSI ISG MEC, the industry specification group for MEC, provides the technical standards for edge computing [135]. MEC is defined by its on-premises location, proximity to end-users, low latency, awareness of location, and contextual information of the network. The advantages of MEC extend to various services and applications that go beyond the capabilities of 5G networks. The integration of MEC with these techniques will enhance the value of MEC systems. Of particular interest is the integration of NOMA with MEC. This combination brings many benefits, including support for numerous users, reduced latency and energy usage of users, and improved achievements for complex network scenarios such as millimeter wave mMIMO [136]. The authors of [136] indicate that a combination of MEC and NOMA could provide several benefits, including:

- A significant improvement in user satisfaction and network performance by leveraging golden opportunities.
- Reinforcement of the services and applications that are supported by 5G networks.
- Provision of low latency transmission.
- Flexible combination with many existing technologies, like MIMO, mMIMO, and millimeter wave communications, leading to increased connectivity, SE, energy efficiency, and computing capability.

The integration of MEC with NOMA and CF-mMIMO has the potential to improve computing capabilities and SE and reduce task delay. Specifically, CF communications can boost reliability, with MEC serving as a suitable platform for managing delay-sensitive applications. For example, the authors of [137] introduce a novel CF-mMIMO system with edge computing, together with a cloud data center and several APs, and derives the probability of successful edge computing. The authors observe that for a given successful edge computing probability, the energy usage is reduced with larger AP density, rather than with more antennas per AP. With a focus on maximizing the sum data rate, [85] studies MEC in CF edge networks where NOMA is used to cluster users. A clustered NOMA-based adaptive relay coordinated transmission scheme is found to outperform direct transmission, relay AF transmission, and relay DF transmission schemes.

5. Conclusions

This paper provided a thorough review of CF-mMIMO-NOMA, with a focus on identifying research gaps and the possibility of integration with other enabling technologies. A detailed theoretical description was provided, in addition to a simple system model and description of the operation.

Notably, TDD operation is more appropriate than FDD because TDD allows the exploitation of channel reciprocity for hardware calibration. A comprehensive literature review found that no work has investigated the use of FDD with CF-mMIMO-NOMA. The problems of error propagation and imperfect SIC still pose a challenge for CF-mMIMO-NOMA. As a result, the system performs more poorly for a low number of users. Although some authors [73,76] have investigated switching techniques to implement OMA in regions with low users, the switching algorithm is dependent upon the length of the coherence time. In addition, few works consider multi-antenna users, despite current and future UEs requiring many antennas. This limits practicability and is an area in which more research should be carried out.

The integration of CF-mMIMO-NOMA with certain 6G-enabling technologies was discussed. AI has attracted substantial interest in recent works. Its application in the physical layer could allow the full potential of CF-mMIMO-NOMA to be realized, and improve resource allocation, which remains a persistent challenge. The integration of MEC with CF-mMIMO and NOMA could lead to further increases in connectivity, SE, energy efficiency, and computing capability. In addition, RISs could be employed in CF-mMIMO systems can address the issue of unsatisfactory energy efficiency caused by the extensive deployment of base stations, which incurs significant costs for hardware and power sources. While each of these integrations seems inevitable, there remain many challenges that must be addressed to optimize performance.

In conclusion, this review has provided a thorough examination of CF-mMIMO-NOMA, covering its motivation, current state-of-the-art, system model, research challenges, and the potential integration with other technologies that enable 6G.

Author Contributions: Conceptualization—A.A. and J.I.; Methodology—A.A.; Software—A.A.; Validation—A.A. and J.I.; Investigation—A.A.; Resources—A.A. and J.I.; Writing-original draft—A.A.; Writing-review and editing—A.A. and J.I.; Visualization—A.A. and J.I.; Supervision—J.I.; Project administration—J.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not Applicable.

