Overview of Interference Management in Massive Multiple- Input Multiple-Output Systems for 5G Networks

Yves Fidèle Aikoun
Hohai University https://orcid.org/0009-0002-2018-468X
Sheng Yin
shengyin@hhu.edu.cn
Hohai University

Systematic Review

Keywords: massive MIMO, 5G networks, NOMA, JT-CoMP, wireless communications

Posted Date: March 20th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3306477/v2

Additional Declarations: The authors declare no competing interests.

Review

Overview of Interference Management in Massive Multiple-Input Multiple-Output Systems for 5G Networks

Yves Fidèle Aikoun and Sheng Yin*

College of Information Science and Engineering, Hohai University, 210098 Nanjing, China
Correspondence: shengyin@hhu.edu.cn (S.Y.)

Abstract: New challenges have arisen in the field of wireless communication systems since the advent of Fifth Generation (5G) networks and the main objectives of delivering higher data rates, ultra-low latency, more reliability, and massive network capacity. While massive multiple-input multiple-output (MIMO) systems have emerged as a key technology for delivering a more uniform user experience to multiple users simultaneously, the deployment of multiple radiated elements in a confined area can lead to interference between the transmitted and received signals, resulting in the degradation of system efficiency. Therefore, interference management techniques are essential to mitigate this impact and enhance system performance. This review aims to explore interference management techniques in massive MIMO systems, including beamforming methods, non-orthogonal multiple access (NOMA), and joint transmission coordinated multi-point (JT-CoMP). The objective is to analyze the performance of these techniques in terms of the network capacity, coverage, and reliability and to compare their effectiveness in reducing interference. The valuable insights gained from this investigation will serve to inform the design and optimization of these systems.

Keywords: Index terms–massive MIMO; 5G networks; NOMA; JT-CoMP; wireless communications

1. Introduction

The rapid evolution of mobile communication networks has led to the development of the fifth generation of mobile networks. 5G NR (New Radio) is a new radio access technology developed by the 3rd Generation Partnership Project (3GPP) for the fifth-generation mobile networks, designed to be the global standard for the air interface powered by the New Radio technology to create wireless connectivity between the mobile equipment and the 5G network. New Radio was originally specified in 3GPP release 15 and is based on the same Orthogonal Frequency Division Multiplexing (OFDM) transmission scheme as the 4G LTE networks [1]. At its core, 5G utilizes a combination of various key technologies to deliver its transformative capabilities. One of the fundamental technologies is the use of higher frequency bands, including millimeter waves, providing a wider spectrum for data transmission and higher multi-Gbps peak data speeds [2,3]. These higher frequency bands enable faster data rates and increased availability, addressing the ever-growing demand for bandwidth-intensive applications and services, thereby supporting the massive Internet of Things (IoT) ecosystem and ensuring seamless connectivity in densely populated areas.

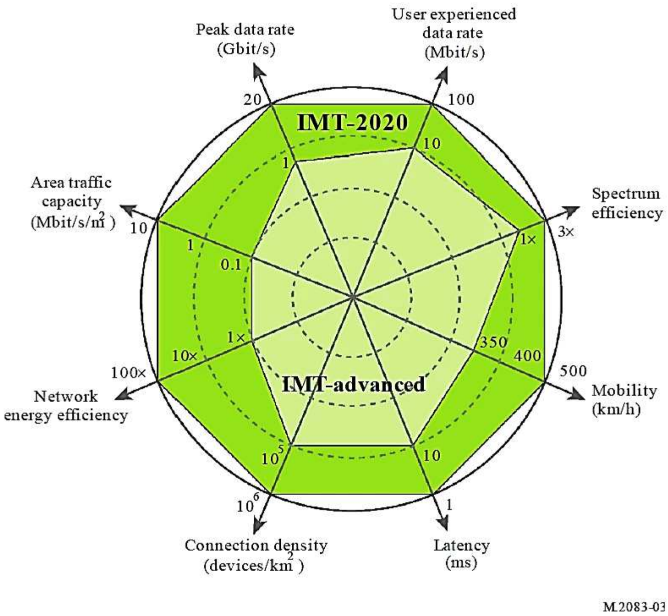

Figure 1. Enhancements of key performance requirements from IMT-Advanced to IMT-2020.

In addition to the key performance indicators mentioned in Figure 1, another important aspect of 5G is the utilization of advanced antenna technology known as massive MIMO. Massive MIMO involves the deployment of large arrays of antennas at both the base stations (gNodeB) and user terminals to improve spectral efficiency, energy efficiency, and reliability in wireless communications. However, massive MIMO systems also face several challenges, such as interference management, channel estimation, beamforming design, hardware complexity, and power consumption. Interference management is one of the most critical issues in massive MIMO systems, as it affects the performance and quality of service (QoS) of the users. Interference can arise from various sources, such as co-channel users, adjacent cells, non-orthogonal signals, and hardware impairments. Therefore, it is important to develop effective techniques to mitigate interference and improve the signal-to-interference-plus-noise ratio (SINR) of the desired users. This study examines three key techniques, such as beamforming, NOMA, and JT-CoMP. Each of these techniques offers a unique approach to mitigating interference and enhancing systems' overall performance.

2. Research Methodology

Our research adopts a systematic literature review approach to comprehensively overview the existing literature on interference management, massive MIMO systems, and 5G networks. This approach facilitates the identification, selection, and analysis of relevant studies to address targeted research inquiries. These inquiries encompass the origins of interference, principal interference management methodologies utilized within massive MIMO systems—especially in the context of the 5G NR—the effects of interference management techniques on system capacity, coverage, and reliability, as well as the comparative merits and drawbacks of various interference management strategies.

The research process involves searches of relevant databases, encompassing academic journals, conference proceedings, and online repositories. Employing a predefined set of keywords and combinations thereof—such as massive MIMO, 5G networks, interference management, beamforming, NOMA, JT-CoMP and their variations—ensures thorough exploration of the relevant literature.

Key databases, including but not restricted to IEEE Xplore, arXiv, ScienceDirect, and Google Scholar, are harnessed for their extensive repositories of peer-reviewed publications and scholarly articles pertinent to the realms of wireless communication and signal processing. The selection process adheres to stringent inclusion and exclusion criteria, ensuring the relevance and quality of the chosen papers. These criteria encompass factors such as alignment with interference management in massive MIMO systems, publication date, peer-review status, technological significance, and accessibility of full-text articles.

Relevant data, including the study objectives, methodologies, key findings, and conclusions, are extracted from selected papers for comparative analysis and the robust synthesis of insights.

To enhance reader comprehension, the review initiates with an exploration of the fundamental theories underlying massive MIMO systems, signal processing, and beamforming. This foundational understanding sets the stage for a comprehensive survey and analysis of these techniques. Subsequently, basic simulations are conducted using MATLAB (2023b) and its Communication Toolbox to illustrate the theories' efficacy. Specifically, the simulations focus on linear precoding techniques and the non-orthogonal multiple access (NOMA) sum rate. By integrating theoretical insights with practical demonstrations, the review aims to provide a holistic understanding of the subject matter.

2.1. Precoding Techniques Simulation

The precoders/decoders simulation is based on spatial multiplexing schemes wherein the data stream is subdivided into independent sub-streams, one for each transmit antenna employed. Consequently, these schemes provide a multiplexing gain and do not require explicit orthogonalization as needed for space–time block coding [4]. However, spatial multiplexing requires powerful decoding techniques at the receiver though, and this design highlights two ordered Successive Interference Cancellation (SIC) detection schemes [5].

