This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

# Overview of Precoding Techniques for Massive MIMO

**MAHMOUD A. ALBREEM[1], (Senior Member, *IEEE*), ALAA ALHABBASH[2], (Member, *IEEE*), AMMAR M. ABU-HUDROUSS [3], (Senior Member, *IEEE*), and SALAMA IKKI[4], (Senior Member, *IEEE*)**

[1]Department of Electronics and Communications Engineering, A'Sharqiyah University, Ibra 400, Oman, (e-mail: mahmoud.albreem@asu.edu)
[2]Palestinian ICT Research Agency, Gaza, Palestine, and with A'Sharqiyah University; (e-mail: alaa.alhabbash@asu.edu.om)
[3]Electrical and Smart-System Engineering Department, the School of Engineering, Islamic University of Gaza, Gaza, Palestine, (e-mail: ahdrouss@iugaza.edu.ps)
[4]the Department of Electrical Engineering, Lakehead University, Thunder Bay, ON, P7B 5E1, Canada, (e-mail: sikki@lakeheadu.ca)

Corresponding author: Mahmoud A. Albreem (e-mail: mahmoud.albreem@asu.edu).

This research has been financially supported in part by the Research Council (TRC) of the Sultanate of Oman (agreement No. TRC/BFP/ASU/01/2018) and in part by the Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

**ABSTRACT** Massive multiple-input multiple-output (MIMO) is playing a crucial role in the fifth generation (5G) and beyond 5G (B5G) communication systems. Unfortunately, the complexity of massive MIMO systems is tremendously increased when a large number of antennas and radio frequency chains (RF) are utilized. Therefore, a plethora of research efforts has been conducted to find the optimal precoding algorithm with lowest complexity. The main aim of this paper is to provide insights on such precoding algorithms to a generalist of wireless communications. The added value of this paper is that the classification of massive MIMO precoding algorithms is provided with easily distinguishable classes of precoding solutions. This paper covers linear precoding algorithms starting with precoders based on approximate matrix inversion methods such as the truncated polynomial expansion (TPE), the Neumann series approximation (NSA), the Newton iteration (NI), and the Chebyshev iteration (CI) algorithms. The paper also presents the fixed-point iteration-based linear precoding algorithms such as the Gauss-Seidel (GS) algorithm, the successive over relaxation (SOR) algorithm, the conjugate gradient (CG) algorithm, and the Jacobi iteration (JI) algorithm. In addition, the paper reviews the direct matrix decomposition based linear precoding algorithms such as the QR decomposition and Cholesky decomposition (CD). The non-linear precoders are also presented which include the dirty-paper coding (DPC), Tomlinson-Harashima (TH), vector perturbation (VP), and lattice reduction aided (LR) algorithms. Due to the necessity to deal with a high consuming power by the base station (BS) with a large number of antennas in massive MIMO systems, a special subsection is included to describe the characteristics of the peak-to-average power ratio precoding (PAPR) algorithms such as the constant envelope (CE) algorithm, approximate message passing (AMP), and quantized precoding (QP) algorithms. This paper also reviews the machine learning role in precoding techniques. Although many precoding techniques are essentially proposed for a small-scale MIMO, they have been exploited in massive MIMO networks. Therefore, this paper presents the application of small-scale MIMO precoding techniques for massive MIMO. This paper demonstrates the precoding schemes in promising multiple antenna technologies such as the cell-free massive MIMO (CF-M-MIMO), beamspace massive MIMO, and intelligent reflecting surfaces (IRSs). In-depth discussion on the pros and cons, performance-complexity profile, and implementation solidity is provided. This paper also provides a discussion on the channel estimation and energy efficiency. This paper also presents potential future directions in massive MIMO precoding algorithms.

**INDEX TERMS** 5G, massive MIMO, precoding, complexity, channel estimation, CF-M-MIMO, beamspace massive MIMO, IRS, energy efficiency

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



## I. INTRODUCTION

RECENTLY, there has been a tremendous increase in demands for faster internet access as well as instant access to multimedia services [1], [2]. For instance, employment of smart healthcare, smart cities, self-driving cars, and smart energy systems, has reached stages wherein it is portended that there will be nearly 39 billion active devices by the end of 2025 [3], [4]. Figure 1 shows a significant increment in the number of active connected devices in the last few years. The most need characteristics are high capacity, high data rates, high spectral efficiency, and high energy efficiency are needed [3], [4].

Therefore, the fifth generation (5G) and beyond 5G (B5G) communication systems have employed new technologies in the context of the multiple-input multiple-output (MIMO) schemes. Recently, remarkable research efforts have been done to develop the conventional multi-antenna transmission techniques to achieve high spectral efficiency and high link reliability [5], [6]. A well-known combination and efficacious technologies have been introduced in 5G such as the ultra-dense networks (UDNs), the machine-to-machine (M2M) communication, the centimeter wave (cmWave) or millimeter wave (mmWave), the spectrum sharing (SS), the internet of things (IoT), and the massive MIMO smart cities [7]–[10].

The mmWave technology corresponds to 30-300 GHz frequency band and has its own propagation characteristics [11], [12]. It has a larger bandwidth compared to the conventional sub-6 GHz massive MIMO systems, but with an extra pathloss [11]. In the literature of massive MIMO systems, there are many research efforts concentrating on mobile broadband type-high rate issues with large data packets. The other application of interest is the massive M2M communications which have become the prevailing communication model of IoT [13].

The massive M2M communications have been deployed swiftly in the past few years and have a large number of connected machines that are only sporadically active [13]–[15]. Furthermore, several detection schemes are utilized in the receivers of MIMO systems, where various antennas transmit multiple interfering signals, to separate the data symbols which are corrupted by noise and interference [16].

In massive MIMO, the base station (BS) serves a large number of single or few antenna terminals in the same band of frequency [1], [17]. The main characteristic of the conventional sub-6 GHz massive MIMO system is that the number of antennas in the BS is distinctly larger than the number of antennas for all user terminals within each cell [16], [18]. In addition, each user's equipment (UE) in the massive MIMO systems has its own processing unit to detect the data.

Considering the large number of antennas in the BS, the small processing ability in each UE leads to an intractable detection process and needs a very large processing time. Here, the amazing mission of the precoding technology is appeared to transform the detection mechanism from the receiver side

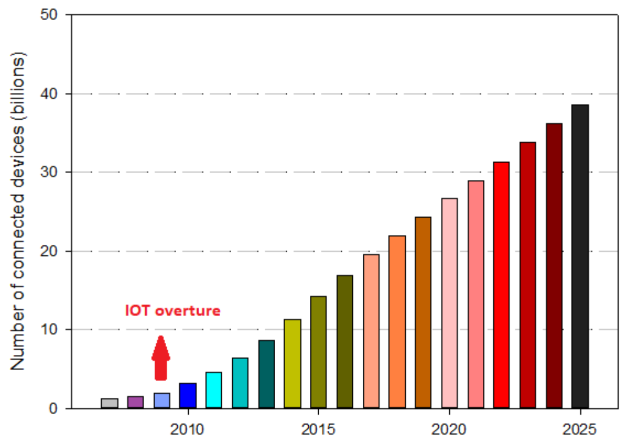


FIGURE 1: Global growth in the total number of active connected devices in the wireless services.

into the transmitter side where a large processing ability can be offered [19], [20]. Although the BS has strong and high processing ability, the demand to find lower complexity precoding algorithms is still required.

This survey focuses on the massive MIMO notion and various types of precoding technologies for systems operating below 6 GHz carrier frequency. It covers the algorithms of linear precoding, non-linear precoding, the peak-to-average power ratio precoding (PAPR), and machine learning in precoding algorithms. In addition, it will comprehensively describe the complexity and implementation issues of each precoding algorithm.

### A. RELEVANT PRIOR ART

In last few years, several research papers were published to address the issues of massive MIMO systems [16], [21]–[30]. In [21], a comprehensive survey of the linear precoding algorithms for massive MIMO for various cell scenarios has been introduced. It also addressed some of the designing issues and practical implementations of precoding algorithms. But it did not include the issues of the non-linear precoding, the PAPR precoding, and the machine learning role in precoding algorithms.

In [22], the effect of the pilot contamination and impairments of hardware in massive MIMO systems are discussed. Furthermore, it reviewed the potential reasons for pilot contamination, e.g. non-reciprocal transceivers and hardware impairments. It also classified the pilot contamination according to various mitigation techniques as a pilot-based tactic and a subspace-based tactic. In [23], a comprehensive survey of linear precoding techniques for massive MIMO systems under a single-cell (SC) scenario is provided. The performance of various linear precoding techniques is compared and analyzed in terms of sum-rate, and spectral efficiency.

In [24], a survey of the mmWave massive MIMO system challenges and benefits was introduced. As it addressed the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



boosting in user throughput, spectral efficiency, and energy efficiency. It also considers the effects of the modulation scheme, the signal waveform, the multiple access technique, the user scheduling algorithm, the fronthaul design, the antenna array architecture, and the precoding algorithm. Nevertheless, it is concluded that the performance of mmWave massive MIMO system in practical scenarios and real-life applications still under intense research until this moment.

In [25], the propagation channels of massive MIMO systems are extensively investigated and main differences from the traditional MIMO systems are discussed. In addition, it reviewed the characteristics, measurements, and channel models. Few futurity channels models directions for massive MIMO systems are also presented and analyzed. It is concluded that the propagation channels will still an open research direction in the advent few years. In [26], a comprehensive review of the various embodiments of digital and analog beamforming designs by employing average channel state information (CSI) has been presented. The hybrid beamforming design has feasible limits of the number of radio frequency (RF) chains. In addition, it is shown that the hybrid beamforming designs are favorable for diminishing the cost of hardware and the overhead of training. Nevertheless, the hybrid beamforming design is considered as a trade-off between performance and complexity in the various applications designs and channel characteristics.

A survey in [16] introduced a detailed clarification of the fundamentals of massive MIMO detection, and recited the past twelve-year history of massive MIMO detection. The authors offered an extensive review and milestones in the development of optimal, near-optimal, linear, approximate inversion based massive MIMO detection algorithms. Furthermore, the authors have briefly explored some of the non-linear small-scaled MIMO detectors and their applicability in the massive MIMO systems. In addition, recent improvements in detection process with incorporated machine learning

In [27], a comprehensive review of various prominent mmWave massive MIMO systems, like multiple access technologies, hybrid precoding and combining, cell-free massive MIMO (CF-M-MIMO), non-orthogonal multiple access (NOMA), and simultaneous wireless information and power transfer (SWIPT) technologies is presented. In [28], a brief overview of massive MIMO localization has been provided. With respect of performing localization of massive MIMO systems, user's localization methods and refined channel estimation routines have been advanced. Several spatial signatures of users can be employed in massive MIMO systems to meet the demands of 5G technology and to specify the locations of the users.

In [29], a survey of the detection techniques in uplink (UL) massive MIMO systems is introduced. The authors concluded that the research efforts on the detection techniques for UL massive MIMO systems are still in an early phase. There are lots of considerable and imperious issues that need to be resolved in the future, e.g. using the deep learning algorithms for detection techniques, and finding suitable detection techniques to work in Hetnets wireless communications.

The authors in [30] over-viewed the artificial intelligence (AI) implementation, and addressed several issues in the massive MIMO systems. AI can be exploited to improve the user experience and effectively utilize the radio resources. The BS in massive MIMO systems needs to produce and sense huge data for communication. Thus, the demand for a new technology, that can learn and predict the system requirements, has been increased. Hence, more accuracy with lower complexity can be offered by employing an effective AI method to massive MIMO systems.

While the above research papers discuss a number of key issues of massive MIMO systems, none of them extensively review the precoding techniques. However, most of these techniques focus only on the linear precoding detection algorithms.

### B. CONTRIBUTION AND OUTLINE

In this paper, an extensive survey on precoding algorithms related to the massive MIMO systems is introduced. Our particular focus is on performance and complexity trade-off as well as the practical implementation of general precoding algorithms. Although the survey in [21] is extensive, the primary focus of the paper was only on linear precoding in massive MIMO systems for different cell scenarios. For instance, the linear matrix inversion approximation precoder and fixed-point iteration-based linear precoding algorithms are not covered in [21]. In addition, the non-linear precoding algorithms, the machine learning based precoders, and the PAPR precoding algorithms are also not reviewed in [21].

To our best knowledge, this is the first survey to review the most types of precoding algorithms considering only massive MIMO systems. In the literature of massive MIMO systems, there is a plethora of precoding algorithms. The target of this survey is to offer insights on such algorithms to a generalist of MIMO communication systems. This paper is also demonstrating the use of massive MIMO for B5G where promising technologies are flashing such as the CF-M-MIMO, beamspace massive MIMO, and intelligent reflecting surfaces (IRSs). Table 1 compares this paper with other prior relevant articles.