Conflicts of Interest: Author Antonio Apiyo was employed by the company Nokia. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. You, X.; Wang, C.X.; Huang, J.; Gao, X.; Zhang, Z.; Wang, M.; Huang, Y.; Zhang, C.; Jiang, Y.; Wang, J.; et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Sci. China Inf. Sci.* **2021**, *64*, 110301. [[CrossRef](#)]
2. Saghezchi, F.B.; Rodriguez, J.; Vujicic, Z.; Nascimento, A.; Huq, K.M.S.; Gil-Castiñeira, F. Drive towards 6G. In *Enabling 6G Mobile Networks*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 3–35. [[CrossRef](#)]
3. Alsabah, M.; Naser, M.A.; Mahmmod, B.M.; Abdhussain, S.H.; Eissa, M.R.; Al-Baidhani, A.; Noordin, N.K.; Sait, S.M.; Al-Utaibi, K.A.; Hashim, F. 6G Wireless Communications Networks: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 148191–148243. [[CrossRef](#)]
4. Shin, W.; Vaezi, M.; Lee, B.; Love, D.J.; Lee, J.; Poor, H.V. Non-Orthogonal Multiple Access in Multi-Cell Networks: Theory, Performance, and Practical Challenges. *IEEE Commun. Mag.* **2017**, *55*, 176–183. [[CrossRef](#)]
5. Saito, Y.; Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Li, A.; Higuchi, K. Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, Germany, 2–5 June 2013; pp. 1–5. [[CrossRef](#)]
6. Dai, L.; Wang, B.; Ding, Z.; Wang, Z.; Chen, S.; Hanzo, L. A Survey of Non-Orthogonal Multiple Access for 5G. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2294–2323. [[CrossRef](#)]
7. Cai, Y.; Qin, Z.; Cui, F.; Li, G.Y.; McCann, J.A. Modulation and Multiple Access for 5G Networks. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 629–646. [[CrossRef](#)]
8. Zhang, J.; Chen, S.; Lin, Y.; Zheng, J.; Ai, B.; Hanzo, L. Cell-Free Massive MIMO: A New Next-Generation Paradigm. *IEEE Access* **2019**, *7*, 99878–99888. [[CrossRef](#)]
9. Elhoushy, S.; Ibrahim, M.; Hamouda, W. Cell-Free Massive MIMO: A Survey. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 492–523. [[CrossRef](#)]

10. Chen, S.; Zhang, J.; Zhang, J.; Björnson, E.; Ai, B. A survey on user-centric cell-free massive MIMO systems. *Digit. Commun. Netw.* **2022**, *8*, 695–719. [[CrossRef](#)]
11. He, H.; Yu, X.; Zhang, J.; Song, S.; Letaief, K.B. Cell-free massive MIMO for 6G wireless communication networks. *J. Commun. Inf. Netw.* **2021**, *6*, 321–335. [[CrossRef](#)]
12. Obakhena, H.I.; Imoize, A.L.; Anyasi, F.I.; Kavitha, K. Application of cell-free massive MIMO in 5G and beyond 5G wireless networks: A survey. *J. Eng. Appl. Sci.* **2021**, *68*, 13. [[CrossRef](#)]
13. Shi, E.; Zhang, J.; Chen, S.; Zheng, J.; Zhang, Y.; Ng, D.W.K.; Ai, B. Wireless energy transfer in RIS-aided cell-free massive MIMO systems: Opportunities and challenges. *IEEE Commun. Mag.* **2022**, *60*, 26–32. [[CrossRef](#)]
14. Ammar, H.A.; Adve, R.; Shahbazpanahi, S.; Boudreau, G.; Srinivas, K.V. User-Centric Cell-Free Massive MIMO Networks: A Survey of Opportunities, Challenges and Solutions. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 611–652. [[CrossRef](#)]
15. Zheng, J.; Zhang, J.; Du, H.; Niyato, D.; Ai, B.; Debbah, M.; Letaief, K.B. Mobile Cell-Free Massive MIMO: Challenges, Solutions, and Future Directions. *arXiv* **2023**, arXiv:2302.02566.
16. Edfors, O.; Liu, L.; Tufvesson, F.; Kundargi, N.; Nieman, K. Massive MIMO for 5G: Theory, implementation and prototyping. In *Signal Processing for 5G: Algorithms and Implementations*; Wiley: Hoboken, NJ, USA, 2016; pp. 189–230.
17. Zhao, L.; Zhao, H.; Zheng, K.; Xiang, W. *Massive MIMO in 5G Networks: Selected Applications*; Springer: Berlin/Heidelberg, Germany, 2018. [[CrossRef](#)]
18. Björnson, E.; Hoydis, J.; Sanguinetti, L. Massive MIMO networks: Spectral, energy, and hardware efficiency. *Found. Trends[®] Signal Process.* **2017**, *11*, 154–655. [[CrossRef](#)]