The simulation compares the Bit Error Rate (BER) of Maximum Ratio Combining (MRC), Zero-Forcing (ZF) and Minimum Mean Squared Error (MMSE) alongside the optimal Maximum-Likelihood (ML) receiver. The evaluation is carried out in the context of spatial multiplexing schemes, employing an uncoded quadrature phase-shift keying (QPSK) modulation over independent transmit–receive 2 × 2 MIMO links affected by flat Rayleigh fading. Channel knowledge is assumed to be perfect at the receiver, with no feedback to the transmitter.

2.2. NOMA Sum Rate Simulation

In order to comprehensively assess the performance of NOMA in the specific context of our study, we employed 64 antennas at the base station, serving three users, each equipped with a single antenna, all operating in an environment characterized by Rayleigh fading-independent channel conditions.

Table 1. MIMO-NOMA simulation parameters.

System Parameters	
Number of Antennas at BS	64
Number of Users	3
Modulation Scheme	2nd Order QPSK
Bandwidth	1 MHz
Channel	Rayleigh fading-independent channel for each user
Distances of User to BS	Known

Based on the parameters outlined in Table 1, our simulation setup aimed to evaluate the key performance metrics:

Sum Rate vs. OMA Sum Rate: We conducted simulations to compare the sum rate achieved in our MIMO-NOMA design against the sum rate attained through orthogonal

multiple access (OMA). This comparison allowed us to quantify the benefits of employing NOMA in enhancing the overall system throughput.
Individual User Rate: We analyzed the rate achieved by each user within the MIMO-NOMA configuration. This assessment provided insights into the fairness and efficiency of the NOMA allocation strategy across users.
BER Analysis: Our simulations encompassed an assessment of the BER for the MIMO-NOMA system under the specified conditions. This allowed us to gauge the system's error performance and its sensitivity to channel impairments.

3. Massive MIMO Systems

Point-to-Point MIMO emerged in the late 1990s as the simplest form of MIMO, where a base station equipped with an antenna array (M) serves a terminal also equipped with an antenna array (K). Different terminals are orthogonally multiplexed via a combination of time- and frequency-division multiplexing [6] (pp. 6–8). In each channel end, a vector is transmitted and a vector is received. In the presence of additive white Gaussian noise at the receiver, Shannon theory yields the following formulas for the link spectral efficiency (in b/s/Hz) [6] (pp. 6–8).

$$C^{ul} = log_2 \left| I_M + \frac{\rho_{ul}}{K} HH^H \right|, \tag{1}$$

$$C^{dl} = log_2 \left| I_M + \frac{\rho_{dl}}{M} HH^H \right|, \tag{2}$$

where H is a $\mathbb{C}^{M\times K}$ matrix that represents the frequency response of the channel between the base station antenna array and the terminal antenna array; ρ_{ul} and ρ_{dl} are the uplink and downlink signal-to-noise ratios (SNRs), which are proportional to the corresponding total radiated powers. The receiver needs to know H for spectral efficiency values in (1) and (2), but the transmitter does not. However, performance can be improved if the transmitter also knows the channel state information (CSI). Increasing the link spectral efficiency may theoretically be achieved by using large arrays at both the transmitter and receiver [7].

The idea of a multiuser MIMO is for a single base station to serve a multiplicity of terminals using the same time-frequency resources. Effectively, the multiuser MIMO scenario is obtained from the Point-to-Point MIMO setup by breaking up the K-antenna terminal into multiple autonomous terminals [6] (pp. 8–10). The uplink and downlink sum spectral efficiencies are given by:

$$C^{ul} = log_2 |I_M + \rho_{ul} HH^H|, \tag{3}$$

$$C^{dl} = \max_{v_k \geq 0} log_2 |I_M + \rho_{dl} H D_v H^H|, \tag{4}$$

where ν is a vector and $\sum_{k=1}^{k} v_k \leq 1$. For a given ρ_{ul}, the total uplink power is K times greater than for the Point-to-Point MIMO model [6] (pp. 8–10). The possession of CSI is crucial to both (3) and (4). It is worth to note that the terminal antennas in the point-to-point case can cooperate, whereas the terminals in the multiuser case cannot, which does not compromise the uplink sum spectral efficiency as seen by comparing (1) and (3). Also, the downlink capacity (4) may exceed the downlink capacity in (2) for Point-to-Point MIMO, because (4) assumes that the base station knows H, whereas (2) does not. Multiuser MIMO has two fundamental advantages over Point-to-Point MIMO. First, it is much less sensitive to assumptions about the propagation environment. For example, the line-of-sight (LoS) conditions are stressing for Point-to-Point MIMO, but not for multiuser MIMO. Second, multiuser MIMO requires only single-antenna terminals. Notwithstanding these virtues, two factors seriously limit the practicality of multiuser MIMO in its originally conceived form. First, to achieve the spectral efficiencies in (3) and (4) requires complicated signal processing by both the base station and the terminals. Second, and more seriously, on the downlink both the base station and the terminals must know H, which requires substantial resources to be set aside for the transmission of pilots in both directions. For these reasons, the original form of multiuser MIMO is not scalable either with respect to M or to K [6,7] (pp. 8–10).

Massive MIMO is a useful and scalable version of multiuser MIMO [8]. According to Equations (3) and (4), further growth of the number of base station antennas (M) only yields logarithmically increasing throughputs while incurring linearly increasing amounts of time spent on training. There are three fundamental distinctions between massive MIMO and conventional multiuser MIMO. First, only the base station learns the frequency response of the channel (H). Second, the number of base station antennas is typically much larger than the K users, although this does not have to be the case. Third, simple linear signal processing is used both on the uplink and on the downlink owing to the channel hardening. These features render massive MIMO scalable with respect to the number of base station antennas, M [6] (pp. 10–15).

Cooperative and non-cooperative massive MIMO are two distinct configurations of coordination among the base stations or access points [9]. A cooperative massive MIMO system involves distributed antennas on different base stations to form a virtual antenna array to achieve MIMO communications [9]. Both CSI and data are shared among the collaborating BSs through backhaul links. This contributes to interference cancellation, and data are passed to the scheduled downlink users cooperatively from the BSs (sometimes using beamforming) [10]. Therefore, cooperative massive MIMO can improve the system performance by exploiting the spatial domain of mobile fading channels, mitigating the inter-cell interference, enhancing the coverage and reliability, and overcoming the limitations of conventional MIMO systems that require multiple antennas on each device [9]. Cooperative massive MIMO can be implemented using techniques such as coordinated multipoint (CoMP) or cell-free massive MIMO (CF mMIMO) [9,11].

Non-cooperative massive MIMO uses multiple antennas on each access point to serve multiple users without coordination with other access points [9], which can improve the system performance by using linear precoding and detection techniques that require the knowledge of the channel state information at the transmitter or receiver [12].

While massive MIMO allows simultaneous communication with multiple users, enhanced spectral efficiency, and better QoS, it also introduces new challenges related to interference management. With a high density of antennas in different cells and the number of users, the interference among the transmitted and received signals becomes more pronounced, potentially degrading the system performance. Therefore, effective interference management techniques are essential to harness the full potential of massive MIMO systems in 5G NR networks.

The rest of this paper is structured as follows. In Section 4, we introduce some signal processing methods, such as channel estimation, signal detection, and precoding, which are used to improve the signals' transmission and reception. Section 5 is devoted to explaining the beamforming method, which employs antenna arrays to steer the signals toward a desired direction. The NOMA and JT-CoMP methods are discussed in Section 6. These methods are used to cancel the interference caused by multiple users sharing the same spectrum or multiple base stations serving the same user. The conclusions follow in Section 7.