The major contributions of this paper are summarized as:

- This paper reviews the massive MIMO precoding algorithms and introduces their performance-complexity profile so that the reader can find the differences between various precoding algorithms with a wider range of potential solutions. It starts off with a dive into the literature of precoders for massive MIMO systems. Then, it introduces the benefits and challenges of precoded massive MIMO systems. It then discusses the basic linear precoders in massive MIMO systems.
- This paper surveys the corresponding linear precoding solutions for massive MIMO systems starting with pre-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



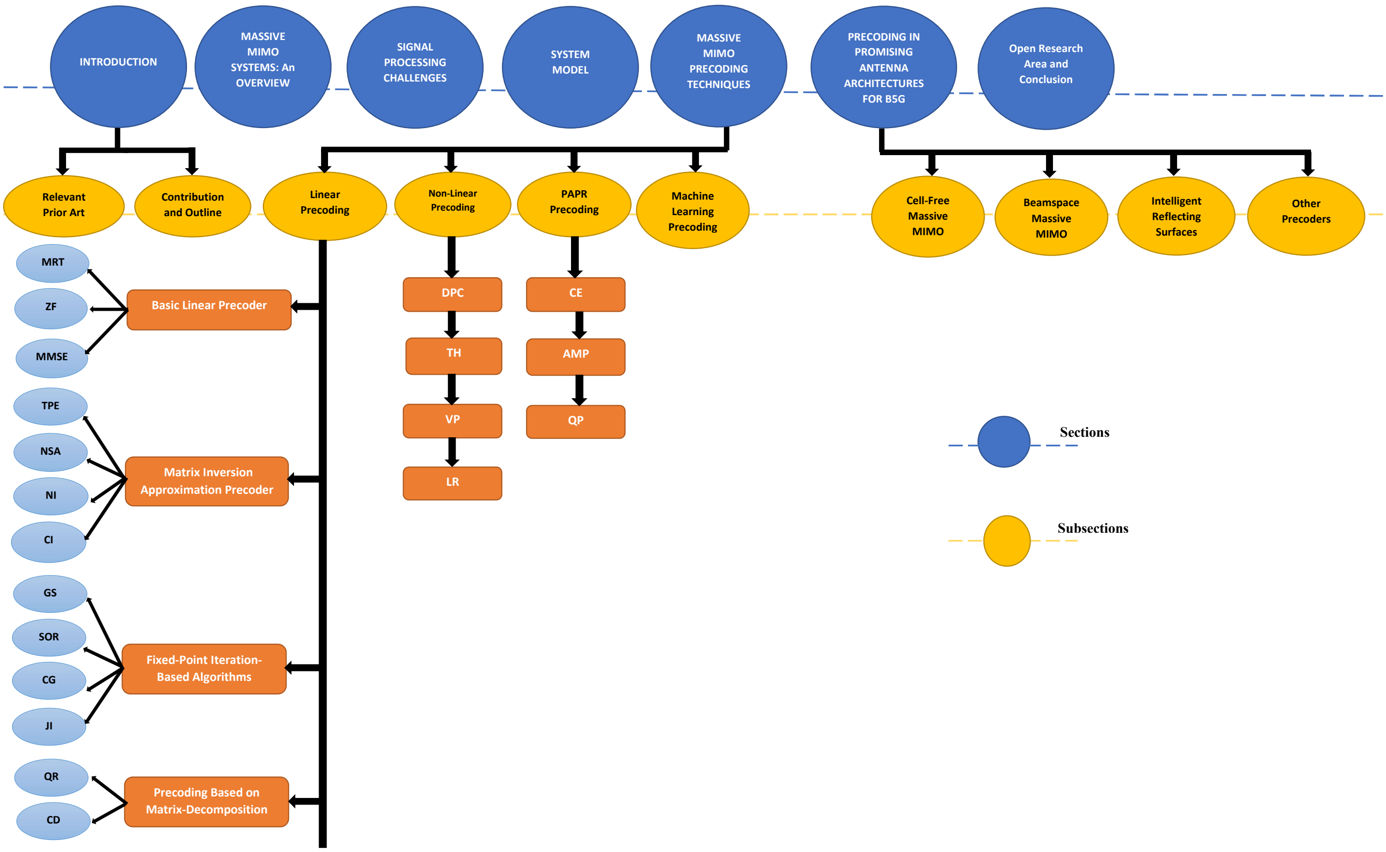


FIGURE 2: Outline of the paper.

coders with approximate matrix inversion methods such as the truncated polynomial expansion (TPE) algorithm, the Neumann series approximation (NSA) algorithm, the Newton iteration (NI) algorithm, and the Chebyshev iteration (CI) algorithm.

- This paper surveys the fixed-point iteration-based linear precoding algorithms for massive MIMO systems such as the Gauss-Seidel (GS) algorithm, the successive over relaxation (SOR) algorithm, the conjugate gradient (CG) algorithm, and the Jacobi iteration (JI) algorithm.
- This paper reviews the direct algorithms-matrix decomposition based linear precoding algorithms such as the QR decomposition algorithm and the Cholesky decomposition (CD) algorithm.
- This paper comprehensively surveys the non-linear precoders such as the dirty-paper coding (DPC) algorithm, the Tomlinson-Harashima (TH) algorithm, the vector perturbation (VP) algorithm, and the lattice reduction aided (LR) algorithm. Thus, due to the necessity to deal with high consuming power by the BS with a large number of antennas in massive MIMO systems, we have dedicated a special subsection to describe the characteristics of the PAPR algorithms as the constant envelope (CE) algorithm, the approximate message passing (AMP) algorithm, and the quantized precoding (QP) algorithms.
- This paper reviews the precoding in promising multiple antenna technologies: the CF-M-MIMO, beamspace massive MIMO, and the IRSs.
- This paper reviews the potential of machine learning role in precoding algorithms.
- Finally, this work is also discussing the advantages and disadvantages of each precoder based on the performance-complexity profile as well as the implementation solidity.

Section II presents the benefits and challenges of massive MIMO systems. Section IV describes the massive MIMO system model. Section V illustrates the precoding algorithms for massive MIMO systems. Finally, section VII concludes the paper and introduces the open research area in the precoding process for massive MIMO systems. For convenient reading, the outline of the paper is depicted in Fig. 2.

## II. MASSIVE MIMO SYSTEMS: AN OVERVIEW

Massive MIMO systems are an expansion of the MIMO technology which has been introduced since the third gener-

TABLE 1: Prior Relevant Articles

| **Ref.** | Linear Precoding | Matrix Inversion Approximation | Fixed-Point Iteration | Non-linear Pre-coding | PAPR Pre-coding | Machine Learning | Channel Estimation | Hardware Efficiency | CF-M-MIMO | Beamspace | IRS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [21] | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| [22] | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ |
| [23] | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| [16] | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ |
| This Work | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

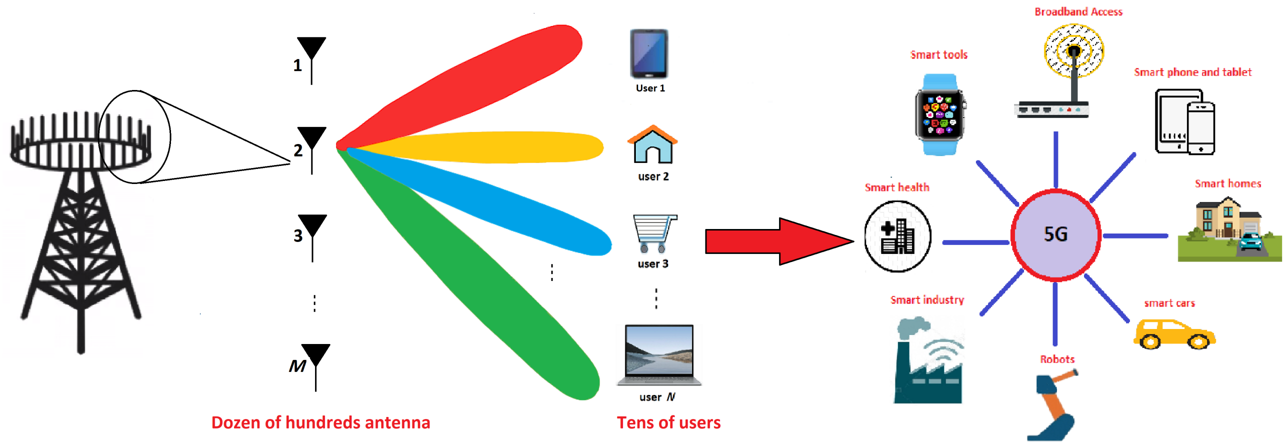


FIGURE 3: Massive MIMO system beamforming and its services.

ation (3G) communication systems. Massive MIMO involves hundreds to thousands of antennas occupied at the BS to serve simultaneously many user terminals [31]. Figure 3 shows a massive MIMO scenario to create directed beams in specific small area to serve one or few users [32] and can be employed to obtain the following benefits:

- ***Spectral efficiency***: Massive MIMO systems achieve a high spectral efficiency by exploiting a large antenna array to originate more multiplexing gain [33]. Consequently, each user equipment has an individual down-beam which leads to offering spectral efficiencies ten times higher than that in the conventional MIMO technology [34].
- ***Energy efficiency***: In massive MIMO systems, the gain of transmitted signals is increased to the position of candidate users by pointing the beam of the antenna array into a small region. Consequently, the massive MIMO systems radiate less power and are more energy-efficient systems [1]. Moreover, the transmit power is significantly reduced when the number of transmit antennas is increased [35]. By dint of the huge number of antennas in massive MIMO systems, a BS can make several beams at the same time and directly pointing them to a particular user or more [32]. Then, the resources can be used repeatedly in the same specific area. Thus, the throughput could be increased without increasing the transmit power by increasing the number of transmit antennas [36]. Massive MIMO systems have the ability to reduce the transmitted power 1000 times below conventional MIMO and to maximize the data rates at the same time [37].
- ***User tracking***: As massive MIMO systems point narrow signal beams to the users; user tracking become more reliable and accurate [32].
- ***Cost efficiency***: Massive MIMO systems are constructed with cheap ultra lower power amplifiers, which abstract the need for expensive bulky electronic equipment. Furthermore, it eliminates the need for bulky coaxial cables which connecting the BS components [1]. These are a glance of low-cost features in massive MIMO systems which reduce the system implementation cost [38].
- ***Reliability***: A large number of antennas in massive MIMO systems advances high diversity gain, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



TABLE 2: Summary of Massive MIMO System, its Features, Advantages, and Challenges.

| Massive MIMO Systems | |
|---|---|
| Features | Dozen of hundreds antennas per BS.<br>BS serves multiple received terminals.<br>Low transmitted power per antenna.<br>Number of antennas is much larger than number of received terminals.<br>Small-size antennas.<br>High directive gains. |
| Advantages | High capacity.<br>High throughput.<br>High resistance against noise.<br>High scalability.<br>High energy efficiency.<br>High spectral efficiency.<br>High reliability.<br>High multiplexing gain.<br>High security.<br>Low fading affects.<br>Low latency.<br>Low implementation cost.<br>Low bit error rate (BER)<br>Low signal processing.<br>User tracking. |
| Challenges | Pilot contamination.<br>Channel estimation.<br>Signal detection.<br>Signal precoding.<br>User scheduling.<br>Hardware impairments. |

increases the link reliability and elasticity against fading [39].

- ***Robustness***: Massive MIMO systems are more robust against internal jamming and unintended interference. Also, they have the ability to avoid one or few antenna failures as a result of large number of antennas [40].
- ***Enhanced security***: In massive MIMO systems, the large number of antennas can be used to cancel the signals from wilful jammers [41]. Moreover, massive MIMO systems are also inherently strong against hackers and passive eavesdropping attacks due to the orthogonal channels of receivers and narrow beams [42].
- ***Simple signal processing***: For Massive MIMO systems with large antennas array, the BS immensely surpasses the number of received terminals. This leads to make the column vectors of the propagation matrix asymptotically orthogonal under the most favorable propagation assumption and that eliminates the interference effects, fast fading, uncorrelated noise, and thermal noise, and hence simplifies the signal processing [43].

## III. SIGNAL PROCESSING CHALLENGES

In massive MIMO, a large number of antenna elements leads to a utilization of random matrix theory to confirm the approximation accuracy of large dimension settings [44]. However, the mutual orthogonality among the vector-valued channels is one of major properties and is called as "favorable propagation". However, the channel condition number was used to indicate whether the propagation is favorable or not. In other words, it is a proxy for how favorable the channel is [45].

The work in [46] shows that the condition number could lead to a misleading conclusion when the norms of the channel vectors are not equal. The favorable propagation has a great impact to maximize the information rate under a power constraint [46]. In [47], the favorable propagation condition was analysed in a generic channel model for LOS and NLOS propagation where a general steering matrix-propagation matrix model was employed. The energy efficiency is also highly affected by the statistical CSI.