19. Marzetta, T.L.; Yang, H. *Fundamentals of Massive MIMO*; Cambridge University Press: Cambridge, UK, 2016.
20. Björnson, E.; Sanguinetti, L.; Wymeersch, H.; Hoydis, J.; Marzetta, T.L. Massive MIMO is a reality—What is next? Five promising research directions for antenna arrays. *Digit. Signal Process.* **2019**, *94*, 3–20. [[CrossRef](#)]
21. Borges, D.; Montezuma, P.; Dinis, R.; Beko, M. Massive mimo techniques for 5g and beyond—Opportunities and challenges. *Electronics* **2021**, *10*, 1667. [[CrossRef](#)]
22. Chataut, R.; Akl, R. Massive MIMO systems for 5G and beyond networks—Overview, recent trends, challenges, and future research direction. *Sensors* **2020**, *20*, 2753. [[CrossRef](#)] [[PubMed](#)]
23. Larsson, E.G.; Edfors, O.; Tufvesson, F.; Marzetta, T.L. Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* **2014**, *52*, 186–195. [[CrossRef](#)]
24. Araújo, D.C.; Maksymyuk, T.; de Almeida, A.L.; Maciel, T.; Mota, J.C.; Jo, M. Massive MIMO: Survey and future research topics. *IET Commun.* **2016**, *10*, 1938–1946. [[CrossRef](#)]
25. Zheng, K.; Ou, S.; Yin, X. Massive MIMO channel models: A survey. *Int. J. Antennas Propag.* **2014**, *2014*, 848071. [[CrossRef](#)]
26. Albreem, M.A.; Habbash, A.H.A.; Abu-Hudrouss, A.M.; Ikki, S.S. Overview of Precoding Techniques for Massive MIMO. *IEEE Access* **2021**, *9*, 60764–60801. [[CrossRef](#)]
27. Fatema, N.; Hua, G.; Xiang, Y.; Peng, D.; Natgunanathan, I. Massive MIMO Linear Precoding: A Survey. *IEEE Syst. J.* **2018**, *12*, 3920–3931. [[CrossRef](#)]
28. Lu, L.; Li, G.Y.; Swindlehurst, A.L.; Ashikhmin, A.; Zhang, R. An Overview of Massive MIMO: Benefits and Challenges. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 742–758. [[CrossRef](#)]
29. Kebede, T.; Wondie, Y.; Steinbrunn, J.; Kassa, H.B.; Kornegay, K.T. Precoding and Beamforming Techniques in mmWave-Massive MIMO: Performance Assessment. *IEEE Access* **2022**, *10*, 16365–16387. [[CrossRef](#)]
30. Zhang, X.; Vaezi, M. Deep Learning Based Precoding for the MIMO Gaussian Wiretap Channel. In Proceedings of the 2019 IEEE Globecom Workshops (GC Wkshps), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6. [[CrossRef](#)]
31. Shi, J.; Wang, W.; Yi, X.; Gao, X.; Li, G.Y. Deep Learning-Based Robust Precoding for Massive MIMO. *IEEE Trans. Commun.* **2021**, *69*, 7429–7443. [[CrossRef](#)]
32. Pereira de Figueiredo, F.A. An Overview of Massive MIMO for 5G and 6G. *IEEE Lat. Am. Trans.* **2022**, *20*, 931–940. [[CrossRef](#)]
33. Marzetta, T.L. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [[CrossRef](#)]
34. Akyildiz, I.F.; Kak, A.; Nie, S. 6G and Beyond: The Future of Wireless Communications Systems. *IEEE Access* **2020**, *8*, 133995–134030. [[CrossRef](#)]
35. Tataria, H.; Shafi, M.; Molisch, A.F.; Dohler, M.; Sjöland, H.; Tufvesson, F. 6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities. *Proc. IEEE* **2021**, *109*, 1166–1199. [[CrossRef](#)]
36. Viswanathan, H.; Mogensen, P.E. Communications in the 6G Era. *IEEE Access* **2020**, *8*, 57063–57074. [[CrossRef](#)]
37. Ziegler, V.; Viswanathan, H.; Flinck, H.; Hoffmann, M.; Räisänen, V.; Hätönen, K. 6G Architecture to Connect the Worlds. *IEEE Access* **2020**, *8*, 173508–173520. [[CrossRef](#)]
38. Nayebi, E.; Ashikhmin, A.; Marzetta, T.L.; Yang, H. Cell-Free Massive MIMO systems. In Proceedings of the 2015 49th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 8–11 November 2015; pp. 695–699. [[CrossRef](#)]
39. Ngo, H.Q.; Ashikhmin, A.; Yang, H.; Larsson, E.G.; Marzetta, T.L. Cell-Free Massive MIMO: Uniformly great service for everyone. In Proceedings of the 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Stockholm, Sweden, 28 June–1 July 2015; pp. 201–205. [[CrossRef](#)]
40. Ngo, H.Q.; Ashikhmin, A.; Yang, H.; Larsson, E.G.; Marzetta, T.L. Cell-Free Massive MIMO Versus Small Cells. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 1834–1850. [[CrossRef](#)]

41. Shaik, Z.H.; Björnson, E.; Larsson, E.G. Cell-Free Massive MIMO with Radio Stripes and Sequential Uplink Processing. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [\[CrossRef\]](#)
42. Yang, H.; Marzetta, T.L. Energy Efficiency of Massive MIMO: Cell-Free vs. Cellular. In Proceedings of the 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018; pp. 1–5. [\[CrossRef\]](#)
43. Interdonato, G.; Björnson, E.; Quoc Ngo, H.; Frenger, P.; Larsson, E.G. Ubiquitous cell-free massive MIMO communications. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 1–13. [\[CrossRef\]](#)
44. Ngo, H.Q.; Tran, L.N.; Duong, T.Q.; Matthaiou, M.; Larsson, E.G. On the Total Energy Efficiency of Cell-Free Massive MIMO. *IEEE Trans. Green Commun. Netw.* **2018**, *2*, 25–39. [\[CrossRef\]](#)
45. Demir, Ö.T.; Björnson, E.; Sanguinetti, L. Foundations of user-centric cell-free massive MIMO. *Found. Trends[®] Signal Process.* **2021**, *14*, 162–472. [\[CrossRef\]](#)
46. Wang, P.; Xiao, J.; Ping, L. Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems. *IEEE Veh. Technol. Mag.* **2006**, *1*, 4–11. [\[CrossRef\]](#)
47. Kumaresan, S.P.; Tan, C.K.; Ng, Y.H. Extreme Learning Machine (ELM) for Fast User Clustering in Downlink Non-Orthogonal Multiple Access (NOMA) 5G Networks. *IEEE Access* **2021**, *9*, 130884–130894. [\[CrossRef\]](#)
48. Kizilirmak, R.C.; Bizaki, H.K. Non-orthogonal multiple access (NOMA) for 5G networks. *Towards Wirel. Netw. Phys. Layer Perspect.* **2016**, *83*, 83–98.