4. Signal Processing Methods

The quality of the wireless transmission depends on the knowledge of the CSI at both ends of the link, which reflects the effects of shadowing, scattering, fading, and path loss on the signal propagation channel. In massive MIMO systems, the pilot signals are used to estimate the CSI in both the uplink and downlink directions to optimize the equalizer at the receiver. The length of the pilot signals should be enough to match the number of antennas at the transmitter or the number of users. However, pilot contamination arises as a new matter, which happens when different users in neighboring cells use the same pilot signals. This leads to interference and error in the channel estimation. To solve this problem, various channel estimation methods have been proposed that use advanced algorithms to select the best channel responses from the received pilot signals, often with less pilots than users.

4.1. Channel Estimation

Massive MIMO systems require periodic channel estimation to cope with the non-stationary wireless channel. The channel estimation can be performed in different ways depending on the duplexing mode, the direction of the CSI feedback, and the number of antennas and users. Time division duplexing (TDD) is one of simplest duplexing modes, where the channel is estimated in one direction and applied in both directions assuming reciprocity. TDD systems have the advantages of an independent CSI acquisition time and coherent antenna processing at the BS [7,13]. The BS needs the CSI for both downlink precoding and uplink detection, which can be obtained by using pilot signals proportional to the number of transmit antennas.

In frequency division duplexing (FDD), the uplink and downlink use different frequency bands, which means a different CSI for each link. The uplink channel estimation at the BS is performed by letting all the users send different pilot sequences. To obtain the CSI for the downlink channel, the BS transmits pilot symbols to all the users. The users respond by the estimated CSI for the downlink channels [13]. The time required for uplink pilot transmission is independent of the number of antennas at the BS; however, the bandwidth resource required to transmit the downlink pilot symbols is proportional to the number of antennas at the BS. As the number of antennas at the BS increases, a valuable uplink band is required to transmit the downlink pilot symbols. Therefore, the downlink channel estimation becomes infeasible for FDD systems [7], but the TDD approach can resolve this issue due to channel reciprocity, as only the CSI for the uplink needs to be estimated. This is used by the BSs to detect the uplink data and to generate beamforming vectors for downlink data transmission. In addition, linear MMSE-based channel estimation can provide near-optimal performance with low complexity [14], even if the pilot sequences employed by users in neighboring cells may no longer be orthogonal to those within the cell due to the limited channel coherence time, leading to a pilot contamination problem [15].

In 4G and 5G, channel quantization, codebook-based CSI feedback, and CS-based CSI feedback form the foundation for FDD system implementation. Each approach addresses the challenge of conveying channel information while managing the trade-offs between accuracy and feedback overhead.

Codebooks have been an indispensable part of wireless communication standards since the first release of Long-Term Evolution in 2009. They offer an efficient way to acquire the CSI for multiple antenna systems. Nowadays, a codebook is not limited to a set of pre-defined precoders but refers to a CSI feedback framework, which is more and more sophisticated [16]. The meaning of codebook extends to the whole CSI report mechanism, which helps the base station compute the precoding matrix with the feedback from the UEs. The codebook is known to the user and the BS. The user searches for the codeword that is the closest to the downlink CSI and feeds back the corresponding index to the BS. Upon receiving the index, the BS can obtain the channel by looking up the shared codebook [17]. The limitations of this approach stem from the complexity of the search algorithm and the lower accuracy of the CSI.

Compressive Sensing (CS)-based CSI feedback involves the concept of sparse representation, where a signal can be expressed as a linear combination of a few essential components from a predefined and overcomplete dictionary [18]. The CSI matrix exhibits sparsity in specific domains, including time, spatial, spatial–temporal, and spatial–frequency domains [18]. This means that not all the details of the signal need to be explicitly conveyed; instead, a sparse representation captures the most representative information. CS can be used to reduce the overhead of the downlink CSI. Given that the number of scatter clusters is much smaller than that of the transmit antennas at the BS in massive MIMO systems, the CSI matrix can be represented by much fewer parameters, and the spatial domain turns into the sparse angular domain using discrete Fourier transform (DFT) [17].

4.2. Signal Detection

Linear and non-linear detectors are two types of detectors for massive MIMO systems, where the transmitter and the receiver need to recover the transmitted data from the received signals. These detectors differ in the complexity and accuracy of the signal detection, as well as in the techniques they use to exploit the CSI [19].

Linear detectors use linear operations such as matrix inversion, multiplication, or projection to perform signal detection with low complexity and high parallelism, but they suffer from performance degradation due to noise enhancement, interference, or channel ill-conditioning [15]. Well-known detectors, such as matched-filter (MF) receivers, zero-forcing (ZF) receivers, and minimum mean-square-error (MMSE) receivers can asymptotically achieve capacity as the number of antennas at the BS is large enough compared to the number of users and the channel vectors from different users are independent [7].

In addition, signal detection can be performed using non-linear operations such as tree search, lattice reduction, or optimization, known as non-linear detectors. Non-linear detectors have high accuracy and can approach the optimal maximum likelihood (ML) performance, but they have high complexity and low parallelism, especially when the number of antennas, users, or subcarriers is great [19]. Lattice Reduction (LR)-based linear detection improves the performance of ordinary linear detection. However, they use a linear transformation on an equivalent system model obtained using LR techniques instead of using it on the received signal model. The new channel matrix has more orthogonality than the old one.

4.3. Precoding Techniques

MIMO is a technique that uses multiple channels between the BS and users with suitable space–time coding to enhance the system throughput. However, the massive MIMO system operates by space division multiplexing with the knowledge of the CSI of every link between the BS and a user.

Precoding in massive MIMO systems is essentially a beamforming approach that enables multi-stream transmission [13]. By considering practical antenna array structures, the received signals from different terminals are combined in the uplink using appropriate decoding. The more the antennas are used, the finer the spatial focusing can be, so that a large array is built in practice. The trade-off is between power consumption and performance, especially in high bit rate scenarios. The use of non-linear but power-efficient RF front-end amplifiers can reduce power consumption, but it may cause signal distortion. Therefore, the transmit signal should have a low peak-to-average-power-ratio (PAPR), which can be achieved by using various PAPR reduction techniques and precoding [20,21]. On the other hand, the use of low-complexity precoding methods in large-scale systems can reduce the computational complexity of the precoder, but it may have suboptimal performance or higher noise enhancement. Then, detector/precoder designs with enhanced power consumption and low estimation complexity are difficult to obtain but extremely important, particularly when the number of antennas increases.

Linear and non-linear precoding techniques are both applicable for conventional MIMO systems. Linear precoders are simple but provide poor BER performance. By contrast, non-linear detectors provide a reasonable BER performance but have a high computational complexity. The best-known non-linear precoder techniques that can be used for multiuser MIMO systems are the dirty-paper-coding (DPC), vector perturbation (VP), and lattice-aided methods [22]. Such precoders can be used to obtain better performances compared to using linear precoders at the cost of higher estimation complexity. However, when the number of antennas at the BS increases, linear precoders, such as MRC, ZF, and MMSE, become nearly optimal. Therefore, it is more practical to use low-complexity linear precoding techniques in massive MIMO systems [7].