In [48], an iterative power allocation algorithm has been proposed based on sequential optimization, fractional optimization, and random matrix theory. However, many chal-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



lenges are kick-in when a large number of arrays are utilized such as:

- ***Pilot contamination***: Channel estimation is crucial in massive MIMO systems [49]. To estimate the UL channel, the received terminal sends orthogonal pilot signals to the BS. Furthermore, thanks to the channel reciprocity feature in massive MIMO systems, the BS estimates the downlink (DL) channel towards the received terminal [50]. When the pilot signals in the cell and neighboring cells are orthogonal, the BS acquires the delicate estimation of the channel. However, in specific resources, the number of orthogonal pilot signals is limited [49].
  As a result, this imposes the reuse concept of the orthogonal pilots. Subsequently, the inter-cell interference appears due to the use of the same set of orthogonal pilots, and the BS will receive a linear combination of channel responses from the cell and its neighboring cells. This is known as a pilot contamination [51]. The pilot contamination reduces the achievable throughput. The same phenomenon occurs in the DL channel, the BS directs the beamforming signal towards the received terminals in its cell in addition to undesired received terminals in the neighboring cells [52], [53].
- ***Channel estimation***: Channel state presents the channel response realization. However, CSI refers to the knowledge of the channel states at the BSs. It is assumed that the statistical CSI of random variables are available anywhere in the network. In addition, as the channels change, instantaneous CSI about current realizations need to be acquired [7]. In other words, CSI depicts the signal propagation, and has information about the communication link between the transmitter and the receiver. CSI elucidates the combined effects of fading, scattering, power decay, and so forth. The performance of massive MIMO systems increases rapidly with the minimum number of transmitting or receiving antennas when the CSI is idealistic [54].
  Pilot signaling is a method to acquire the CSI where the antenna transmits a predefined pilot signal. However, the transmission can be simultaneously received by any antenna and compared with the known pilot signal to estimate the channel from the transmitting antenna. In case of two transmitting antennas, two orthogonal pilot signals are normally needed. In other words, as the number of transmit antennas increases, the number of orthogonal pilot signals will also increases. The receive antennas can "listen" to the pilots simultaneously where individual channels were estimated to the transmitters [7]. In other words, within the UL transmission, the channel is estimated by the BS with thanks to orthogonal pilot signals which are sent by the received terminals. Within the DL transmission, the BS transmits pilot signals towards the received terminals, and the received terminals recognize the estimated channel information for the DL transmission. Pilot signaling could cause an overhead because every pilot signal could have been a signal where payload data is carried. Therefore, the overhead based on pilot signaling should be considered and minimized.
  The CSI, in the systems using frequency division duplexing (FDD), requires to be estimated during both the UL and DL transmissions. FDD systems use different frequency bands and consider different CSI corresponding to each band [49]. In order to obtain the required CSI, the BS sends training symbols to the users. Each user estimates the channel coefficients and transmits the estimated channel vector back to the BS [55]. Duplex distance is usually utilized to separate the UL and DL channels which may lead to a performance consideration. The required time to transmit the DL pilot signals is proportional to the number of BS antennas [49]. This is a real challenge since the complexity scales with the number of antennas. In other words, with a large number of BS antennas in massive MIMO systems, the use of FDD system becomes very hard and infeasible to be implemented in practical applications [49]. Fortunately, using the time division duplexing (TDD) offers a solution for the problem within the DL transmission by exploiting the channel reciprocity feature. The BS can estimate the DL channel from channel information within UL [49].
  In TDD, the UL and DL operate in different time slots but in the same frequency band. At the BS, the design of DL precoder depends on the channel estimate that was obtained in the previous UL slot. In TDD, channel reciprocity is usually assumed. In high speed scenarios, the system has to be designed carefully to avoid a degradation in the precoder performance [56]. A performance comparison between TDD and FDD systems in [57] shows that the TDD is more suitable in massive MIMO. The FDD suffers from a considerable degradation in performance in many channels. FDD can achieve a satisfactory performance in line of sight (LOS) scenario and high Ricean factors.
  However, due to a limited number of orthogonal pilots in massive MIMO systems in a specific cell and neighboring cells, the pilot contamination problem becomes a considerable challenge for a channel estimation [49], [58].
  The Covariance matrix plays a crucial role for resource allocation and pilot contamination. In a plethora of massive MIMO papers, it is commonly assumed to be perfectly known which could lead to misleading conclusions because the matrix dimensions vary with the number of antennas and other statistics based on mobility. In practical scenarios, channels are spatially correlated where channel elements are correlated. In order to apply the MMSE channel estimator, covariance matrix is required [59]. In [59]–[61], the large-dimensional covariance matrix was estimated using a sample covariance

matrix. Principles of robust precoding under imperfect CSI was comprehensively illustrated in [55]. A model for the uncertainty of the CSI can be developed based on a precoder to minimize the transmission power subject to SINR constraints for all channel sets [55], [62]. In [63], a design of robust hybrid analog/digital beamforming systems under imperfect CSI was proposed where a norm-bounded channel error and the MMSE were utilized to capture the imperfect CSI conditions.

In [64], hardware cost and power consumption of a hybrid analog/digital beamforming systems have been lowered by a reduction of training sequence dimension where a limited number of RF chains were utilized. Toeplitz distribution theorem with specific antenna configurations were applied to select the training sequence parameter. In [65], a hybrid precoding scheme based on equal gain transmission and the ZF has been proposed to reduce the hardware cost and processing complexity of massive MIMO systems. Detailed explanations of hybrid precoding with hardware architectures and methods of deployment with the impact of CSI have been comprehensively discussed in [66]. It was shown that the hybrid beamforming has a critical impact in minimizing the hardware cost since a small number RF chains at the transceivers was utilized. Hybrid beamforming is an energy efficient scheme because it reduces the power consumption for each mobile device without a performance degradation.

The channel capacity is highly affected by the instantaneous CSI, perfect CSI, and imperfect CSI. In [67], the relationship between the channel capacity and the CSI in different scenarios was determined when the CSI is unknown at the transmitter, the CSI is perfectly known at the transmitter, and the CSI is imperfect at the transmitter. The impact of CSI on energy efficiency was demonstrated in [67]–[69]. In [68], it was shown that the CSI has a crucial impact on the energy efficiency when transmitting over long link distances. However, its impact is not critical in the short link distance scenario.

- ***Hardware Efficiency***: The high computational complexity of massive MIMO receivers limits the gain that can be obtained in real applications. Therefore, a design of energy-efficient massive MIMO systems has attracted the attention of the research community in both academia and industry [70]–[81]. In [70], a resource allocation for energy-efficient in an orthogonal frequency division multiple access (OFDMA) with a massive MIMO system was considered. In this paper, circuit power consumption, imperfect CSI, and different QoS were taken into consideration. The resource allocation policies were updated based on the realization of path loss and shadowing.

  Numerical results show that the large number of antennas is always useful for the communication system capacity. However, it could not be a cost effective solution for enhancing the performance. In [78], a trade-off between energy efficiency and spectrum efficiency was considered by utilizing the channel states, the transmit power and its allocation. In [81], the selection of optimal subcarrier rates and power allocation to obtain an optimal energy efficiency was comprehensively demonstrated using an iterative approach. The efficiency of power amplifiers was modeled as a function of the number of subcarriers utilized for transmission. The water-filling and link adaption based on CSI could be utilized to maximize the sum rate for a transmission power in a frequency-selective channel. However, traditional water-filling is not the best approach to achieve efficient and reliable subcarrier power allocation. In [79], the subcarrier availability was taken into consideration to maximize the transmission rate and energy efficiency. The proposed approach has outperformed the water-filling scheme. In [80], a new energy-efficient approach was proposed to reduce the computational complexity of [78]. Energy optimal link adaption and resource scheduling techniques were derived in closed forms where time average bit-per-Joule metrics were taken into consideration.
- **Data detection**: The large number of antennas in massive MIMO systems causes a high computational complexity and reduces the achievable throughput within the signal detection. Moreover, all signals which are transmitted from received terminals superimpose at the BS and cause interference, which reduces the throughput and spectral efficiency [82]. The maximum likelihood (ML) detector achieves the optimum performance and has a strong ability to minimize the probability of error. Due to a large number of antennas in massive MIMO systems, the ML detector has an illicit complexity [21]. Comprehensive research has been done to find the optimal data detection method for massive MIMO systems that can achieve a preferable throughput performance with low computational complexity.

  The classical non-linear detectors such as the successive interference cancellation (SIC) [83] and the sphere decoder (SD) [84]–[86] yield acceptable performance. Though, a large number of antennas increases the computational complexity which makes the conventional non-linear detectors impossible for massive MIMO systems. A comprehensive presentation of detection techniques for massive MIMO is presented in [16]. There are many linear detection methods that are considered for the UL detection in massive MIMO systems, such as the zero-forcing (ZF) and minimum-mean-square-error (MMSE) [21], [42], [87]. In ZF detectors, the inter-antenna interference is moderated, but the additive noise is increased for tacky conditioned channel matrices. The MMSE detector considers the noise power within the detection process accordingly. The MMSE detector performance exceeds the performance of the ZF detector [88].

  Though the ZF and the MMSE based detectors give a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



good throughput performance, they comprise a matrix inversion during the processing which makes detection methods computationally inefficient for a large number of antennas [89], [90]. A plethora of research methods has been proposed to design a low complexity detection method for massive MIMO systems [91], [92]. Iterative methods to avoid the exact matrix inversion are considered in the literature to find sub-optimal for the UL detection in massive MIMO systems such as the CG [92], AMP, GS [93], SOR, and least-square regression selection methods [94].

- ***Precoding***: Precoding is a conception of beamforming where the multi-antenna systems support the multi-stream transmission [32]. Precoding performs an imperious technique in massive MIMO systems where it plays a crucial role to reduce the effects of interference and path-loss, and increases the throughput [21], [95]. In massive MIMO systems, the BS can estimate the CSI thanks to the UL pilot signals which are sent from the received terminals. The received CSI at the BS is imperfect and uncontrollable as a result of several environmental obstacles on the wireless channel [21]. Though the BS does not have a perfect CSI, nevertheless the DL performance of the BS broadly depends on the estimated CSI [32]. The massive MIMO's BS exploits the precoding techniques and the estimated CSI to mitigate the interference and increase spectral efficiency [32]. The precoding techniques give a tremendous benefits to massive MIMO systems. Unfortunately, these benefits are coming with a high computational complexity which is directly proportional to the number of antennas. Therefore, a low complexity precoder is imperative to exploit in massive MIMO systems [32], [95]. This aspect is covered in detail in this review.

The features, advantages, and challenges of massive MIMO systems are summarized in Table 2.

## IV. SYSTEM MODEL

In this section, an overview of the DL system model for massive MIMO is presented. It is assumed that a single BS with $M$-transmitted antennas is serving $N$-single antenna received terminals, where $N \leq M$. A frequency-flat channel, which indicates coefficients across $N$-received terminals and $M$-transmitted antennas, is considered. In the TDD mode, the DL transmission has the same channel matrix $\mathbf{H}$ as the UL transmission, within the channel coherence time, due to the channel reciprocity [49]. The channel matrix $\mathbf{H} \in \mathbb{C}^{M\times N}$ can be represented as

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13}\cdots & h_{1N} \\ h_{21} & h_{22} & h_{23}\cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{N2} & h_{N3}\cdots & h_{MN} \end{bmatrix}, \quad \mathbf{H} \in \mathbb{C}^{M\times N}. \tag{1}$$

Elements of $\mathbf{H}$ are drawn from complex Gaussian distributions $\mathcal{CN}(0, 1)$.

For the DL transmission, the upcoming data $\mathbf{a} = [a_1, a_2, \cdots, a_N]^T$, which are taken from a $M$-ary constellation, passed into a precoding stage at the BS. The $M$ BS antennas form their precoded vector by converting $\mathbf{a}$ into $M \times 1$ vector as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M]^T$, and then send it separately to each terminal of $N$-received terminals through the channel. With the assumption of perfect CSI and synchronization at the BS, the precoder can be exploited to point the transmitted signal to its specified received terminal. Figure 4 shows the system model of massive MIMO systems with $M$-transmitted antennas and $N$-received terminals which also indicates the position of the precoding block.

The received vector at received terminals is $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]^T$ which is affected by channel effects and noise. The $N \times 1$ vector of the received signal at the BS can be represented as

$$\mathbf{y} = \mathbf{H^T x} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^{N\times 1}, \tag{2}$$

where $\mathbf{n}$ is $N \times 1$ additive white Gaussian noise (AWGN) vector whose elements are drawn from complex Gaussian distributions $\mathcal{CN}(0, \sigma_n^2)$.

## V. MASSIVE MIMO PRECODING TECHNIQUES

One of the main concepts in massive MIMO systems is a precoding technology that transforms the complexity system from the side of received terminals to the side of BS by using a strong signal processing technology at the transmitter side [19]. Usually, in a real wireless propagation environment, it is difficult to obtain a reliable CSI where the performance of DL transmission largely depends on CSI. The precoding technology can be employed to deal with imperfect CSI.

Many research papers have shown that massive MIMO precoding technology acts as a critical role to out from the bottleneck of breaking down the system's performance by controlling the direction of the beams and points them into a specific received terminal location [19]. Utilization of the precoding technology in massive MIMO systems leads to eliminate/cancel the effects of interference and fading, and increasing the throughput and capacity [96] when the number of antennas approaches infinity. The precoding algorithms can be mainly classified into linear, non-linear, PAPR precoding and machine learning based precoding algorithms, which are covered in detail in this section.

### A. LINEAR PRECODING

Fig. 5 depicts the generalized block diagrams of communication systems with precoding and decoding techniques. The $\mathbf{P}$ is a feedforward matrix of linear precoding, the $\mathbf{B}$ is a feedback matrix of linear precoding, the $\mathbf{K}$ is a feedforward matrix of linear decoding, and the $\mathbf{C}$ is a feedback matrix of non-linear precoding. These matrices specify the required precoding technique from a linear/non-linear or hybrid technique. For instance, when $\mathbf{B}=0$ the generalized precoding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



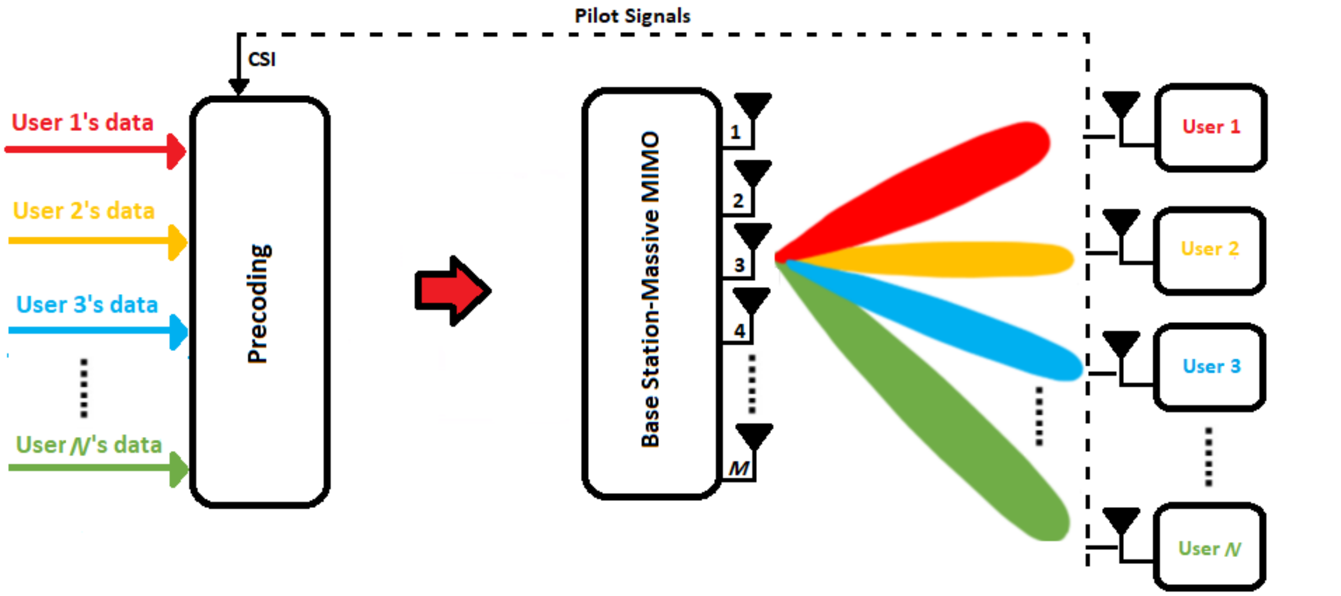


FIGURE 4: System model of massive MIMO systems with $M$- transmitted antennas and $N$-received terminals.