49. Liu, Y.; Qin, Z.; Elkashlan, M.; Ding, Z.; Nallanathan, A.; Hanzo, L. Nonorthogonal Multiple Access for 5G and Beyond. *Proc. IEEE* **2017**, *105*, 2347–2381. [\[CrossRef\]](#)
50. Wang, L.; Şaşıoğlu, E.; Bandemer, B.; Kim, Y.H. A comparison of superposition coding schemes. In Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 2970–2974. [\[CrossRef\]](#)
51. Sen, S.; Santhapuri, N.; Choudhury, R.R.; Nelakuditi, S. Successive interference cancellation: A back-of-the-envelope perspective. In Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks, Monterey, CA, USA, 20–21 October 2010; pp. 1–6.
52. Xu, P.; Ding, Z.; Dai, X.; Poor, H.V. NOMA: An information theoretic perspective. *arXiv* **2015**, arXiv:1504.07751.
53. Tse, D.; Viswanath, P. *Fundamentals of Wireless Communication*; Cambridge University Press: Cambridge, UK, 2005.
54. Mahmoudi, A.; Abolhassani, B.; Razavizadeh, S.M.; Nguyen, H.H. User Clustering and Resource Allocation in Hybrid NOMA-OMA Systems Under Nakagami-m Fading. *IEEE Access* **2022**, *10*, 38709–38728. [\[CrossRef\]](#)
55. Katwe, M.; Singh, K.; Sharma, P.K.; Li, C.P.; Ding, Z. Dynamic User Clustering and Optimal Power Allocation in UAV-Assisted Full-Duplex Hybrid NOMA System. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 2573–2590. [\[CrossRef\]](#)
56. Ding, Z.; Fan, P.; Poor, H.V. Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions. *IEEE Trans. Veh. Technol.* **2016**, *65*, 6010–6023. [\[CrossRef\]](#)
57. Ali, M.S.; Tabassum, H.; Hossain, E. Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems. *IEEE Access* **2016**, *4*, 6325–6343. [\[CrossRef\]](#)
58. Shahjalal, M.; Rahman, M.H.; Ali, M.O.; Chung, B.; Jang, Y.M. User Clustering Techniques for Massive MIMO-NOMA Enabled mmWave/THz Communications in 6G. In Proceedings of the 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Jeju Island, Republic of Korea, 17–20 August 2021; pp. 379–383. [\[CrossRef\]](#)
59. Zhang, H.; Zhang, D.K.; Meng, W.X.; Li, C. User pairing algorithm with SIC in non-orthogonal multiple access system. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6. [\[CrossRef\]](#)
60. Li, Y.; Aruma Baduge, G.A. NOMA-Aided Cell-Free Massive MIMO Systems. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 950–953. [\[CrossRef\]](#)
61. Sayyari, R.; Pourrostam, J.; Niya, M.J.M. Cell-Free Massive MIMO System with an Adaptive Switching Algorithm between Cooperative NOMA, Non-Cooperative NOMA, and OMA Modes. *IEEE Access* **2021**, *9*, 149227–149239. [\[CrossRef\]](#)
62. Ohashi, A.A.; Costa, D.B.d.; Fernandes, A.L.P.; Monteiro, W.; Failache, R.; Cavalcante, A.M.; Costa, J.C.W.A. Cell-Free Massive MIMO-NOMA Systems with Imperfect SIC and Non-Reciprocal Channels. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1329–1333. [\[CrossRef\]](#)
63. Zhang, Y.; Cao, H.; Zhou, M.; Yang, L. Spectral Efficiency Maximization for Uplink Cell-Free Massive MIMO-NOMA Networks. In Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 20–24 May 2019; pp. 1–6. [\[CrossRef\]](#)
64. Nguyen, T.K.; Nguyen, H.H.; Tuan, H.D. Max-Min QoS Power Control in Generalized Cell-Free Massive MIMO-NOMA With Optimal Backhaul Combining. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10949–10964. [\[CrossRef\]](#)
65. Nguyen, T.K.; Nguyen, H.H.; Tuan, H.D. Cell-Free Massive MIMO-NOMA with Optimal Backhaul Combining. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 13–15 January 2021; pp. 455–460. [\[CrossRef\]](#)
66. Zhang, Y.; Cao, H.; Zhou, M.; Yang, L. Non-orthogonal multiple access in cell-free massive MIMO networks. *China Commun.* **2020**, *17*, 81–94. [\[CrossRef\]](#)