Using MRC receivers, BSs attempt to obtain the maximum SNR for every stream and ignore the influence of other multiuser interference. MRC receivers are advantageous in that they simplify signal processing; however, MRCs perform poorly in interference-limited scenarios because they do not address the effects of multiuser interference. In comparison to MRC receivers, ZFs consider multiuser interference in their calculations, but they do not consider noise effects [23]. Zero-forcing precoding performs better in multiuser scenarios by estimating the orthogonal complement of each stream of the multiuser interference. In terms of power consumption and capacity efficiency, conjugate beamforming could obtain better overall computational measures compared to zero-forcing precoding because of the greater number of served terminals. By optimizing the management of transmitted power, conjugate beamforming is more robust than ZF and may thus be preferable regardless of the computational aspects [24].

MMSE precoding is the optimal linear precoding in a massive MIMO downlink system [13]. It minimizes the mean squared error between the transmitted and received signals as well as balances between reducing the interference and preserving the SNR. The MMSE outperforms ZF and MF in multi-cell MIMO systems. When linear precoding methods including MF, ZF and MMSE are used, the transmit signal from the BS can be expressed as:

$$X^{MF} = \frac{1}{\sqrt{\alpha}} H^{*} S_{d}, \tag{5}$$

where α is a power normalization factor (vector normalization is better for ZF, while matrix normalization is better for MF).

$$X^{ZF} = \frac{1}{\sqrt{\alpha}} H^{*} (H^{T} H^{*})^{-1} S_{d}, \tag{6}$$

$$X^{MMSE} = \frac{1}{\sqrt{\alpha}} H^{*} \left(H^{T} H^{*} + \frac{1}{SNIR} I_{k} \right)^{-1} S_{d}, \tag{7}$$

where H^{*} is the Hermitian operation of the channel matrix, S_d is source information vector and I_k is the identity matrix for the kth user.

4.4. Simulation Results

The graphic compares the BER of the MRC, ZF and MMSE alongside the optimal ML receiver. The detailed information about this simulation is given in Section 2.1.

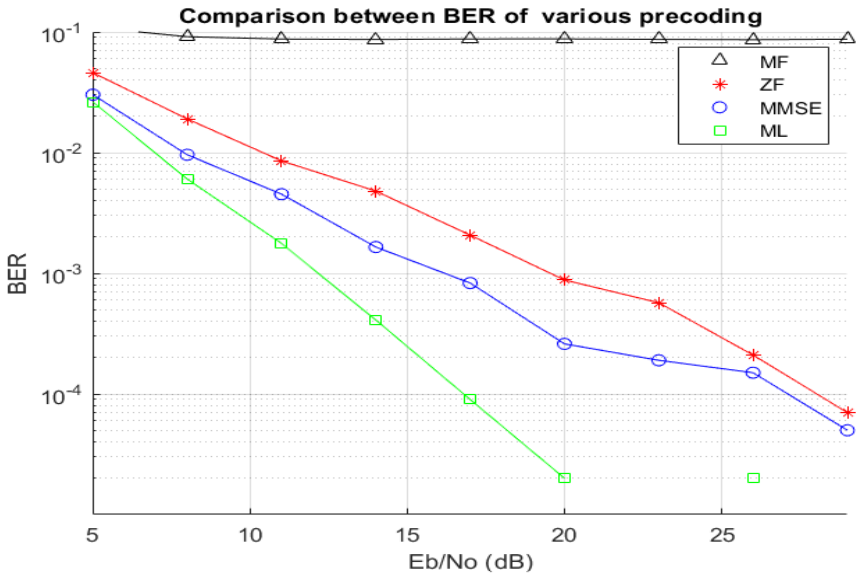

Figure 2. Comparison between the bit error rates of various precodings.

Observations from the results, as depicted in Figure 2, confirm that the ML receiver achieves the highest performance, outperforming MMSE and ZF receivers. Considering the receiver complexity, the ML receiver's computational load increases exponentially with the number of transmit antennas. In contrast, the ZF and MMSE receivers exhibit linear complexity and incorporate successive interference cancellation techniques. Linear precoders are also applicable in multi-cell massive MIMO systems, where cooperating base stations (BSs) collaborate to jointly serve users across various cells, echoing the perspective presented in [25].

5. Beamforming Method

Beamforming is a process formulated to produce the radiated beam patterns of the antennas by completely building up the processed signals in the direction of the desired terminals and cancelling the beams of the interfering signals. This can be achieved by using a finite impulse response (FIR) filter, which can adjust its weights adaptively to achieve optimal beamforming [24]. Beamforming has several benefits for massive MIMO systems, such as higher energy efficiency, better spectral efficiency, and suitability for mm-wave bands. The beamforming concept is widely used in advanced wireless communication systems, such as LTE, and recently in 5G wireless systems.

To communicate with all the users in a multiuser configuration, the zero-forcing beamforming algorithm is applied at the base station due to its low complexity. The BS can communicate to each user via individual beams unless K users are in the same direction, which can create overlap between beams and generate interference. It shows in [26] that the weight w_i generated by the zero-forcing beamforming algorithm for the user i ensures that the BS can communicate with the user via its own beam, and the interference with other users can be avoided if the beam is sharp enough or the users are far from each other. So then, the transmitted signal at the BS is a multiplication of the information signals of all the users and their own weights.

$$X^{BEAM} = \sum_{k=1}^{K} w_k \sqrt{a_k p} S_k, \tag{8}$$

where p denotes the transmit power of BS, a_k the power allocation coefficient of the user k, and S_k represents the information signal of user k. The received signal at the user i still includes the thermal noise $(n_i \in \sigma_i^2 \mathbb{C}^{l\times 1})$ and the signal of others because of the overlap of the beams.

$$Y_i^{BEAM} = H_i w_i \sqrt{a_i p} S_i + H_i \sum_{k=1,\neq i}^{K} w_k \sqrt{a_k p} S_k + n_i, \quad (9)$$

where the weight depends on the channel response matrix H_i and $H_i w_i = 0\ (i \neq k)$, $H_i w_i = I$ in case the users are far from each other. The SINR of user i is the result of $(1-\rho^2)$ due to the channel estimation error, and $\gamma_{ik}\rho^2$ is due to the overlap of the beams of users i and k.

$$SINR_i^{BEAM} = \frac{\rho^2 \Omega_i a_i \delta}{(1-\rho^2)\Omega_i a_i \delta + \sum_{k=1,\neq i}^{K} (1-\rho^2+\gamma_{ik}\rho^2)\Omega_i a_k \delta + 1}, \quad (10)$$

where $\delta = \frac{p}{\sigma^2}$, Ωi denotes the variance of the channel gain between the BS and user i, and the overlap factor of the beams γ_{ik}. As the number of antennas at the BS is not enough to generate a sharp beam and many users are close to each other in a high-density user environment, the interference signal remains and depends on an overlap factor of the beams. In addition, most of the several well-known beamforming algorithms have assumed perfect knowledge of the CSI at both transceivers, and they do not adapt to the changes in the transmission environment and the relative position between transceivers.

Several works have been conducted to classify the beamforming techniques according to their characteristics, such as analogue beamforming, digital beamforming, and hybrid (digital and analogue) beamforming [27,28], which is a most considerate technique for 5G to partition beamforming between the digital and RF domains to reduce the cost associated with the number of RF signal chains (number of antennas). Ref. [29] classified the beamforming methods into switched beamforming and adaptive beamforming, as show the figure.