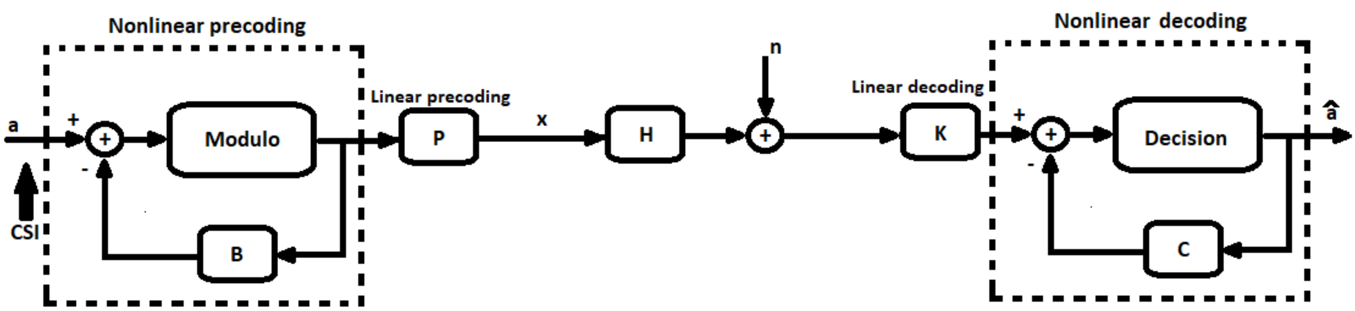


FIGURE 5: Generalized block diagram of communication systems with precoding and decoding techniques.

technique acts as the linear precoding technique [97]. The Modulo arithmetic is used to adjust the average power [98].

Thus the transmitted signal for the $N$ users in the DL transmission, where $M > N$, can be expressed as:

$$\mathbf{x} = \sqrt{\rho}\mathbf{P}\mathbf{a}, \quad \mathbf{x} \in \mathbb{C}^{M\times 1}, \tag{3}$$

where $\mathbf{P}$ is a $M \times N$ linear precoding matrix, $\mathbf{a}$ is a $N \times 1$ transmitted vector before precoding process, and $\sqrt{\rho}$ is the transmitted average power. The precoding matrix $\mathbf{P}$ is related to $\mathbf{H}$. In the TDD mode, the DL channel is the transpose of the $\mathbf{H}$ [49], and the $N \times 1$ vector at $N$-received terminals becomes

$$\begin{aligned} \mathbf{y} &= \mathbf{H}^{\mathbf{T}}\mathbf{x} + \mathbf{n}, \\ &= \sqrt{\rho}\; \mathbf{H}^{\mathbf{T}}\mathbf{P}\mathbf{a} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^{N\times 1}. \end{aligned} \tag{4}$$

In general, the $\mathbf{P}$ matrix of basic precoding techniques contains a matrix inversion operation which leads to high computational complexity, especially, if $N$ is not greater enough than $M$ [98]. According to the manner of dealing with the matrix inversion process, the linear precoding technique can be classified into basic linear precoding, linear precoder based on the matrix inversion approximation, linear precoder based on fixed-point iterations, and linear precoder based on matrix-decomposition.

1) Basic Linear Precoding Algorithms:

The basic linear precoder mainly depends on multiplying the transmitted signal $\mathbf{a}$ with the precoding matrix $\mathbf{P}$. The basic linear precoder has $\mathcal{O}(N^3)$ computational complexity which is comparable to the exact matrix inversion complexity [16], [99], [100].

a) Maximum Ratio Transmission (MRT) Algorithm: The MRT aims to maximize the gain of signal into a specific receive terminal. It is the counterpart of the matched filtering (MF) and conjugate beamforming (CB) [21]. The MRT precoding matrix formula is

$$\mathbf{P}_{\text{MRT}} = \sqrt{\beta}\left(\mathbf{H}^{*}\right), \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



where $\beta$ is a scaling power factor and $\mathbf{H}^*$ is the complex conjugate of $\mathbf{H}$ matrix [21]. Thus, the received signal becomes

$$\mathbf{y}_{\text{MRT}} = \sqrt{\beta}\sqrt{\rho}\ \mathbf{H}^{\mathbf{T}}\ \mathbf{H}^*\ \mathbf{a} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^{N\times 1}. \quad (6)$$

The MRT algorithm achieves the sum capacity of a massive MIMO system when the number of $M$ is much larger than $N$, and $M$ grows to infinity ($M >> N$ and $M \to \infty$). In general, the MRT algorithm performance is close to optimal when the inter-user interference (IUI) is trivial compared to the noise (noise-limited systems). In the MRT algorithm, when the values of $M$ and $N$ are comparable, the system experiences a strong IUI. Thus, the throughput of each user becomes low which degenerates the massive MIMO concept [21]. Another amazing feature of the MRT algorithm is that each antenna in the BS can perform its signal processing locally [101]. That allows a decentralized construction for the large number of antennas and leads to a great flexible system [21], [102], [103].

b) Zero-Forcing (ZF) Algorithm: The ZF algorithm is a common algorithm of fundamental precoding techniques. It is the counterpart of the channel inversion. A ZF algorithm mitigates the interference caused of other users by pointing the signal beam into the intended user whereas nulling the other directions where other users are located [104]. This nulling is performed by multiplying the user data with the following ZF precoding matrix

$$\mathbf{P}_{\text{ZF}} = \sqrt{\beta}\mathbf{H}^*\left(\mathbf{G}^{-1}\right), \quad (7)$$

where $\mathbf{G} = \mathbf{H}^{\mathbf{T}}\mathbf{H}^*$ is a *Gram* matrix whose diagonal components indicate power imbalance throw the channel, and non-diagonal components indicate the mutual correlations among the channels. While the number of transmit antennas grows to infinity in massive MIMO systems, $\mathbf{G}$ goes to become an identity matrix and the matrix inversion computations can be simplified. The received signal of the ZF algorithm can be expressed as

$$\mathbf{y}_{\text{ZF}} = \sqrt{\beta}\sqrt{\rho}\ \mathbf{H}^{\mathbf{T}}\mathbf{H}^*\left(\mathbf{G}^{-1}\right)\ \mathbf{a} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^{N\times 1}. \quad (8)$$

The ZF algorithm performance is close to optimal when the noise is trivial compared to the IUI. The ZF algorithm is considered to be practical when neglecting the AWGN in the massive MIMO channel model, while the massive MIMO precoding algorithm becomes much simpler to implement. Unfortunately, the noise is not negligible in a real situation and utilization of the ZF algorithm in massive MIMO systems may not give an optimal solution. The ZF algorithm may achieve accurate results at high signal-to-noise ratio (SNR) [21], [104], [105].

c) Minimum Mean Square Error (MMSE) Algorithm: The MMSE algorithm exploits the benefits of the MRT and ZF algorithms and achieves a balance between them [21]. Therefore, it has an acceptable performance in moderate noise and interference systems [105]. The MMSE algorithm is the counterpart of the regularized ZF (RZF), signal-to-leakage-and-interference ratio (SLNR) [106], eigenvalue-based beamforming, and transmit Wiener filtering [104]. The MMSE algorithm is created by using the mean square error method in the signal to minimize the error filtering between the transmitted symbols from the BS and the received terminal.

The MMSE precoding matrix formula is

$$\mathbf{P}_{\text{MMSE}} = \sqrt{\beta}\ \mathbf{H}^*\left(\mathbf{G} + \mathbf{V} + \lambda\,\mathbf{I}_N\right)^{-1}, \quad (9)$$

where $\lambda$ is a positive regularizing factor which depends on the system dimensions, the noise variance, and uncertainty of channel at the transmitter. The matrix $\mathbf{V}$ is a $N \times N$ deterministic Hermitian non-negative definite matrix. When $\mathbf{V} = \mathbf{0}$, a balance occurs between increasing the channel gain toward intended received terminals (at a large value of $\lambda$ ) and eliminating the IUI (at a small value of $\lambda$ ). The MMSE algorithm performs as the ZF algorithm at $\lambda \to 0$, and as the MF algorithm at $\lambda \to \infty$ [21], [107]. The received signal of the MMSE algorithm is

$$\mathbf{y}_{\text{MMSE}} = \sqrt{\beta}\sqrt{\rho}\,\mathbf{H}^{\mathbf{T}}\mathbf{H}^*\left(\mathbf{G} + \mathbf{V} + \lambda\,\mathbf{I}_N\right)^{-1}\mathbf{a} + \mathbf{n}. \quad (10)$$

However, the computation of the ZF and MMSE precoding matrix comprises the inversion of a very large-dimension matrix, particularly for large values of $M$ and $N$ [21], [108]. Therefore, it is quite important to offer a method to diminish the complexity of the basic precoding algorithms [109].

2) Linear Precoder Based on the Matrix Inversion Approximation:

A large number of $M$ compared to $N$ leads to make $\mathbf{G}$ as a diagonal dominant, where the non-diagonal components go to zero and diagonal components become close to $M$ [82], [110]. A matrix inversion of $\mathbf{G}$ requires a high computational complexity. There is a plethora of research to approximate or avoid the matrix inversion of $\mathbf{G}$ rather than computing it [111]. In addition to the high complexity of a matrix inversion, a defy in matrix inversion is the inversion of nearly singular and ill-conditioned matrix [112]. To beat the inveterate noise boost, advanced precoders with approximate/avoid matrix inversion methods are required.

a) Truncated Polynomial Expansion (TPE) Algorithm: The TPE precoding algorithm aims to achieve a similar performance of the MMSE algorithm with low computational complexity. The TPE algorithm exploits an approximation of known precoding matrices instead of the matrix inversion in the MMSE algorithm to balance between complexity and achievable data rate via different truncation orders [21], [113]. The TPE precoding matrix formula is

$$\mathbf{P}_{\text{TPE}} = \sum_{j=0}^{J-1} p_j \mathbf{H}^*\left(\mathbf{G}\right)^j, \quad (11)$$

where $p_j$ is a scalar coefficient, and $J$ is the number of terms of the precoder polynomial. Thus, proper adjusting value of $J$ leads to a smooth transition between the traditional low-complexity MRT ($J$ = 1) and the high-complexity RZF ($J$ = min($M$, $N$)) precoding. Clearly, the flexibility of the TPE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



algorithm appeases in the ability to easily tailor the hardware complexity or changed it dynamically by increasing and reducing $J$ depending on the high or low SNR, respectively [21], [114].

b) Neumann Series Approximation (NSA) Algorithm: The NSA algorithm can tackle the high computational complexity by employing the Neumann series in the matrix inversion approximation [107]. In the NSA algorithm, the inversion process in traditional inversion algorithms such as the MMSE algorithm is replaced by a series of matrix vector multiplications (sum of powers) which has a simple flow of data and can be highly parallelized [115]. The NSA algorithm has more energy efficiency and low complexity than the other traditional inversion algorithms i.e. ZF, and MMSE algorithms [100], [116], [117]. This is because the fact that the $\mathbf{G}$ matrix becomes diagonally dominant as the value $(M/N)$ ratio increases [107]. The diagonal $\mathbf{G}$ matrix can be used as initial matrix in the Neumann series. When the $\mathbf{G}$ matrix is non-diagonally dominant, due to low value of $(M/N)$ ratio and high antenna correlation, the NSA algorithm experiences a slow convergence of the Neumann series [113].