67. Rezaei, F.; Heidarpour, A.R.; Tellambura, C.; Tadaion, A. Underlaid Spectrum Sharing for Cell-Free Massive MIMO-NOMA. *IEEE Commun. Lett.* **2020**, *24*, 907–911. [\[CrossRef\]](#)

68. Rezaei, F.; Tellambura, C.; Tadaion, A.A.; Heidarpour, A.R. Rate Analysis of Cell-Free Massive MIMO-NOMA with Three Linear Precoders. *IEEE Trans. Commun.* **2020**, *68*, 3480–3494. [[CrossRef](#)]
69. Zhang, J.; Fan, J.; Ai, B.; Ng, D.W.K. NOMA-Based Cell-Free Massive MIMO Over Spatially Correlated Rician Fading Channels. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [[CrossRef](#)]
70. Zhang, X.; Wang, J.; Poor, H.V. Statistical QoS Provisioning Over Cell-Free M-MIMO-NOMA Based 5G+ Mobile Wireless Networks in the Non-Asymptotic Regime. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; pp. 1–5. [[CrossRef](#)]
71. Zhang, J.; Fan, J.; Zhang, J.; Ng, D.W.K.; Sun, Q.; Ai, B. Performance Analysis and Optimization of NOMA-Based Cell-Free Massive MIMO for IoT. *IEEE Internet Things J.* **2022**, *9*, 9625–9639. [[CrossRef](#)]
72. Le, Q.N.; Nguyen, V.D.; Dobre, O.A.; Nguyen, N.P.; Zhao, R.; Chatzinotas, S. Learning-Assisted User Clustering in Cell-Free Massive MIMO-NOMA Networks. *IEEE Trans. Veh. Technol.* **2021**, *70*, 12872–12887. [[CrossRef](#)]
73. Bashar, M.; Cumanan, K.; Burr, A.G.; Ngo, H.Q.; Hanzo, L.; Xiao, P. On the Performance of Cell-Free Massive MIMO Relying on Adaptive NOMA/OMA Mode-Switching. *IEEE Trans. Commun.* **2020**, *68*, 792–810. [[CrossRef](#)]
74. Kusaladharma, S.; Zhu, W.P.; Ajib, W.; Amarasuriya, G. Achievable Rate Analysis of NOMA in Cell-Free Massive MIMO: A Stochastic Geometry Approach. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–6. [[CrossRef](#)]
75. Kusaladharma, S.; Zhu, W.P.; Ajib, W.; Baduge, G.A.A. Achievable Rate Characterization of NOMA-Aided Cell-Free Massive MIMO with Imperfect Successive Interference Cancellation. *IEEE Trans. Commun.* **2021**, *69*, 3054–3066. [[CrossRef](#)]
76. Bashar, M.; Cumanan, K.; Burr, A.G.; Ngo, H.Q.; Hanzo, L.; Xiao, P. NOMA/OMA Mode Selection-Based Cell-Free Massive MIMO. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–6. [[CrossRef](#)]
77. Zargari, S.; Ahmadinejad, H.; Abolhassani, B.; Falahati, A. SWIPT-NOMA in Cell-Free Massive MIMO. In Proceedings of the 2020 28th Iranian Conference on Electrical Engineering (ICEE), Tabriz, Iran, 4–6 August 2020; pp. 1–6. [[CrossRef](#)]
78. Wang, Z.; Zhang, D.; Xu, K.; Xie, W.; Xv, J.; Li, X. NOMA in Cell-Free mMIMO Systems with AP Selection. In Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 21–23 October 2020; pp. 430–435. [[CrossRef](#)]
79. Dang, X.T.; Le, M.T.P.; Nguyen, H.V.; Chatzinotas, S.; Shin, O.S. Optimal User Pairing Approach for NOMA-Based Cell-Free Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2023**, *72*, 4751–4765. [[CrossRef](#)]
80. Dang, X.T.; Le, M.T.P.; Nguyen, H.V.; Shin, O.S. Optimal User Pairing for NOMA-assisted Cell-free Massive MIMO System. In Proceedings of the 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), Nha Trang, Vietnam, 27–29 July 2022; pp. 7–12. [[CrossRef](#)]
81. Nguyen, T.K.; Nguyen, H.H.; Tuan, H.D. Adaptive Successive Interference Cancellation in Cell-free Massive MIMO-NOMA. In Proceedings of the 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Victoria, BC, Canada, 18 November–16 December 2020; pp. 1–5. [[CrossRef](#)]
82. Galappaththige, D.L.; Amarasuriya, G. NOMA-Aided Cell-Free Massive MIMO with Underlay Spectrum-Sharing. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [[CrossRef](#)]
83. Hua, M.; Ni, W.; Tian, H.; Nie, G. Energy-Efficient Uplink Power Control in NOMA Enhanced Cell-Free Massive MIMO Networks. In Proceedings of the 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Xiamen, China, 28–30 July 2021; pp. 7–12. [[CrossRef](#)]