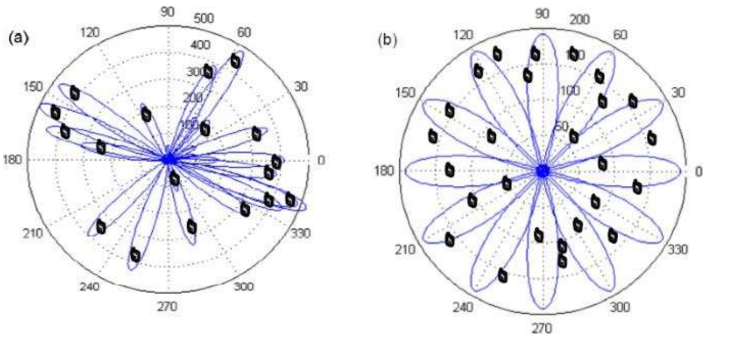

Figure 3. Adaptive beamforming (**a**) and switched beamforming (**b**).

In switched (Figure 3.b) array beamforming, one beam can serve more than one terminal at the same time. However, adaptive (Figure 3.a) array systems can create specific beam shapes and direct the main lobe toward a desired mobile station, and consequently, null toward the interfering signals. Adaptive beamforming is more suitable in high user density environments but requires the BS to update the location of the mobile station, which is a hard task because there may be too many mobile stations in real time that overload the process. It is considerably more difficult to put an adaptive beamforming system into practice than a switched beamforming system [23]; however, the recent studies related to massive MIMO prefer adaptative beamforming to switched beamforming because of its reliability for 5G requirements.

Another factor that can improve the quality-of-service of beamforming techniques is the use of mm-wave bands, which are high-frequency bands. In mm-wave bands, the antenna array is extremely small owing to the size of the wavelength and the beam width being extremely sharp, which means they can only cover a short distance between the base station and the users. Most current wireless communication applications are still focused on narrowband beamforming. However, wideband beamforming becomes important for

IoT machine-to-machine applications that require 5G standards to reach high speeds and high capacities regardless of the signal processing complexity.

6. NOMA and JT-CoMP Methods

By considering the high user density, the conventional multiple access techniques are encountering limitations in managing interference and ensuring resource allocation. To overcome these problems, NOMA and JT-CoMP emerged as major solutions, where they enable power-domain user multiplexing and the exploitation of the channel-gain difference among the users in a cellular system, two features that were not exploited in past cellular systems.

6.1. Non-Orthogonal Multiple Access Method

NOMA is a widely used approach to deal with the problem of multiple nearby users by providing multiple access based on the power domain [26,30,31]. It is considered as a promising multiple access technique for 5G mobile networks due to its superior spectral efficiency [32]. OFDMA and NOMA are both techniques that allow multiple users to share the same radio resource in wireless networks; however, they differ in how they allocate the resource among the users. OFDMA uses orthogonal subcarriers to divide the resource into subchannels and assigns each subchannel to one user at a time, while NOMA uses non-orthogonal signals to superimpose the data of multiple users on the same subchannel and uses different power levels or codes at the transmitter and the SIC method to detect the signals at the receiver.

The NOMA method assigns more power to users with low channel gain and less power to users with high channel gain. The user with low channel gain decodes its own signal by considering the signals of other users as interference [26]. The user with high channel gain applies the SIC, operates following the principle of the NOMA method to separate superimposed symbols and removes the inter-user interference, and then decodes its own signal.

The broadcast information signals simultaneously send to K users with different power levers [26], which can be formulated as follows:

$$X^{NOMA} = \sum_{k=1}^{K} \sqrt{a_k p} S_k, \tag{11}$$

The user i receives the information signals of all the users:

$$Y_i^{NOMA} = H_i \sum_{k=1}^{K} \sqrt{a_k p} S_k + n_i, \tag{12}$$

$$Y_i^{NOMA} = H_i \sqrt{a_i p} S_i + H_i \sum_{k=1,\neq i}^{K} \sqrt{a_k p} S_k + n_i$$

$$SINR_i^{NOMA} = \frac{\rho^2 \Omega_i a_i \delta}{\sum_{k=1}^{i-1}(1-\rho^2)\Omega_i a_k \delta + \sum_{k=i+1}^{K} \Omega_i a_k \delta + 1}, \tag{13}$$

As mentioned in [26], a_k the power allocation coefficient is inversely proportional to the channel gain, the user i firstly utilizes the SIC operation to cancel the signals of users $1, 2, \cdots, i-1$ and then detects its own signal while considering the signal of users $i+1, \cdots, K$ as interference. However, because of the outdated CSIT, the interference from users $1, 2, \cdots, i-1$ is not completely cancelled. Therefore, the achievable rate at user i is given by $R_{i,NOMA} = log_2\left(1 + SINR_i^{NOMA}\right)$.

In contrast, the MIMO-OMA operation divides our transmission into K (number of users) equal time slots and allocates the total power resource to successively transmit to each user within their respective time slot. Therefore, the achievable rate of MIMO-OMA for user i is given by $R_{i,OMA} = {}^{1}/_{K}\, log_2\left(1 + SINR_i^{OMA}\right)$, reflecting the fact that only ${}^{1}/_{K}$ of the time slot is utilized for communication with each user. Conversely, in MIMO-NOMA, the entire time slot is utilized for simultaneous transmission to K users, leading to enhanced efficiency.

6.2. Simulation Results

A base station equipped with 64 antennas served three users located, respectively, at 500 m, 200 m and 50 m from the BS. This simulation parameters are given in Section 2.2.

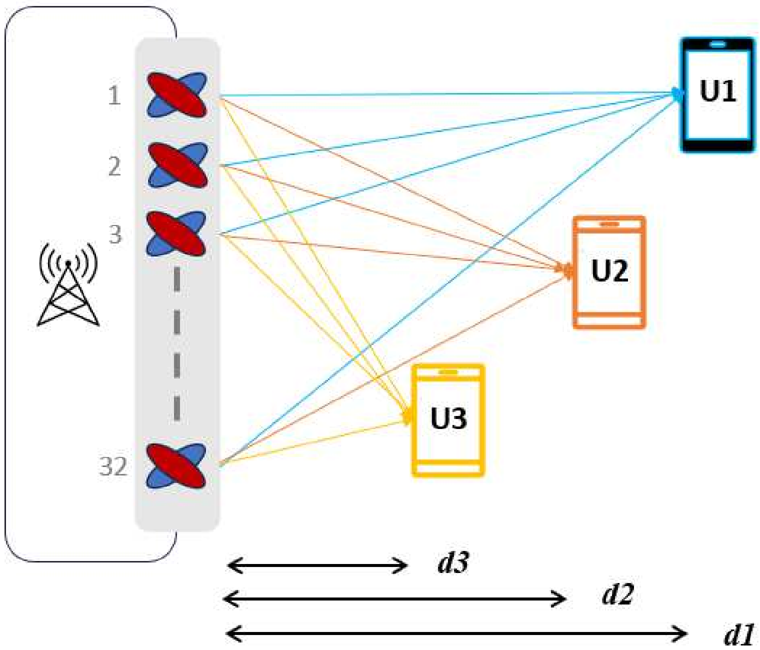

Figure 4. Illustrates 8 × 8 MIMO-NOMA serving 3 users under the Rayleigh fading channel.

The achievable sum rate of MIMO-NOMA is $R_{NOMA} = \sum_{i=1}^{3} R_{i,NOMA}$, while that of MIMO-OMA is $R_{OMA} = \sum_{i=1}^{3} R_{i,OMA}$ respect to the Figure 4.