The NSA algorithm treats with non-diagonal matrix by decomposed $\mathbf{G}$ into the sum of diagonal elements matrix $\mathbf{D}$ and non-diagonal elements matrix $\mathbf{E}$, which needs more iterations to achieve a certain performance [107], [118]. The NSA expands the inverse matrix $\mathbf{G}^{-1}$ as

$$\mathbf{G}^{-1} \approx \sum_{i=0}^{\infty} \left(\mathbf{I}_N - \mathbf{X}^{-1}\mathbf{G}\right)^i \mathbf{X}^{-1}, \tag{12}$$

where $\mathbf{X}$ is the matrix of an initial approximation of $\mathbf{G}^{-1}$. The matrix $\mathbf{X}$ must satisfy

$$\lim_{i\to\infty} \left(\mathbf{I}_N - \mathbf{X}^{-1}\mathbf{G}\right)^i \simeq \mathbf{0}_N. \tag{13}$$

For a reasonable convergence for the NSA algorithm, the value of $\lambda_{max}$ must be less than one ($|\lambda_{max}| < 1$) where $\lambda_{max}$ is the largest magnitude of eigenvalue of the $\left(\mathbf{I}_N - \mathbf{X}^{-1}\mathbf{G}\right)$ matrix [118]. For the non-diagonally matrix $\mathbf{G}$=$\mathbf{D}$+$\mathbf{E}$, the $\mathbf{G}^{-1}$ can be represented as

$$\mathbf{G}^{-1} \approx \sum_{i=0}^{\infty} \left(\mathbf{D}^{-1}\mathbf{E}\right)^i \mathbf{D}^{-1}, \tag{14}$$

where the condition

$$\lim_{i\to\infty} \left(-\mathbf{D}^{-1}\mathbf{E}\right)^i \simeq \mathbf{0}_N, \tag{15}$$

must be satisfied. The NSA algorithm minimizes the computational complexity to $\mathcal{O}(K^2)$ instead of $\mathcal{O}(K^3)$ at the number of iterations lower than 2 ( $i \leqslant 2$) [107], [119]. In [120], a weighted NSA (WNS) algorithm for massive MIMO systems based on the large Wishart matrix properties is proposed to fast up the convergence of the NSA algorithm.

The WNS algorithm offers a noteworthy increase in the convergence and weights are not sensitive to channel investigation in case of i.i.d. at a low to reasonable correlation factor. In [121], by exploiting the properties of the WNS algorithm, a weighted Neumann series-steepest descent (WNS-SD) iterative precoder is proposed to obtain a fast convergence while maintaining low-complexity. Also in [121], an accelerated weighted Neumann series-steepest descent (AWNS-SD) precodig algorithm is proposed. The AWNS-SD algorithm has a remarkable increase in the convergence rates while maintaining low-complexity and guaranteeing a wide range of convergence. The AWNS-SD algorithm has a near performance of the ZF algorithm in only one iterative step for identical massive MIMO systems.

c) Newton Iteration (NI) Algorithm: The NI algorithm is a method used in approximating the matrix inversion. It is the counterpart of the Newton-Raphson method [122]. The matrix inversion estimation at the $i^{\text{th}}$ iteration is

$$\mathbf{X}^{(i)} = \mathbf{X}^{i-1}\left(2\mathbf{I}_N - \mathbf{G}\mathbf{X}^{i-1}\right), \tag{16}$$

while

$$\mathbf{G}^{-1} = \lim_{n\to\infty} \mathbf{X}^{(i)} \; when \; \left\|\mathbf{I}_N - \mathbf{G}\mathbf{X}^0\right\| < 1, \tag{17}$$

where $\mathbf{X}^0$ is an initial rough estimation. Finding the initial value in the NI algorithm is complicated and needs extra calculations. Besides that, the NI algorithm needs a significant number of iterations to have fast convergence [116]. High reliability with quadratic convergence can be offered in the NI algorithm [122].

Similar to the NSA algorithm, the NI algorithm just needs a simple computation to speed the precoding process. Though the NI needs one extra matrix multiplication in each iteration, it converges faster than the NSA algorithm [122]. In [118], to accelerate the convergence, a joint NI algorithm and NSA algorithm (NI-NSA) is proposed. The first iteration of the NI algorithm is exploited to re-extract the NSA algorithm series, which leads to a high probability convergence. Numerical results show that the joint NI-NSA algorithm has a more efficient and speed convergence rate compared to the NSA algorithm, without increasing the computational complexity at later iterations ($i \geqslant 2$). The joint NI-NSA algorithm also derives a high probability convergence condition about $M/N$ ratio.

d) Chebyshev Iteration (CI) Algorithm: The CI algorithm is also approximating the matrix inversion process in linear precoding for massive MIMO systems by employing iterative computation [109]. The matrix inversion estimation at the $i^{\text{th}}$ iteration is

$$\mathbf{X}^{(i)} = \mathbf{X}^{i-1}\left(3\mathbf{I}_N - \mathbf{G}\mathbf{X}^{i-1}\left(3\mathbf{I}_N - \mathbf{G}\mathbf{X}^{i-1}\right)\right), \tag{18}$$

while

$$\mathbf{G}^{-1} = \lim_{n\to\infty} \mathbf{X}^{(i)} \; when \; \left\|\mathbf{I}_N - \mathbf{G}\mathbf{X}^0\right\| < 1. \tag{19}$$

However, finding the initial value in the CI algorithm, as in the NI algorithm, is complicated and related to the eigenvalues of the matrix $\mathbf{G}$. Where, the convergence rate of the CI algorithm is affected by iterative initial values, which are needed to determine carefully [107]. Though the CI algorithm needs two matrix additions and three matrix mul-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



tiplications (18), it converges faster than the NI algorithm [109].

In [109], the optimization of initial values is carried out for the CI algorithm. Thus, the initial values become easier to be acquired. The result of simulation shows that the optimized CI algorithm offers the same achievable average rate as the RZF algorithm's rate after just two iterations. The optimized CI algorithm offers the same performance as the CI algorithm after just one iteration.

In [100], a precoding technique using the joint CI and NSA (CI-NSA) algorithm is proposed to achieve the near-optimal performance. The CI-NSA algorithm optimizes the NSA algorithm by CI method, which converges faster than the other existing NSA precoding algorithms. The computational complexity of the CI-NSA algorithm is similar to the NSA and NI-NSA algorithms. Nevertheless, the CI-NSA algorithm offers a faster convergence rate with fewer iterations and with the same performance. Thus, the CI-NSA algorithm offers a trade-off between performance and complexity.

3) Fixed-Point Iteration-Based Algorithms:

The fixed-point iteration-based algorithms approach to realize **G** by solving a linear precoding equation **x** in **Gx** = **a** iteratively, instead of **x** = $\mathbf{G}^{-1}$**a** directly [113]. Subsequently, precoders based on fixed-point iteration algorithms approach can be described as below:

a) Gauss-Seidel (GS) Algorithm: When $M$ is very large and $M >> N$, the **G** matrix becomes diagonally dominant and meets a symmetric positive definite condition of the GS algorithm. The GS algorithm is mainly used to iteratively solve a linear precoding equation without a matrix inversion. The GS algorithm is also known as the Liebmann algorithm or the method of successive algorithm [20]. By factorized **G** matrix into diagonal matrix **D**, lower-triangular matrix **L** and upper-triangular matrix **U** as **G=D+L+U** [117], the GS algorithm can be presented as

$$\mathbf{x}^{(i)} = (\mathbf{D}+\mathbf{L})^{-1}\left(\mathbf{a}-\mathbf{U}\,\mathbf{x}^{(i-1)}\right). \tag{20}$$

The computational complexity of the GS algorithm is $\mathcal{O}(K^2)$ as a result of using matrix-vector multiplication. Unlike the matrix inversion approximation algorithms, the GS algorithm can be relaxed as selecting an initial vector $\mathbf{x}^0$. $\mathbf{x}^0$=$\mathbf{D}^{-1}$ can be a good initial estimation [123]. Where the GS algorithm exploits the most recent values at each iteration, its BER performance is better than the NSA algorithm with lower complexity at the same number of iterations [124]. In addition, the GS algorithm converges a bit faster than the NSA algorithm [99], [117]. In [125], a GS-based matrix inversion approximation (GSBMIA) algorithm is proposed which simplifies calculations by approximating the matrix inversion process. The GSBMIA algorithm has a similar convergence rate of the GS algorithm. In order to speed up the convergence rate of the GSBMIA algorithm, it is combined with the NI algorithm. Numerical results show that the GSBMIA algorithm and joint algorithm have a faster convergence rate than the approximate matrix inversion algorithms like the NSA and NI algorithms.

b) Successive Over-Relaxation (SOR) Algorithm: The SOR algorithm is proposed to enhance the convergence rate of the GS algorithm by employing a variable relaxation factor of $\omega$. Therefore, the SOR algorithm, like the GS algorithm, offers a good performance at the starting of iteration [116]. By decomposing **G** matrix as **G=D+L+U**, the SOR algorithm's equation is expressed as

$$\mathbf{x}^{(i)} = (\mathbf{D}-\omega\mathbf{L})^{-1}\left[(\omega\mathbf{U}+(1-\omega)\mathbf{D})\mathbf{x}^{(i-1)}+\omega\mathbf{a}\right]. \tag{21}$$

The $\omega$ acts as a decisive function in the convergence rate of the SOR algorithm. If $\omega = 1$, the SOR algorithm works as the GS algorithm. The SOR algorithm is being convergent when $0 < \omega < 2$ [126]. It also outperforms the NSA algorithm with lower complexity [126]. The convergence rate of the SOR algorithm is faster than the GS algorithm [127].

However, the SOR algorithm has a higher complexity than the GS algorithm [128]. In addition, the convergence rate of the SOR algorithm is not fast enough. In [129], a symmetric SOR (SSOR) algorithm is proposed to reduce the complexity of the SOR algorithm. Where each iteration in the SSOR algorithm has two half iterations. The first half iteration is similar to the SOR iteration and the second half iteration is similar to the SOR algorithm with reverse order equations. The performance of the SSOR algorithm is close to the MMSE algorithm performance. In [130], a low-complexity method to enhance the SOR algorithm is proposed to obtain a fast convergence rate compared with other methods. Linear fitting method has been used to determine the best relaxation parameter. However, it obtains almost the same computational complexity. Numerical results show that the proposed SOR algorithm overcomes other iterative algorithms in terms of BER. In [116], the joint SOR matrix inverse and NI (SORMI-NI) algorithm is proposed to get an overall advantage within the iteration. The SORMI iteration is done before the NI iteration to offer the initial value of the NI algorithm and offer more efficient and speed searching. By doing that, the NI algorithm achieves fast convergence at an early stage of the iteration. Besides that, the SORMI algorithm solves the problem of difficultly to isolate $\mathbf{G}^{-1}$**a** in the SOR algorithm.

In [131], four joint algorithms are proposed to find linear precoding factors in massive MIMO systems and then obtain more speed convergence with low complexity. The first algorithm is the joint CI and NSA algorithm (CI-NSA), which accelerates the convergence rate of the NSA algorithm with more delicate inversion. The second algorithm is the SOR-based approximate matrix inversion (SOR-AMI), which offers a direct simplified matrix inversion with the same convergence rate of the traditional SOR algorithm. The third and fourth algorithms are extension and improvement of SOR-AMI and they are called the NI-SOR-AMI and CI-SOR-AMI. These four proposed algorithms offer a near-optimal BER performance of the ZF algorithm. The convergence rate of the proposed CI-NSA algorithm is faster than

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



the conventional NSA algorithm with the same complexity. Likewise, the convergence rates of the CI-SOR-AMI and NI-SOR-AMI algorithms are faster than the conventional SOR algorithm.

c) Conjugate Gradient (CG) Algorithm: The CG algorithm is another good method to solve the linear equations iteratively by avoiding the matrix inversion process. It can achieve the same performance as the MMSE algorithm after a few iterations with lower computational complexity by about one order of magnitude. In addition, the CG algorithm has higher BER performance, higher capacity, and lower computational complexity than the NSA algorithm [127]. The CG algorithm is also known as Lanczos orthogonalization algorithm [107]. The CG iterations can be described as

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \alpha^{(i-1)}\mathbf{F}^{(i)}, \tag{22}$$

where $\mathbf{F}^{i}$ is the conjugate direction matrix related to $\mathbf{G}$, as

$$\left(\mathbf{F}^{(i)}\right)^{\mathbf{H}} \mathbf{G}\ \mathbf{F}^{(j)} = 0, \quad \text{for} \quad i \neq j, \tag{23}$$

where $\alpha^{(i-1)}$ is a scalar factor and denotes the step size and $\mathbf{G}$ must be a symmetric positive definite matrix to make the CG algorithm convergent. The CG algorithm contains several division processes and a large number of iterations, and the degree of parallelism is low [91], [107], [118]. In the CG algorithm, the zero vector is usually used as an initial solution [107].

In [107], a new algorithm named as a three-term-recursion conjugate gradient (TCG) algorithm is proposed. It mainly aims to make the algorithm to have fast convergence besides reducing the computational complexity by selecting a good initial value. Also, the TCG algorithm has higher parallelism than the conventional CG algorithm.

In [132], a novel low complexity algorithm for the linear precoding in massive MIMO systems based on the CG algorithm with asymmetric $\mathbf{G}$ matrix is proposed. The algorithm has two versions of the CG algorithms: the first version is conjugate gradient squared (CGS) and the second is Bi-conjugate gradient (Bi-CG). The two novel algorithms overcome the conventional CG algorithm in terms of convergence speed and BER performance.

d) Jacobi Iteration (JI) Algorithm: The JI is a simple iterative algorithm used to find a solution of $\hat{\mathbf{X}} = \mathbf{G}^{-1}\mathbf{a}$. By decomposing $\mathbf{G}$ into a diagonal matrix $\mathbf{D}$ and off-diagonal matrix $\mathbf{R}$, the estimated signal can be presented as

$$\mathbf{x}^{(i)} = \mathbf{D}^{-1}\left[\mathbf{a} + (\mathbf{D} - \mathbf{G})\,\mathbf{x}^{(i-1)}\right], \tag{24}$$

which must satisfy

$$\lim_{i\to\infty}\left(\mathbf{I} - \mathbf{D}^{-1}\mathbf{G}\right)^{i} = 0. \tag{25}$$

The initial matrix of the JI precoder can be presented as

$$\mathbf{x}^{(0)} = \mathbf{D}^{-1}\mathbf{a}. \tag{26}$$

The JI algorithm has lower performance and lower convergence rate than the GS and SOR algorithms [99], [121], [124]. Conversely, the JI algorithm enjoys parallelism and effective hardware implementation and has $\mathcal{O}(K^2)$ computational complexity which is lower than the complexity of the NSA, GS, and SOR algorithms [117], [121], [128].