84. Zhang, X.; Zhu, Q. NOMA and User-Centric Based Cell-Free Massive MIMO over 6G Big-Data Mobile Wireless Networks. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [[CrossRef](#)]
85. Peng, X.; Yuan, X.; Zhang, K. Clustered NOMA-based downlink adaptive relay coordinated transmission scheme for future 6G cell-free edge network. *Peer-to-Peer Netw. Appl.* **2022**, *15*, 612–625. [[CrossRef](#)]
86. Rafieifar, A.; Ahmadinejad, H.; Falahati, A. IRS-aided NOMA in a Cell Free Massive MIMO System. In Proceedings of the 2022 30th International Conference on Electrical Engineering (ICEE), Tehran, Iran, 17–19 May 2022; pp. 874–879. [[CrossRef](#)]
87. Mendoza, C.F.; Schwarz, S.; Rupp, M. Cluster Formation in Scalable Cell-free Massive MIMO Networks. In Proceedings of the 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Thessaloniki, Greece, 12–14 October 2020; pp. 62–67. [[CrossRef](#)]
88. Failache, R.; Ohashi, A.; Fernandes, A.; Rodrigues, R.; Cavalcante, A.; Weyl, J. A Fair Comparison Between OMA and NOMA For Cell-Free Massive MIMO Systems. *arXiv* **2020**, arXiv:2002.03819.
89. Tan, F.; Wu, P.; Wu, Y.C.; Xia, M. Energy-Efficient Non-Orthogonal Multicast and Unicast Transmission of Cell-Free Massive MIMO Systems with SWIPT. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 949–968. [[CrossRef](#)]
90. Wang, W.; Wang, C.; Pang, M.; Wang, W.; Jiang, F. Fast Signal Reconstruction Based on Compressed Sensing in NOMA-Aided Cell-Free Massive MIMO. In Proceedings of the GLOBECOM 2022-2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 1–6.

91. Abdallah, A.; Mansour, M.M. Efficient Angle-Domain Processing for FDD-Based Cell-Free Massive MIMO Systems. *IEEE Trans. Commun.* **2020**, *68*, 2188–2203. [\[CrossRef\]](#)
92. Abdallah, A.; Mansour, M.M. Angle-Based Multipath Estimation and Beamforming for FDD Cell-free Massive MIMO. In Proceedings of the 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, 2–5 July 2019; pp. 1–5. [\[CrossRef\]](#)
93. Van Trees, H.L. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*; John Wiley & Sons: New York, NY, USA, 2002.
94. Huang, Y.; Zhang, C.; Wang, J.; Jing, Y.; Yang, L.; You, X. Signal Processing for MIMO-NOMA: Present and Future Challenges. *IEEE Wirel. Commun.* **2018**, *25*, 32–38. [\[CrossRef\]](#)
95. Dong, P.; Zhang, H.; Li, G.Y.; Gaspar, I.S.; NaderiAlizadeh, N. Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 989–1000. [\[CrossRef\]](#)
96. He, H.; Wen, C.K.; Jin, S.; Li, G.Y. Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 852–855. [\[CrossRef\]](#)
97. Boccardi, F.; Tosato, F.; Caire, G. Precoding schemes for the MIMO-GBC. In Proceedings of the 2006 International Zurich Seminar on Communications, IEEE, Zurich, Switzerland, 22–24 February 2004; pp. 10–13.
98. Heath, R.W.; Gonzalez-Prelcic, N.; Rangan, S.; Roh, W.; Sayeed, A.M. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 436–453. [\[CrossRef\]](#)
99. Björnson, E.; Kountouris, M.; Bengtsson, M.; Ottersten, B. Receive Combining vs. Multi-Stream Multiplexing in Downlink Systems with Multi-Antenna Users. *IEEE Trans. Signal Process.* **2013**, *61*, 3431–3446. [\[CrossRef\]](#)
100. Budhiraja, I.; Kumar, N.; Tyagi, S. Cross-Layer Interference Management Scheme for D2D Mobile Users Using NOMA. *IEEE Syst. J.* **2021**, *15*, 3109–3120. [\[CrossRef\]](#)
101. Islam, S.M.R.; Avazov, N.; Dobre, O.A.; Kwak, K.S. Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 721–742. [\[CrossRef\]](#)
102. Meng, R.; Ye, Y.; Xie, N.G. Multi-objective optimization design methods based on game theory. In Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, IEEE, Jinan, China, 7–9 July 2010; pp. 2220–2227.