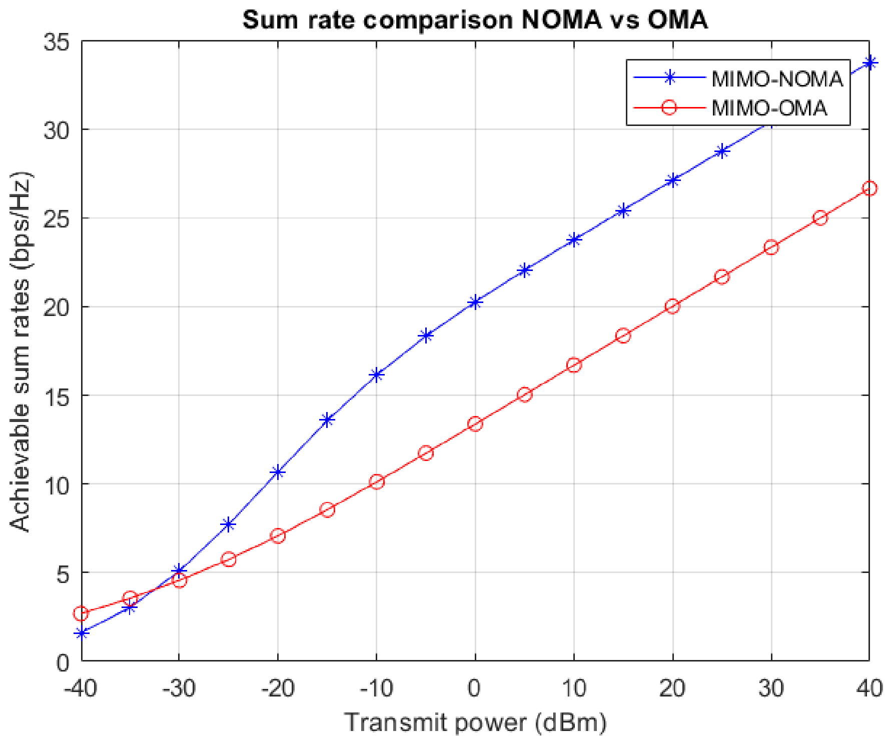

Figure 5. Sum rate comparison NOMA vs. OMA.

As shown in the Figure 5, MIMO-NOMA achieves a higher sum rate than MIMO-OMA because it allows the users to share the same frequency resource at the same time. As it shows in the individual user rate graph, the weak user exhibits saturation in its achievable rate beyond a transmit power of 10 dBm. This consistent characteristic is observed universally across all the NOMA networks. The interference encountered by the weak user results in the saturation of its achievable rate. However, this saturation poses no issue if the required data rate of the weak user is below the saturation limit. In contrast, this issue is absent in OMA, as the weak user does not contend with interference from simultaneous transmissions.

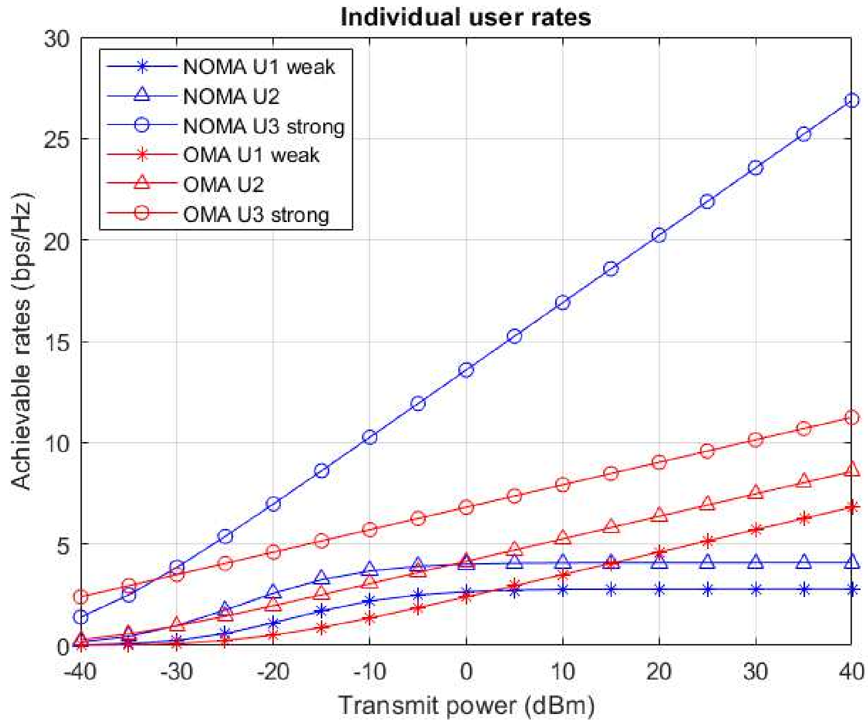

Figure 6. Comparison of the individual user data rate.

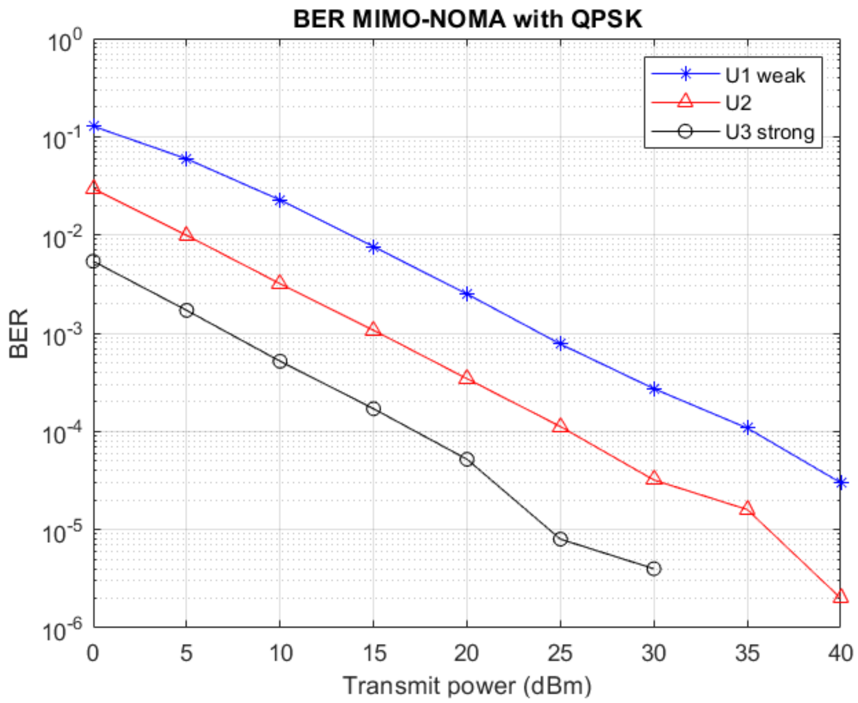

Figure 7. Users' BER comparison.

From the Figure 6 and 7, the user 1 has the worst BER of the three users, as it receives interference from both user 2 and user 3. User 2 has a moderate BER, as it only receives interference from user 3. User 3 has the best BER, as it does not face any interference.

In [32,33], the authors made a proposal of a combination of the NOMA and beamforming methods especially for downlink multi-users systems. They considered a multi-antenna base station that uses beamforming to eliminate the interference between clusters and makes a trade-off between the complexity and system performance. The use of

NOMA with the SIC approach was intended to remove the interference within clusters. Consequently, the inter-cluster and intra-cluster interferences were completely cancelled. Therefore, combining the NOMA and beamforming methods can enhance the spectral efficiency of the system. However, it is essential to note that while NOMA effectively allocates resources to users with lower channel gains, it may also result in a reduced SINR for these users due to the interference experienced. So, despite the overall spectral efficiency improvement, careful consideration is needed to mitigate the potential reduction in the SINR for users with lower channel gains.