In [133], a joint JI and steepest descent algorithm (JI-SD) is proposed to obtain a good direction of searching for the JI algorithm to increase the convergence rate. The convergence rate of the JI-SD algorithm is not met for the large number of $M$ in massive MIMO systems. In [134], a new joint JI and CG (JI-CG) algorithm is proposed to speed up the convergence rate of the JI algorithm. In the joint algorithm, to discover a more delicate searching direction for the JI algorithm, the CG algorithm is employed two times. The JI-CG algorithm overcomes the JI-SD in terms of BER or at least has similar BER with faster convergence rate at lower complexity and latency.

4) Precoding Based on Matrix-Decomposition:

The direct algorithms-matrix decomposition precoder for massive MIMO systems is conventionally used for the matrix inversion process instead of using an explicit matrix inversion in small-scale MIMO systems [135]. It is numerically stable over the basic linear precoder algorithms such as the MRT, ZF, and MMSE algorithms. Besides that, it can be employed to offer a modular design, where the inversion process can be dispensed between different parts [136]. However, the employing of the direct algorithms-matrix decomposition in the massive MIMO systems has a considerable computational complexity. Where the direct algorithms-matrix decomposition needs to decompose the $\mathbf{G}$ matrix into a multiplication of small matrices as in the QR algorithm and the Chelosky decomposition algorithm [136].

In spite of the significance of the direct algorithms-matrix decomposition, their analysis of complexity for massive MIMO systems and the differentiation with existing fixed-point iteration-based algorithms and matrix inversion approximation algorithms is lacking in the literature [135], [137].

a) QR Decomposition Algorithm: The QR decomposition algorithm can be applied to get the solution of (10) as

$$\begin{aligned}\mathbf{y}_{\mathrm{MMSE}} &= \sqrt{\beta}\sqrt{\rho}\,\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left(\mathbf{G} + \mathbf{V} + \lambda\,\mathbf{I}_{N}\right)^{-1}\mathbf{a} + \mathbf{n}\\ &= \sqrt{\beta}\sqrt{\rho}\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left(\mathbf{QR}\right)^{-1}\mathbf{a} + \mathbf{n}\\ &= \sqrt{\beta}\sqrt{\rho}\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left(\mathbf{R}^{-1}\mathbf{Q}^{\mathbf{H}}\right)\mathbf{a} + \mathbf{n},\end{aligned} \tag{27}$$

where $\mathbf{Q}$ is $N \times N$ unitary matrix and contains orthogonal columns and $\mathbf{R}$ is an $N \times N$ upper triangular matrix.

The block diagonalization algorithm (BD) is a famous linear precoding decomposition algorithm for DL transmission in multi-user MIMO (MU-MIMO) systems [138]. The BD algorithm offers good performance but with a high computational complexity where each user needs to employ two singular value decomposition (SVD) operations. In [139], the QR decomposition algorithm is used with the BD algorithm to reduce the complexity, by using the QR operation instead of the first SVD operation. Where this joint algorithm known as the QR decomposition based BD algorithm (QR-BD).

Also in [139], a new algorithm, QR decomposition and Gram Schmidt (QR-GS) algorithm, is introduced for DL transmission in multi-user MIMO (MU-MIMO) systems. The QR-GS algorithm offers similar performance as the BD and QR-BD algorithms with a significant reduction in computational complexity.

In [140], an improved precoding algorithm for large-scale MU-MIMO systems is proposed. This algorithm is a joint algorithm that consists of the BD algorithm with QR decomposition of the ZF matrix and is known as the QR-ZF-BD algorithm. The QR-ZF-BD algorithm employs the ZF algorithm and the QR decomposition algorithm instead of a complex SVD process in the classical BD algorithm. The QR-ZF-BD algorithm has two stages. Firstly, it uses the QR decomposition algorithm to minify the multi-user interference (MUI). Secondly, it uses the ZF and QR decomposition algorithms again to increase the spectral efficiency. The result of the simulation shows that the QR-ZF-BD algorithm offers better spectral efficiency than other recent decomposition algorithms.

In [141], An MMSE-based QR-BD (QR-MMSE-BD) precoding algorithm was proposed which has lower complexity than both the BD and QR-BD algorithms.

b) Cholesky Decomposition (CD) Algorithm: The CD algorithm can be applied to get the solution of (10) as

$$\begin{aligned}\mathbf{y}_{\text{MMSE}} &= \sqrt{\beta}\sqrt{\rho}\,\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left(\mathbf{G}+\mathbf{V}+\lambda\,\mathbf{I}_{N}\right)^{-1}\mathbf{a}+\mathbf{n}\\ &= \sqrt{\beta}\sqrt{\rho}\;\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left(\mathbf{L}\mathbf{L}^{\mathbf{H}}\right)^{-1}\;\mathbf{a}+\mathbf{n}\\ &= \sqrt{\beta}\sqrt{\rho}\;\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\left((\mathbf{L}^{\mathbf{H}})^{-1}(\mathbf{L})^{-1}\right)\;\mathbf{a}+\mathbf{n},\end{aligned} \tag{28}$$

where a matrix $\mathbf{L}$ is the lower triangular matrix. In [142], utilization of the CD algorithm with Sherman-Morrison strategy (CSM) is proposed. The CSM algorithm contributes to solving the problem of basic linear precoding algorithms in massive MIMO systems which need to get a large-size matrix inversion operation and have high computational complexity. The CSM algorithm offers a near-optimal performance of the MMSE algorithm by iteratively decomposing the large-size matrix inversion process. The result of simulation shows that the CSM algorithm has better BER and sum-rate performance than the NSA and SOR algorithms with fewer operations and reducing the computational complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N^2)$.

For smooth readability and comparison, the computational complexity of the linear precoding algorithms is presented in Table 3 and their pros and cons are comprehensively reviewed in Table 4.

### B. NON-LINEAR PRECODING

As mentioned above, there are two main classes of signal precoders for massive MIMO systems: linear precoders and non-linear precoders. Though, the linear precoder algorithms have the advantages of low complexity, their insufficiency in precoding accuracy cannot be neglected, particularly when $M/N$ is close or equal to one [110]. The optimum signal precoder is the ML precoder. Unfortunately, the ML precoder's complexity increases exponentially with the increase $M$, so it is unattainable to implement in massive MIMO systems [152]. This section introduces the most used non-linear massive MIMO signal precoding algorithms, and its related advantages and disadvantages.

1) Dirty-Paper Coding (DPC):

The DPC algorithm was alluded by Costa in 1983, which evidenced that the capacity of the theoretical channel can be offered and the interference can be annulled when the interference is known at the transmitter side [153]. In MU-MIMO systems, when the precoding matrix is designed for the $n^{\text{th}}$ received terminal, the interference that comes from the first up to $(n-1)^{\text{th}}$ received terminals are deemed to be annulled. Besides that, the DPC algorithm can offer a remarkable performance without needing extra power in the transmission side and without sharing CSI with the receiver side. However, the DPC algorithm is impracticable, because it needs an infinite length of codewords and sophisticated signal processing [154], [155].

The DPC algorithm has been proposed to offer the optimum DL sum-rate for massive MIMO systems, where the idea of the DPC algorithm is that the sum-rate of a system is equal to the sum-rate of a free-interference system when the interference is known at the transmitter side [154]. The sum-rate of the DPC algorithm can be presented as

$$C = \max_{\mathbf{W}} \log_2 \det(\mathbf{I}_N + \mathbf{H}^{*}\mathbf{W}\mathbf{H}^{\mathbf{T}}) \quad \text{bits/s/Hz}, \tag{29}$$

where $\mathbf{W}$ is a $N \times N$ diagonal power allocation matrix, and $\sum \text{diag}(\mathbf{W}) = 1$. In [156], a novel non-linear precoding algorithm known as ZF-DPC is proposed. It is a suboptimal DPC algorithm with lower complexity. The ZF-DPC algorithm is based on the QR decomposition of the channel where it is assumed that users have a single receive antenna.

In [20], the performance of the ZF-DPC algorithm has been scrutinized in the condition of the rayleigh fading model, and QAM modulation with 100,000 Monte-Carlo trials. The ZF-DPC algorithm can overcome the conventional MMSE algorithm by approximately 3 dB for 64-QAM at a higher SNR.

2) Tomlinson-Harashima (TH) Precoding:

The TH precoding algorithm is a suboptimal implementation algorithm of the DPC algorithm, which is a combination of the DPC algorithm and the modulo arithmetic [157], [158]. The TH algorithm is proposed by Tomlinson and Harashima in 1972 [157], [159]. The TH algorithm is originally an equalization process proposed to repeal the ISI [160]. Besides that, the TH algorithm can be used to clear the sub-channels interference in MIMO systems [98]. Though the TH algorithm experiences a loss of performance in contrast with the DPC algorithm, it has a practical implementation. The TH algorithm has more complexity when compared to linear precoding algorithms but efficaciously eschew the noise amplification [98]. The TH algorithm has three prime components, the feedforward filter, the feedback filter, and

TABLE 3: Computational Complexity of the Linear Precoding Algorithms, $i$ is the number of the required iteration.

| Algorithm | Computational complexity | Complexity order |
|---|---|---|
| ZF | $N^3 + 2MN^2 + MN + M$ [100] | $\mathcal{O}(N^3)$ |
| MMSE | $3N^3 + 2N^2 + NM^2 + MN^2 + M$ [116] | $\mathcal{O}(N^3)$ |
| NSA | $(i-2)N^3 + MN^2 + N^2 + 2MN + M$ [117] | $\mathcal{O}(N^2)$ for $i \leqslant 2$ |
| WNS | $(i-2)N^3 + MN^2 + N^2 + 2MN + M + N$ [120] | $\mathcal{O}(N^3)$ |
| NI | $i2N^3 + N^2 + MN + M$ [116] | $\mathcal{O}(N^3)$ |
| GS | $i4N^2 + NM + M$ [134] | $\mathcal{O}(N^2)$ |
| CG | $i(4N^2 + 10N) + NM + M$ [134] | $\mathcal{O}(N^2)$ |
| CI | $2N^3 + i8N^2 + 2N^2 + 2MN^2 + 2MN + 2$ [113] | $\mathcal{O}(N^3)$ |
| JI | $i(4N^2 - 2N) + NM + M$ [134] | $\mathcal{O}(N^2)$ |
| SOR | $i(4N^2 + 4N) + NM + M$ [128] | $\mathcal{O}(N^2)$ |
| Improved SOR | $i(N^2 + N) + NM + M$ [130] | $\mathcal{O}(N^2)$ |
| SSOR | $i(8N^2 + 8N) + NM + M$ [128] | $\mathcal{O}(N^2)$ |
| SORMI | $i(N^3 + N^2) + NM + M$ [116] | $\mathcal{O}(N^3)$ |
| WNS-SD | $i(N^2 + 2N) + 2N^2 + 7N + MN + M + 7$ [121] | $\mathcal{O}(N^2)$ |
| AWNS-SD | $i(N^2 + 2N) + 4N^2 + 17N + MN + M + 7$ [121] | $\mathcal{O}(N^2)$ |
| NI-NSA | $(i-2)N^3 + 3MN + N + M$ [100] | $\mathcal{O}(N^2)$ for $i \leqslant 2$ |
| CI-NSA | $(i-2)N^3 + 4MN + N + M$ [100] | $\mathcal{O}(N^2)$ for $i \leqslant 2$ |
| JI-SD | $i(4N^2 - 2N) + 12N + MN + M$ [134] | $\mathcal{O}(N^2)$ |
| SORMI-NI | $(2-i)N^3 + 2N^2 + NM + M$ [116] | $\mathcal{O}(N^2)$ for $i \geqslant 2$ |
| SOR-AMI | $iN^2 + 2NM + M$ [131] | $\mathcal{O}(N^2)$ |
| NI-SOR-AMI | $2N^3 + (i-1)N^2 + 2NM + M$ for $i \leqslant 3$ and $2N^3 + N^2 i + 2NM + M$ for $i > 3$ [131] | $\mathcal{O}(N^3)$ |
| CI-SOR-AMI | $3N^3 + iN^2 + 2NM + M$ [131] | $\mathcal{O}(N^3)$ |
| JI-CG | $2(4N^2 + 8N) + NM + M$ for $i = 1$ and $i(4N^2 - 2N) + 24N + NM + M$ for $i \geqslant 2$ [134] | $\mathcal{O}(N^2)$ |
| CSM | $4N^2 - 3N - 1$ [142] | $\mathcal{O}(N^2)$ |

the modulo arithmetic [98], [161].