103. Awoyemi, B.S.; J. Maharaj, B.T.; Alfa, A.S. Solving resource allocation problems in cognitive radio networks: A survey. *EURASIP J. Wirel. Commun. Netw.* **2016**, *2016*, 176. [\[CrossRef\]](#)
104. Blum, C.; Puchinger, J.; Raidl, G.R.; Roli, A. Hybrid metaheuristics in combinatorial optimization: A survey. *Appl. Soft Comput.* **2011**, *11*, 4135–4151. [\[CrossRef\]](#)
105. Polegre, A.A.; Riera-Palou, F.; Femenias, G.; Armada, A.G. Channel Hardening in Cell-Free and User-Centric Massive MIMO Networks with Spatially Correlated Ricean Fading. *IEEE Access* **2020**, *8*, 139827–139845. [\[CrossRef\]](#)
106. Kim, S.; Shim, B. FDD-Based Cell-Free Massive MIMO Systems. In Proceedings of the 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 25–28 June 2018; pp. 1–5. [\[CrossRef\]](#)
107. Kim, S.; Choi, J.W.; Shim, B. Downlink Pilot Precoding and Compressed Channel Feedback for FDD-Based Cell-Free Systems. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 3658–3672. [\[CrossRef\]](#)
108. Han, T.; Zhao, D. Downlink channel estimation in FDD cell-free massive MIMO. *Phys. Commun.* **2022**, *52*, 101614. [\[CrossRef\]](#)
109. Han, T.; Zhao, D. On the Performance of FDD Cell-Free Massive MIMO with Compressed Sensing Channel Estimation. In Proceedings of the 2021 IEEE 21st International Conference on Communication Technology (ICCT), Tianjin, China, 13–16 October 2021; pp. 238–242. [\[CrossRef\]](#)
110. da Silva, B., Jr.; Mairton, J.; Skouroumounis, C.; Krikidis, I.; Fodor, G.; Fischione, C. *Energy Efficient Full-Duplex Networks*; The Institution of Engineering and Technology (IET): Stevenage, UK, 2020.
111. Gazestani, A.H.; Ghorashi, S.A.; Mousavinasab, B.; Shikh-Bahaei, M. A survey on implementation and applications of full duplex wireless communications. *Phys. Commun.* **2019**, *34*, 121–134. [\[CrossRef\]](#)
112. Rong, B. 6G: The next horizon: From connected people and things to connected intelligence. *IEEE Wirel. Commun.* **2021**, *28*, 8. [\[CrossRef\]](#)
113. Budhiraja, I.; Kumar, N.; Tyagi, S.; Tanwar, S.; Han, Z.; Piran, M.J.; Suh, D.Y. A Systematic Review on NOMA Variants for 5G and Beyond. *IEEE Access* **2021**, *9*, 85573–85644. [\[CrossRef\]](#)
114. Ding, J.; Nemati, M.; Pokhrel, S.R.; Park, O.S.; Choi, J.; Adachi, F. Enabling Grant-Free URLLC: An Overview of Principle and Enhancements by Massive MIMO. *IEEE Internet Things J.* **2022**, *9*, 384–400. [\[CrossRef\]](#)
115. Uwaechia, A.N.; Mahyuddin, N.M. A Comprehensive Survey on Millimeter Wave Communications for Fifth-Generation Wireless Networks: Feasibility and Challenges. *IEEE Access* **2020**, *8*, 62367–62414. [\[CrossRef\]](#)
116. Kripa, M.; Mathews, L. A Survey on Modulation and Multiple Access Techniques For Exploiting Heterogeneous User Mobility Profiles. *IOSR J. Electron. Commun. Eng.* **2020**, *15*, 47–50.
117. Liu, Y.; Zhang, S.; Mu, X.; Ding, Z.; Schober, R.; Al-Dhahir, N.; Hossain, E.; Shen, X. Evolution of NOMA Toward Next Generation Multiple Access (NGMA) for 6G. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 1037–1071. [\[CrossRef\]](#)
118. Zhou, X.; Ying, K.; Gao, Z.; Wu, Y.; Xiao, Z.; Chatzinotas, S.; Yuan, J.; Ottersten, B. Active Terminal Identification, Channel Estimation, and Signal Detection for Grant-Free NOMA-OTFS in LEO Satellite Internet-of-Things. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 2847–2866. [\[CrossRef\]](#)

119. Merin Joshiba, J.; Judson, D.; Albert Raj, A. 5G Modulation techniques—A systematic literature survey. In Proceedings of the International Conference on Futuristic Communication and Network Technologies, Chennai, India, 6–7 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 351–372.