6.3. Joint Transmission Coordinated Multi-Point

The concept of joint transmission coordinated multi-point (JT-CoMP) emerges as a promising solution to overcome the challenges posed by the increasing user density in modern wireless networks. JT-CoMP addresses these challenges through a collaborative approach that leverages the strengths of multiple base stations (BSs) working in tandem to transmit data to a single user device at the same time in order to improve the signal quality and reduce the interference. There are different types of JT-CoMP, such as coherent and non-coherent transmission, centralized and distributed scheduling, and static and dynamic clustering. These types differ in how the base stations coordinate their transmissions, how the user device combines the received signals, and how the network selects the best base stations to serve the user device [34,35]. At the heart of JT-CoMP lies the principle of coordinated multi-point transmission, where neighboring base stations form a coordinated cluster to jointly serve users within their overlapping coverage areas. This cooperative approach contrasts starkly with traditional cellular networks, where each base station independently serves its designated users. JT-CoMP harnesses the spatial diversity of massive MIMO to ensure that users experience enhanced signal quality, higher data rates, and consistent connectivity, even at the cell edges [36,37].

One of the primary advantages of JT-CoMP is its capability to effectively mitigate interference. By converting the interfering signals from the other base stations into useful signals for the user device, the signal overlap is minimized and unwanted interference is nullified. This way, the user device can combine the received signals from different base stations and achieve a higher SINR and throughput. Furthermore, JT-CoMP contributes to load balancing across base stations. Dynamic resource allocation and user association ensure that the available network resources are optimally distributed, avoiding network congestion and ensuring a more equitable distribution of traffic [34,35]. However, effective cooperation among base stations requires accurate synchronization, low latency backhaul links, and the exchange of the real-time CSI. Ensuring seamless coordination without introducing additional latency demands advanced network architecture and management. Additionally, JT-CoMP's complexity increases as the number of cooperating base stations and users grows.

Some research proposes a combination of JT-CoMP and NOMA for joint transmission coordination and interference management by comparing the performance of JT-CoMP-NOMA with two benchmark schemes, such as JT-CoMP-OMA and NOMA without CoMP [38]. The authors show that JT-CoMP-NOMA can achieve a higher network sum rate than both benchmark schemes, especially when the number of users and base stations is great. JT-CoMP-NOMA can improve the fairness among users by increasing the minimum rate of cell-edge users, who suffer from severe inter-cell interference. This work also demonstrates that JT-CoMP-NOMA can effectively exploit the spatial diversity and multiplexing gains of CoMP and NOMA, respectively. However, there are some limitations and challenges of JT-CoMP-NOMA, such as the high computational complexity and the large feedback overhead due to the CSI imperfection.

7. Conclusions

In the pursuit of delivering reliable, high-quality connectivity in environments characterized by a high user density, massive MIMO systems prove instrumental in meeting the stringent requirements of 5G networks. Through a comprehensive exploration of the three main interference management approaches, beamforming harnesses spatial precision to enhance signal quality, while NOMA introduces power-based interference mitigation through dynamic user power allocations and JT-CoMP facilitates coordinated base station cooperation to nullify inter-cell interference.

The investigation demonstrates that combining these techniques holds immense promise for overcoming the interference challenges in high-density scenarios. The synergistic integration of beamforming, NOMA, and JT-CoMP may effectively contribute to achieving the full potential of 5G networks. Additionally, combining precoding and equalization yields positive outcomes for canceling interference in multiuser MIMO systems.

However, there are several interconnected design issues that need careful consideration and resolution, such as developing more advanced signal processing techniques for accurate detection with low complexity, using deep learning for channel estimation to reduce the training time in FDD systems, exploring the use of mm-wave frequency bands, and applying adaptive beamforming to improve the channel gain, capacity, received power, and reduce latency.

Author Contributions: Y.F.A. and S.Y. both contributed significantly to the conception, design, and execution of the study. Y.F.A. was responsible for the primary investigation, simulations, and manuscript writing. S.Y., as the corresponding and second author, provided valuable guidance throughout the research process, including study supervision, methodology refinement, and manuscript review. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding, And the APC was funded by Dr. Yi Mao.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

5G NR	Fifth-Generation New Radio
MIMO	Multiple-Input Multiple-Output
NOMA	Non-Orthogonal Multiple Access
JT-CoMP	Joint Transmission Coordinated Multiple-Point
3GPP	3rd Generation Partnership Project
OFDM	Orthogonal Frequency Division Multiplexing
LTE	Long-Term Evolution
4G	Fourth Generation
IoT	Internet of Things
SINR	Signal-to-Interference-Plus-Noise Ratio
SIC	Successive Interference Cancellation
BER	Bit Error Rate
MRC	Maximum Ratio Combining
ZF	Zero-Forcing
MMSE	Minimum Mean-Squared-Error
ML	Maximum Likelihood
QPSK	Quadrature Phase-Shift Keying
OMA	Orthogonal Multiple Access
SNR	Signal-to-Noise Ratios
CSI	Channel State Information
LoS	Line-of-Sight
CF mMIMO	Cell-Free Massive MIMO
BS	Base Station
QoS	Quality of Service

TDD	Time Division Duplexing
FDD	Frequency Division Duplexing
MF	Matched-Filter
LR	Lattice Reduction
PAPR	Peak-to-Average-Power-Ratio
RF	Radio Frequency
DPC	Dirty-Paper-Coding
VP	Vector Perturbation
FIR	Finite Impulse Response