Figure 6 shows the block diagram of the TH algorithm where it is assumed that $M$ is equal to $N$ [98]. Based on LQ decomposition, the TH algorithm can be carried out by decomposing $\mathbf{H^T}$ to the multiplying of a lower triangular matrix $\mathbf{L}$ and a unitary matrix $\mathbf{Q}$ as [98]

$$\mathbf{H^T} = \mathbf{L}\,\mathbf{Q}. \tag{30}$$

A matrix $\mathbf{K}$ is a scalar matrix which weighting a coefficient of each sub-stream and has a diagonal format [98]. The diagonal elements of $\mathbf{K}$ are the inverse of the diagonal elements of $\mathbf{L}$ matrix [98]. The $\mathbf{K}$ matrix can be represented as

$$\mathbf{K} = \begin{bmatrix} l_{11}^{-1} & 0 & 0\cdots & 0 \\ 0 & l_{22}^{-1} & 0\cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0\cdots & l_{NN}^{-1} \end{bmatrix}, \tag{31}$$

where the $l_{ij}$ for $i, j = 0, ..., N$ is a diagonal element of $\mathbf{L}$ matrix. The matrix $\mathbf{B}$ is a feedback (pre-cancellation) matrix, which is a lower triangular matrix and all diagonal elements are ones [98]. The feedback matrix $\mathbf{B}$ is used to remove the previous stream interference from the immediate stream and it can be represented as [98]

$$\mathbf{B} = \mathbf{K}\,\mathbf{L}. \tag{32}$$

The matrix $\mathbf{F}$ is a feedforward matrix which is a conjugate transpose of $\mathbf{Q}$ and utilized to save a transmitted power constant and to compel the spatial causality at the transmitter side [98], [162].

$$\mathbf{F} = \mathbf{Q^H}. \tag{33}$$

The pre-cancellation process in the TH algorithm leads to an increase of the power of each stream layer [98]. Modulo arithmetic is used in the transmitter and the receiver sides to adjust the average power [98], [162]. The modulo operation is tightly related to the utilized constellation $A$. Assuming that the $M$-ary square of QAM constellation is exploited to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



TABLE 4: Pros and Cons of the Linear Precoding Algorithms in Massive MIMO Systems

| Algorithm | Pros | Cons |
|---|---|---|
| MRT | • It supports locally signal processing in each antenna in the BS [101].<br>• More flexibility.<br>• It has a close to optimal performance in the noise-limited system [101].<br>• If the propagation matrix has nearly orthogonal columns, it works properly [21]. | • It is not robust against the IUI [21], [127].<br>• It does not able to offer a full diversity at a high spectral efficiency [127].<br>• It offers a low performance if the channel is an ill-conditioned [16]. |
| ZF | • It has a higher data rate than the MRT algorithm [127].<br>• It eliminates the IUI [127].<br>• It offers a trade-off between performance and complexity.<br>• It has a close to optimal performance in the interference-limited system [16]. | • It does not consider the effects of noise [116].<br>• It has a complicated matrix inversion process with a large ratio of $(M/N)$ [16].<br>• It amplifies the noise [21], [127].<br>• It offers a low performance if the channel is ill-conditioned [21], [127].<br>• It dos not have the ability to enhance the diversity gain or reduce computational complexity when the ratio of $(M/N)$ goes to one [16]. |
| MMSE | • It offers a trade-off between the MRT and ZF algorithms [21], [116].<br>• It has a better performance in a noisy environment than ZF algorithm [16].<br>• It eliminates the IUI [143].<br>• It considers the effects of noise [16], [116].<br>• It offers a near-optimal performance with a proper channel propagation case [32]. | • The **G** matrix is must be a symmetric positive definite [144].<br>• It has a complicated matrix inversion method especially when $M$ is very high [143].<br>• It dos not have the ability to enhance the diversity gain or reduce the computational complexity when the ratio of $(M/N)$ goes to one [101], [145]. |
| TPE | • It has more stability and better performance than the the NSA algorithm [121].<br>• It has an optimized polynomial coefficient to speed up the convergence rate [129]. | • It needs an additional computational complexity over the NSA algorithm [121].<br>• It has a larger delay than the NSA algorithm [121].<br>• The determination of the optimized coefficient is complicated and must be updated in each iteration [121]. |
| NSA | • It contains matrix multiplication and addition processes only without division processes [118].<br>• It has simple hardware implementations [120].<br>• It has a near-optimal performance of the ZF algorithm when the ratio of $(M/N)$ is large [16], [100].<br>• It has a marginal lower complexity than the ZF algorithm when the ratio of $(M/N)$ is large [100].<br>• It does not need a complicated optimization parameter as the TPE algorithm [129].<br>• It can be highly parallelized [115].<br>• It has a simple data-flow [115]. | • It converges only when the ratio of $(M/N) > 5.83)$ [118], [121].<br>• It has a slow convergence rate [120].<br>• When $M$ is not large enough, it shows the same computational complexity of the MMSE algorithm [121].<br>• It has a significant loss of performance in large massive MIMO systems [146]. |
| WNS | • It has a faster convergence rate than the conventional NSA algorithm [120].<br>• The computations of its optimized weights is not affected by instantaneous channel realization [120].<br>• It has the same cost of complexity of the conventional NSA algorithm [120].<br>• It has a near-optimal performance compared to optimal exact inversion algorithms at a fourth term of series [120]. | • It can not converge when $N$ is very large in high mobility environment [120].<br>• The **G** matrix is must be a symmetric positive definite [121].<br>• It needs additional calculations to find a suitable optimized weights [121]. |
| NI | • It has a high degree of parallelism [147].<br>• The matrix inversion process is easy to be realized [109], [116], [147].<br>• It has a fast convergence rate during the later stage of iterations [116].<br>• The complexity can be controlled only by the number of iterations [116], [118]. | • Its initial value calculations is very complicated [116], [118].<br>• It has two order of convergence (slow convergence rate) [109], [116].<br>• It is a high complexity algorithm compared to the other iterative algorithms [109]. |
| CI | • It has a fast convergence rate during the whole algorithm iterations [100].<br>• It has a third-order convergence which is faster than the NI algorithm [100], [109].<br>• It can offer the same achievable rate of the MMSE algorithm after just two iterations [109].<br>• It needs fewer iterations to obtain the same performance of the NI algorithm at the same initial value [109]. | • It needs an accurate selection of initial value to guarantee convergence [107], [109].<br>• It has a higher complexity than the NI algorithm [109]. |
| GS | • It has a faster convergence rate than the JI and NSA algorithms [123], [124], [148].<br>• It offers a capacity-approaching performance of conventional linear precoders by an iterative manner without complex matrix inversion and with few number of iterations [113], [148].<br>• It employs the optimal relaxation factor to increase the convergence rate [149].<br>• It can offer the same achievable rate of the MMSE algorithm after just two iterations [149].<br>• It has a better performance on the beginning iterations [116].<br>• It has lower computational complexity than the MMSE algorithm by one order of magnitude [116], [144].<br>• It can converge for any initial value [148].<br>• It depends on the last updated value in each iteration [148].<br>• It has a double convergence rate of the conventional JI algorithm [150]. | • It is difficult to be implemented in parallel [121], [134].<br>• It needs additional calculations to find a suitable optimized factor.<br>• The **G** matrix is must be a Hermitian positive definite [144].<br>• The **G** matrix is must be a symmetric positive definite [144], [150]. |
| SOR | • It offers a near-optimal performance of the MMSE algorithm with a few number of iterations [128].<br>• It employs a variable relaxation factor to enhance the convergence rate [116], [128].<br>• It can offer a good performance at beginning of iterations [116].<br>• Its convergence rate is faster than the JI and GS algorithm rate [127]. | • It is very sensitive to relaxation factor [128].<br>• Its performance is degraded in the later iterations.<br>• The **G** matrix is must be a symmetric positive definite [126].<br>• It converges only when $0 < \omega < 2$ [128].<br>• It is difficult to find its optimal relaxation factor [130].<br>• It has a less degree of parallelism than that in the JI algorithm [151]. |
| CG | • It offers a fast convergence rate with at most $N$ number of iterations [91].<br>• It offers higher capacity-approaching, BER performance, and lower complexity than the conventional NSA algorithm [127].<br>• It offers a near-optimal performance of the ZF algorithm with a few number of iterations [127].<br>• It offers zero value of relative residue [132].<br>• It has lower computational complexity than the MMSE algorithm by one order of magnitude [127].<br>• It has an accurate gradient searching direction [132]. | • It has many division processes [150].<br>• The **G** matrix must be a symmetric positive definite [127].<br>• It has a higher complexity in each gradient computation [134]. |
| CGS | • It can converge when **G** matrix is an asymmetric positive definite [132].<br>• It has a higher convergence rate with lower relative residue than the CG algorithm [132].<br>• It does not need to compute the transpose of **G** matrix as in the BI-CG algorithm [132].<br>• It has a double convergence rate of the BI-CG algorithm with a lower complexity [132].<br>• Its relative residue is equal to zero in the $(N-1)^{\text{th}}$ iteration or before [132]. | • It has an irregular behavior of convergence [132].<br>• It needs to make a balance between the convergence rate and accuracy [132]. |
| BI-CG | • It is more accurate than the CGS [132].<br>• It has a lesser relative residue than the CGS algorithm [132].<br>• It can converge when **G** matrix is an asymmetric positive definite [132].<br>• It has a higher convergence rate with lower relative residue than the CG algorithm [132]. | • It needs to compute the transpose of **G** matrix [132].<br>• The **G** matrix is must be a symmetric positive definite [132].<br>• It needs to make a balance between the convergence rate and accuracy [132]. |
| JI | • Its convergence rate increase linearly with increasing $M$ [121].<br>• It offers a near-optimal performance of the ZF algorithm with few iterations [146].<br>• It offers capacity-approaching of the ZF algorithm few iterations [146].<br>• It has lower computational complexity than the ZF algorithm by one order of magnitude [146].<br>• It has an easy implementation [121].<br>• It can offer the same achievable rate of the MMSE algorithm after just two iterations [149]. | • It converges only when the ratio of $(M/N) > 5.8)$ [146], [125].<br>• The **G** matrix is must be a symmetric positive definite [146].<br>• It has lesser robustness and speed than the GS algorithm [117].<br>• It has a high latency [134].<br>• It has a low convergence rate because of the possible oscillations [128]. |
| SSOR | • It offers a near-optimal performance of the ZF algorithm [128].<br>• It has a lesser sensitivity to the relaxation parameter than the SOR algorithm [128], [129].<br>• It offers more flexibility for practical systems [129].<br>• It has lower computational complexity than the ZF algorithm by one order of magnitude without performance loss [129].<br>• It optimizes the relaxation parameter, which is dependent just on the dimension of massive MIMO system [129].<br>• It is suitable for a fast time-varying system [129].<br>• It has a higher performance than the NSA and TPE algorithms [129]. | • It needs accurate calculations for the optimized relaxation factor [129].<br>• The **G** matrix is must be a Hermitian positive definite [129].<br>• It has a double complexity cost over the conventional SOR algorithm [128]. |

| | | |
|---|---|---|
| SOR-AMI | • It has a direct simple matrix inversion process [131].<br>• It has lower computational complexity than the conventional SOR, while maintains the same convergence rate [131].<br>• It has a higher convergence rate than the CI-NSA algorithm [131].<br>• It offers a near-optimal performance of the ZF algorithm [131]. | • It has a lower convergence rate than the CI-SOR-AMI, and NI-SOR-AMI algorithms [131]. |
| NI-SOR-AMI | • It has a faster convergence rate than the SOR, SOR-AMI, and CI-NSA algorithms [131].<br>• It offers a near-optimal performance of the ZF algorithm under correlated and uncorrelated channels [131].<br>• It is favorite because of its lower complexity and more accurate inversion result [131].<br>• It offers an efficient direction for the searching process for the SOR-AMI algorithm [131].<br>• It has a high probability of convergence [131]. | • It has a lower convergence rate than the CI-SOR-AMI algorithm [131]. |
| CI-SOR-AMI | • It has a faster convergence rate than the SOR, SOR-AMI, CI-NSA, and NI-SOR-AMI algorithms [131].<br>• It offers a near-optimal performance of the ZF algorithm under correlated and uncorrelated channels [131].<br>• It offers an efficient direction for the searching process for the SOR-AMI algorithm [131].<br>• It has a high probability of convergence [131]. | • It is not preferable in practice due to its high complexity and not accurate inversion result compared to the NI-SOR-AMI algorithm [131]. |
| NI-SOR | • It has a global speed convergence during the iterations [116].<br>• It has more effective and faster searching than the NI and SOR algorithms [116].<br>• It needs a lower hardware cost than the NI and SOR algorithms [116]. | • It needs accurate calculations for the optimized relaxation factor of the SOR algorithm to determine the initial value of the joint algorithm [116]. |
| NI-SORMI | • It is robust for different ratio of $(M/N)$ [116].<br>• It has lower computational complexity than the MMSE algorithm [116].<br>• It speeds up the convergence rate of the NI algorithm [116]. | • It has a high computational complexity compared with the JI, NI, and SOR algorithms [116]. |
| CI-NSA | • It has a balance between fewer iterations and a performance [131].<br>• It offers a near-optimal performance of the ZF algorithm under correlated and uncorrelated channels [100], [131].<br>• It offers a good direction for searching process for the NSA algorithm [100], [131].<br>• It improves the convergence rate of the NSA algorithm [100], [100].<br>• It has a more accurate inversion result than the conventional NSA algorithm [100], [131].<br>• It has a high probability of convergence [100], [131].<br>• It has a smaller spectral radius [100].<br>• It can offer a better BER with a fewer number of iterations than the other existing iterative algorithms [100], [131].<br>• Wide range of convergence [100]. | • It has a lower convergence rate than the CI-SOR-AMI, NI-SOR-AMI, and SORMI algorithms [131].<br>• It converges only when the ratio of $(M/N) > 5.83)$ [131]. |
| WNS-SD | • It has a high hardware efficiency [121].<br>• It offers a near-optimal performance of the ZF algorithm [121].<br>• It can accelerate the convergence rate with a high probability when the ratio of $(M/N)$ is small [121].<br>• It offers a high parallelism and low complexity [121].<br>• It meets a strict latency requirement in massive MIMO systems [121].<br>• It offers a wide range of convergence [121].<br>• At first stage of the WNS algorithm, the SD algorithm offers an efficient convergence direction [121]. | • It needs additional calculations to find a suitable optimized weights.<br>• The **G** matrix must be a symmetric positive definite [121]. |
| AWNS-SD | • It needs just one iteration to reach a near-optimal performance of the ZF algorithm [121].<br>• It offers a high parallelism and low complexity [121].<br>• It meets a strict latency requirement in massive MIMO systems [121].<br>• It offers a wide range of convergence [121].<br>• It quit improves the convergence rate of the WNS-SD algorithm [121].<br>• Its convergence rate overcomes the other competitive algorithms like the NSA, WNS-SD, and JI-SD algorithms, while preserves lower complexity [121]. | • The **G** matrix must be a symmetric positive definite [121]. |
| JI-SD | • It offers a well direction for searching process for the JI algorithm [121].<br>• It significantly speeds the convergence rate over the JI algorithm [121].<br>• It has a higher speed of convergence rate over the NSA, SD, CG, and JI algorithms [134]. | • It can not converge when the ratio of $(M/N)$ is small [121]. |
| NI-NSA | • It speeds up the convergence rate of the NSA algorithm [118], [118].<br>• It offers a good direction for the searching process for the NSA algorithm [118].<br>• It can offer a very high convergence probability [118], [118].<br>• It has a two-order of convergence [100], [118]. | • Its convergence rate is lower than the convergence rate of the CG, GS, NSA, and CI-NSA algorithms [118], [131].<br>• It converges with high probability only when the ratio of $(M/N) > 5.83)$ [118], [118]. |
| JI-CG | • It offers an accurate direction of searching for iterations of the JI algorithm [134].<br>• It has a better BER performance than the NSA, GC, JI, and JI-SD algorithms with a faster convergence rate [134].<br>• It has a simple hardware implementation [134].<br>• It has a reasonable latency [134]. | • The **G** matrix must be a symmetric positive definite [134]. |