120. Mohammadi, M.; Ngo, H.Q.; Matthaiou, M. Cell-Free Massive MIMO Meets OTFS Modulation. *IEEE Trans. Commun.* **2022**, *70*, 7728–7747. [[CrossRef](#)]
121. Idris, M.S.; Ali, D.M.; Razak, N.I.A.; Idris, A.; Ahmad, H. Performance analysis of NOMA using different coding techniques. *J. Phys. Conf. Ser.* **2020**, *1502*, 012002. [[CrossRef](#)]
122. Hu, F.; Chen, B. Channel Coding Scheme for Relay Edge Computing Wireless Networks via Homomorphic Encryption and NOMA. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 1180–1192. [[CrossRef](#)]
123. Marques da Silva, M.; Dinis, R.; Martins, G. On the Performance of LDPC-Coded Massive MIMO Schemes with Power-Ordered NOMA Techniques. *Appl. Sci.* **2021**, *11*, 8684. [[CrossRef](#)]
124. Wu, Z.; Lu, K.; Jiang, C.; Shao, X. Comprehensive Study and Comparison on 5G NOMA Schemes. *IEEE Access* **2018**, *6*, 18511–18519. [[CrossRef](#)]
125. Zhu, K.; Wu, Z. Comprehensive Study on CC-LDPC, BC-LDPC and Polar Code. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Seoul, Republic of Korea, 6–9 April 2020; pp. 1–6. [[CrossRef](#)]
126. Niu, K.; Dai, J. Polar codes and polar processing for 6G wireless systems. *J. Commun.* **2020**, *41*, 9–17.
127. Pan, C.; Ren, H.; Wang, K.; Kolb, J.F.; El Kashlan, M.; Chen, M.; Di Renzo, M.; Hao, Y.; Wang, J.; Swindlehurst, A.L.; et al. Reconfigurable Intelligent Surfaces for 6G Systems: Principles, Applications, and Research Directions. *IEEE Commun. Mag.* **2021**, *59*, 14–20. [[CrossRef](#)]
128. Wang, J.; Tang, W.; Han, Y.; Jin, S.; Li, X.; Wen, C.K.; Cheng, Q.; Cui, T.J. Interplay Between RIS and AI in Wireless Communications: Fundamentals, Architectures, Applications, and Open Research Problems. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 2271–2288. [[CrossRef](#)]
129. Le, C.B.; Do, D.T.; Sur, S.N. Reconfigurable intelligent surface (RIS)-assisted wireless systems: Potentials for 6g and a case study. In *Advances in Communication, Devices and Networking: Proceedings of ICCDN 2020*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 367–378. [[CrossRef](#)]
130. Zhang, Y.; Di, B.; Zhang, H.; Lin, J.; Li, Y.; Song, L. Reconfigurable Intelligent Surface Aided Cell-Free MIMO Communications. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 775–779. [[CrossRef](#)]
131. Liu, H.; Yuan, X.; Zhang, Y.J.A. PHY-Layer Design Challenges in Reconfigurable Intelligent Surface Aided 6G Wireless Networks. In *6G Mobile Wireless Networks*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 53–81.
132. Zhao, Y.; Zhai, W.; Zhao, J.; Zhang, T.; Sun, S.; Niyato, D.; Lam, K.Y. A comprehensive survey of 6G wireless communications. *arXiv* **2020**, arXiv:2101.03889.
133. Shafin, R.; Liu, L.; Chandrasekhar, V.; Chen, H.; Reed, J.; Zhang, J.C. Artificial Intelligence-Enabled Cellular Networks: A Critical Path to beyond-5G and 6G. *IEEE Wirel. Commun.* **2020**, *27*, 212–217. [[CrossRef](#)]
134. Zhang, S.; Zhu, D. Towards artificial intelligence enabled 6G: State of the art, challenges, and opportunities. *Comput. Netw.* **2020**, *183*, 107556. [[CrossRef](#)]
135. Kekki, S.; Featherstone, W.; Fang, Y.; Kuure, P.; Li, A.; Ranjan, A.; Purkayastha, D.; Jiangping, F.; Frydman, D.; Verin, G.; et al. MEC in 5G networks. *ETSI White Pap.* **2018**, *28*, 1–28.
136. Pham, Q.V.; Fang, F.; Ha, V.N.; Piran, M.J.; Le, M.; Le, L.B.; Hwang, W.J.; Ding, Z. A Survey of Multi-Access Edge Computing in 5G and beyond: Fundamentals, Technology Integration, and State-of-the-Art. *IEEE Access* **2020**, *8*, 116974–117017. [[CrossRef](#)]
137. Mukherjee, S.; Lee, J. Edge Computing-Enabled Cell-Free Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2884–2899. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.