References

1. Adnan Ghayas. 5G New Radio: What is 5G NR Technology and How Does It Work? Available online: commsbrief.com (accessed on 1 October 2023).
2. ITU. M.2083, Recommendation ITU-R, IMT Vision—Framework and Overall Objectives of The Future Development of IMT for 2020 and Beyond. Available online: https://www.itu.int/rec/R-REC-M.2083-0-201509-I/en.itu.int/rec/R-REC-M.2083-0-201509-I/en (accessed on 1 December 2020).
3. National Instruments. 5G Massive MIMO Testbed: From Theory to Reality. Available online: ni.com (accessed on 1 October 2023).
4. Shang, Y.; Xia, X.-G. On fast recursive algorithms for V-blast with optimal ordered SIC detection. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 2860–2865, https://doi.org/10.1109/twc.2009.081466.
5. Li, X.; Huang, H.C.; Lozano, A.; Foschini, G.J. Reduced-complexity detection algorithms for systems using multi-element arrays. In *Globecom'00-IEEE. Global Telecommunications Conference. Conference Record (Cat. No. 00CH37137)*; IEEE: New York, NY, USA, 2000; pp. 1072–1076. https://doi.org/10.1109/GLOCOM.2000.891302.
6. Marzetta, T.L.; Larsson, E.G.; Yang, H. *Fundamentals of Massive MIMO*; Cambridge University Press: Cambridge, UK, 2016.
7. 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. https://doi.org/10.1109/jstsp.2014.2317671.
8. Marzetta, T.L. How Much Training is Required for Multiuser Mimo? In Proceedings of the 2006 Fortieth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 29 October 2006; pp. 359–363. https://doi.org/10.1109/ACSSC.2006.354768.
9. 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. https://doi.org/10.1109/ACCESS.2021.3124816.
10. Gesbert, D.; Hanly, S.; Huang, H.; Shitz, S.S.; Simeone, O.; Yu, W. Multi-Cell MIMO Cooperative Networks: A New Look at Interference. *IEEE J. Sel. Areas Commun.* **2010**, *28*, 1380–1408. https://doi.org/10.1109/JSAC.2010.101202.
11. Nguyen, T.T.; Nguyen, K.-K. Joint Learning and Optimization-Based Resource Management in Hybrid Network of Cooperative and Non-Cooperative Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2023**, *72*, 12001–12015. https://doi.org/10.1109/TVT.2023.3267660.
12. He, C.; Gitlin, R.D. System performance of cooperative massive MIMO downlink 5G cellular systems. In Proceedings of the 2016 IEEE 17th Annual Wireless and Microwave Technology Conference (WAMICON), Clearwater, FL, USA, 2 June 2016; pp. 1–5. https://doi.org/10.1109/WAMICON.2016.7483862.
13. Hassan, N.; Fernando, X. Massive MIMO Wireless Networks: An Overview. *Electronics* **2017**, *6*, 63. https://doi.org/10.3390/electronics6030063.
14. Hoydis, J.; Brink, S.T.; Debbah, M. Comparison of linear precoding schemes for downlink massive MIMO. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 2135–2139. https://doi.org/10.1109/ICC.2012.6364360.
15. Liu, L.; Peng, G.; Wei, S. *Massive MIMO Detection Algorithm and VLSI Architecture*; Springer Science and Business Media LLC: Dordrecht, The Netherlands, 2019.
16. Qin, Z.; Yin, H. A review of codebooks for CSI feedback in 5G new radio and beyond. *arXiv* **2023**, arXiv:2302.09222.
17. Guo, J.; Wen, C.-K.; Jin, S.; Li, G.Y. Overview of Deep Learning-based CSI Feedback in Massive MIMO Systems. June 2022. Available online: http://arxiv.org/abs/2206.14383 (accessed on 1 December 2023).
18. Qin, Z.; Fan, J.; Liu, Y.; Gao, Y.; Li, G.Y. Sparse Representation for Wireless Communications: A Compressive Sensing Approach *IEEE Signal Process. Mag*. **2018**, *35*, 40–58. https://doi.org/10.1109/MSP.2018.2789521.
19. Albreem, M.A.; Juntti, M.; Shahabuddin, S. Massive MIMO Detection Techniques: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3109–3132, https://doi.org/10.1109/comst.2019.2935810.
20. Albreem, M.A.; Al Habbash, A.H.; Abu-Hudrouss, A.M.; Ikki, S.S. Overview of Precoding Techniques for Massive MIMO. *IEEE Access* **2021**, *9*, 60764–60801. https://doi.org/10.1109/ACCESS.2021.3073325.
21. Liang, Q.; Wang, W.; Liu, X.; Zhenyu; Baoju, N.; , Z. (Eds.) Lecture Notes in Electrical Engineering 878 Communications, Signal Processing, and Systems. Available online: https://link.springer.com/bookseries/7818 (accessed on 1 February 2024).

22. Masouros, C.; Sellathurai, M.; Ratnarajah, T. A Low-Complexity Sequential Encoder for Threshold Vector Perturbation. *IEEE Commun. Lett.* **2013**, *17*, 2225–2228. https://doi.org/10.1109/LCOMM.2013.101813.131811.
23. Ali, E.; Ismail, M.; Nordin, R.; Abdulah, N.F. Beamforming techniques for massive MIMO systems in 5G: overview, classification, and trends for future research. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 753–772. https://doi.org/10.1631/FITEE.1601817.
24. Yang, H.; Marzetta, T.L. Performance of Conjugate and Zero-Forcing Beamforming in Large-Scale Antenna Systems *IEEE J. Sel. Areas Commun.* **2013**, *31*, 172–179. https://doi.org/10.1109/JSAC.2013.130206.
25. Zakhour, R.; Hanly, S.V. Base Station Cooperation on the Downlink: Large System Analysis. *IEEE Trans. Inf. Theory* **2012**, *58*, 2079–2106. https://doi.org/10.1109/TIT.2011.2177770.
26. Phuong, N.T.; Van Son, V.; Hiep, P.T. Combining precoding and equalization for interference cancellation in MU-MIMO systems with high density users. *EURASIP J. Wirel. Commun. Netw.* **2022**, *2022*, 1–19. https://doi.org/10.1186/s13638-022-02118-2.
27. Hur, S.; Kim, T.; Love, D.J.; Krogmeier, J.V.; Thomas, T.A.; Ghosh, A. Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks. *IEEE Trans. Commun.* **2013**, *61*, 4391–4403. https://doi.org/10.1109/TCOMM.2013.090513.120848.
28. Bogale, T.E.; Le, L.B. Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8 December 2014; pp. 4066–4071. https://doi.org/10.1109/GLOCOM.2014.7037444.
29. Gotsis, K.A.; Sahalos, J.N. Beamforming in 3G and 4G Mobile Communications: The Switched-Beam Approach. Available online: www.3gpp.org (accessed on 1 September 2023).
30. Hoang, T.M.; Van Son, V.; Dinh, N.C.; Hiep, P.T. Optimizing duration of energy harvesting for downlink NOMA full-duplex over Nakagami-m fading channel. *AEU Int. J. Electron. Commun.* **2018**, *95*, 199–206. https://doi.org/10.1016/j.aeue.2018.08.020.
31. Hieu, T.C.; Le Cuong, N.; Hoang, T.M.; Quan, D.T.; Hiep, P.T. On Outage Probability and Ergodic Rate of Downlink Multi-User Relay Systems with Combination of NOMA, SWIPT, and Beamforming. *Sensors* **2020**, *20*, 4737. https://doi.org/10.3390/s20174737.
32. Vinh, H.D.; Van Son, V.; Hoang, T.M.; Hiep, P.T. Proposal of Combination of NOMA and Beamforming Methods for Downlink Multi-users systems. In Proceedings of the 2019 3rd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), Hanoi, Vietnam, 21–22 March 2019; pp. 20–24. https://doi.org/10.1109/SIGTELCOM.2019.8696143.
33. Chinnadurai, S.; Selvaprabhu, P.; Jeong, Y.; Sarker, A.L.; Hai, H.; Duan, W.; Lee, M.H. User clustering and robust beamforming design in multicell MIMO-NOMA system for 5G communications. *AEU Int. J. Electron. Commun.* **2017**, *78*, 181–191. https://doi.org/10.1016/j.aeue.2017.05.021.
34. Brueck, S.; Zhao, L.; Giese, J.; Amin, M.A. Centralized scheduling for joint transmission coordinated multi-point in LTE-Advanced. In Proceedings of the 2010 International ITG Workshop on Smart Antennas (WSA), Bremen, Germany, 23–24 February 2010; pp. 177–184. https://doi.org/10.1109/WSA.2010.5456455.
35. Nigam, G.; Minero, P.; Haenggi, M. Coordinated Multipoint Joint Transmission in Heterogeneous Networks. *IEEE Trans. Commun.* **2014**, *62*, 4134–4146. https://doi.org/10.1109/TCOMM.2014.2363660.
36. Nezafati, M.B.; Taki, M.; Svensson, T. MSE minimized joint transmission in coordinated multipoint systems with sparse feedback and constrained backhaul requirements. *EURASIP J. Wirel. Commun. Netw.* **2021**, *2021*, 1–32, https://doi.org/10.1186/s13638-021-01979-3.
37. Merlhe, C.; Gueguen, C.; Lagrange, X. Hybrid Joint-Transmission Multi-Point Coordination for Inter-Cell Interference Management. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; pp. 1–7. https://doi.org/10.1109/VTC2021-Spring51267.2021.9448660.
38. Hendraningrat, D.K.; Narottama, B.; Shin, S.Y. Non-orthogonal multiple access in downlink coordinated multipoint transmissions. *Phys. Commun.* **2020**, *39*, 101017, https://doi.org/10.1016/j.phycom.2020.101017.