get the **a** symbols [162]. Based on the concept of SIC, the elements of pre-signal vector $\tilde{\mathbf{x}}$ could be represented as

$$\tilde{x}_n = \mathrm{MOD}_M\left(a_n - \sum_{l=1}^{n-1} b_{nl}\tilde{x}_{nl}\right),\ n = 1, 2, ..., N \quad (34)$$

where $b_{nl}$ is the element of the matrix **B** and $\mathrm{MOD}_M$ is a modular arithmetic which can be designated as

$$\begin{aligned}\mathrm{MOD}_M(x) = x - 2\sqrt{M}\left\lfloor \frac{1}{2} - \Re\left\{\frac{x}{2\sqrt{M}}\right\}\right\rfloor \\ - \left\lfloor \frac{1}{2} + \Im\left\{\frac{x}{2\sqrt{M}}\right\}\right\rfloor.\end{aligned} \quad (35)$$

Utilization of the modulo operation in the TH algorithm causes some losses of performance as shaping, modulo, and power losses [163], [164]. The loss of modulo could be

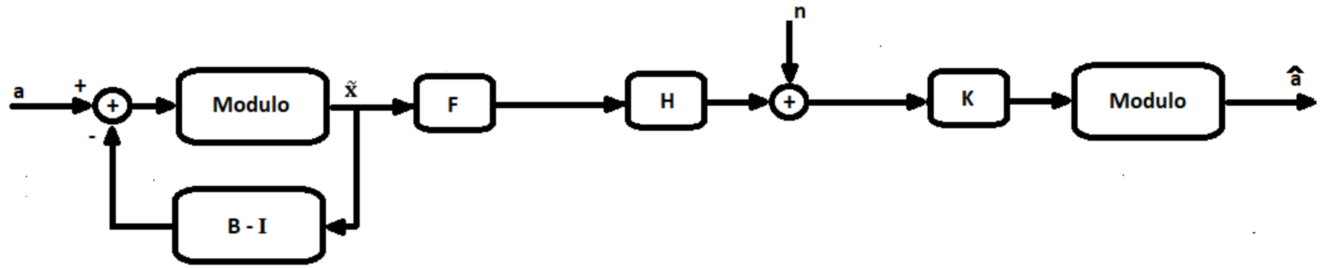


FIGURE 6: The block diagram of the TH algorithm based on LQ decomposition.

avoided by exploiting the large constellation with a suitable decoding algorithm , or by modifying the input signal by adding a perturbation vector [163]. The TH precoder algorithm has a higher computational complexity compared with linear precoding algorithms, where it mainly has three processing steps precancellation matrix, LQ decomposition of $\mathbf{H^T}$, and the feedforward matrix. Furthermore, the use of the TH algorithm with limited CSI is ambiguous [162]. However, utilization of the TH algorithm in massive MIMO systems has two main challenges: the high computational complexity and the instability due to the CSI inaccuracies [162]. In [165], a low-complexity TH algorithm is proposed which obtains a performance near to the conventional TH algorithm performance. Particularly, the proposed algorithm has the same computational complexity of the RZF algorithm for calculating the feedforward and feedback matrix. Per contra, the proposed algorithm may still have a high computational complexity, especially in massive MIMO systems.

In [166], a hybrid low-complexity algorithm which combines the linear and TH algorithms (HL-TH) is proposed. In the HL-TH algorithm, users are classified into groups to diminish the computational complexity, where each group has a lower size of efficacious channels than the realistic channels. The proposed algorithm has two steps. The first step is minimizing the inter-group interference by simply exploiting CSI in an inner linear precoder. In the second step, a TH precoding algorithm successively removes the intra-group interference in each group. The result of simulation shows that the HL-TH algorithm has a higher BER for each user compared to the RZF algorithm and has a substantially lower computational complexity than the low-complexity TH algorithm proposed in [165].

In [162], a novel algorithm for the DL massive MU-MIMO systems is proposed to address the challenges of the conventional TH algorithm. The proposed algorithm is a special case of the HL-TH algorithm but does not rely on grouping and without assuming the users' channel correlation matrices. Furthermore, the proposed algorithm utilizes conventional linear algorithms with a TH algorithm to offer a lower complexity implementation, and offer more resilience for diverse kinds of CSI while maintaining a substantial performance rate. The proposed algorithm increases the sum-rate over linear precoders with delayed and limited CSI and has robustness with smaller size arrays and different propagation conditions.

3) Vector Perturbation (VP) Precoding:

The VP algorithm is proposed in [167] to present an easy encoding technique without explicating the dirty-paper techniques, and viewed as a generalized TH algorithm [167]. The VP algorithm offers a full diversity order with a much lower complexity compared with the DPC algorithm [168]. The VP algorithm offers a near-capacity performance and regularizes a variation on the inversion process, where exploits a sphere encoder to perturb the input data to mitigate the transmitted energy, after that the vector of perturbed data is precoded by a linear front-end precoder [167]–[170]. The transmitter in the VP algorithm chooses the precoding matrix to relieve the IUI and after that finds the perturbation vector according to the criterion of minimizing the unscaled transmitted power [167]. The pseudo-inverse of $\mathbf{H}$ matrix and its regularized version can be used as the generator matrix of the lattice [167].

The perturbation operation in the VP algorithm needs the linear front-end precoding process to select the perturbing vector of the signal to be sent to all the users, which indicates that these two processes are required to be done jointly [167]. Whereas the TH algorithm chooses the scalar integer offset sequentially to be used in the transmitter and it does not execute nearly as well as the VP algorithm selection. The VP algorithm is modifying the transmitted data, instead of modifying the inverse process, by aligning the data symbols at the transmitter to the eigenvalues of the inverse $\mathbf{H}$ matrix on an instantaneous basis [167]. This modifying can be done by discreetly inserting a scalar integer vector offset at the transmitter and that leads to an interference cancellation at the receiver by applying a modulo arithmetic operation. In the VP algorithm, the CSI is assumed to be perfectly known at the transmitter, and each receiver requires only a single pre-arranged scalar which is regarding to the channel SNR [171].

The VP algorithm has the unpretentious interpretation of placement of the largest ingredients of the signal along with the lowest singular values of the inverse channel, and the smallest signal ingredients of the signal along the highest singular values [167]. However, a sphere encoding technique can be utilized in selecting the required vector perturbation [172], [173]. Figure 7 shows the block diagram of the VP algorithm and for simplicity assuming $M$ is equal to $N$, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



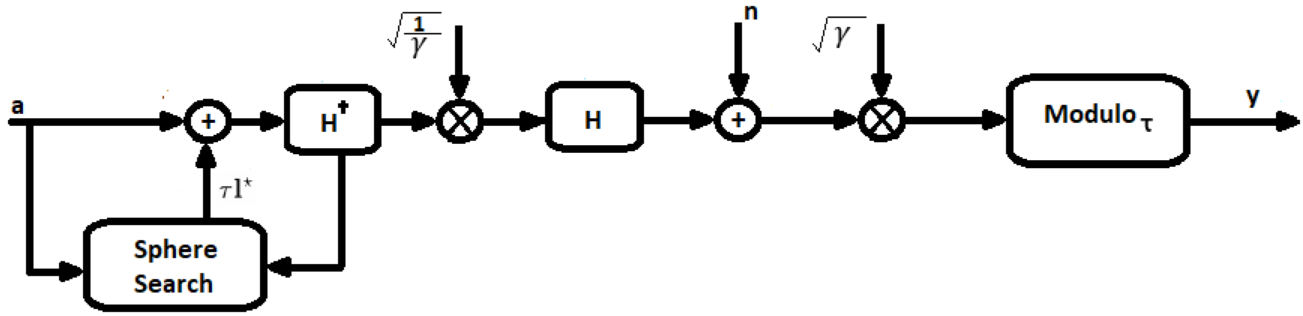


FIGURE 7: Block diagram of the VP algorithm. [174].

the perturbed data $\tilde{\mathbf{a}}$ vector can be represented as [174]

$$\tilde{\mathbf{a}} = \mathbf{a} + \tau \mathbf{l}^*, \tag{36}$$

where $\mathbf{l}^*$ is the $N \times 1$ selected perturbation vector with complex integer entries, and $\tau$ is the absolute value of the maximum magnitude of the constellation symbol which is equal to [169], [172], [174]

$$\tau = 2\left|c\right|_{max} + \Delta, \tag{37}$$

where $\Delta$ is the minimum Euclidean distance of the constellation [169], [172]. Depend on the VP-ZF as an example, the transmitted signal $\mathbf{x}$ can be represented as [174]

$$\mathbf{x} = \frac{1}{\sqrt{\gamma}} \mathbf{H}^* \left(\mathbf{H}^{\mathbf{T}} \mathbf{H}^*\right)^{-1} \tilde{\mathbf{a}}, \tag{38}$$

where $\gamma$ is the scaling factor of transmit power, and is equal to [169], [172], [174]

$$\gamma = \left\| \mathbf{H}^* \left(\mathbf{H}^{\mathbf{T}} \mathbf{H}^*\right)^{-1} \tilde{\mathbf{a}} \right\|_F^2. \tag{39}$$

After that, the signal symbol vector at the receiver is scaled-back and a modulo operator is employed to clear the perturbation affect $\tau \mathbf{l}^*$. The output of the receiver $\mathbf{y}$ is represented as [169], [172], [174]

$$\begin{aligned} \mathbf{y} &= \mathrm{mod}_\tau \left[ \sqrt{\gamma} \left( \frac{1}{\sqrt{\gamma}} (\mathbf{a} + \tau \mathbf{l}^*) + \mathbf{n} \right) \right] \\ &= \mathrm{mod}_\tau \left[ \mathbf{a} + \tau \mathbf{l}^* + \sqrt{\gamma} \mathbf{n} \right], \end{aligned} \tag{40}$$

and with ignoring the effect of $\mathbf{n}$, the received vector $\mathbf{y}$ becomes [169], [170], [174],

$$\mathbf{y} = \mathbf{a} + \mathbf{n}. \tag{41}$$

From (40), $\sqrt{\gamma}$ has the prime role in the perturbation vector $\mathbf{l}^*$ design. The performance of the system degrades significantly with a large value of $\gamma$ [98]. Thus, the value of $\gamma$ is minimized to find the best perturbation vector $\mathbf{l}^*$ as [98], [169], [170], [174]

$$\mathbf{l}^* = \arg\min_{\mathbf{l}} \left\| \mathbf{H}^* \left(\mathbf{H}^{\mathbf{T}} \mathbf{H}^*\right)^{-1} (\mathbf{a} + \tau \mathbf{l}^*) \right\|_F^2. \tag{42}$$

The optimization process in (42) is a $2N$-dimensional real integer lattice problem (NP-hard problem). The sphere search algorithms are commonly exploited to do this minimization process where the computational complexity increases exponentially with $N$.

In [175], a block diagonalized vector perturbation (BD-VP) algorithm that joins the BD and VP algorithms for the MU-MIMO system is proposed. The BD-VP algorithm avoids the needing for a global CSI like in the BD algorithm, and hence diminishes the receiver complexity of each user. It has a comparable performance of diversity to the BD algorithm. However, the BD-VP algorithm has rather high complexity at the transmitter side due to the combination of BD and VB algorithms. In [176], a low-complexity BD-VP algorithm and a user grouping vector perturbation (UG-VP) are proposed to enhance the performance of the BD-VP algorithm. However, the BD-VP and UG-VP algorithms are sub-optimal algorithms where the perturbation process is done independently for each user/group. To clarify the performance loss, the authors in [169] proposed a new joint VP algorithm (JVP) that achieves a considerable performance with the conventional VP algorithm and can be exploited in the adaptive modulation system.

In [177], a novel VP precoding algorithm assisted by reactive tabu search (RTS) for the large-scale MU-MIMO systems is proposed where the RTS is an iterative local neighborhood search technique. The proposed algorithm has the ability to efficiently flee from penurious local minima and achieves a quasi-optimal performance with a significant complexity reduction compared to the conventional VP algorithm.

In [178], a novel thresholded VP (TVP) algorithm is proposed to offer a tradeoff between performance and complexity for the VP algorithm in small and large scale MU-MIMO systems. A threshold for specific performance is exploited to decrease the sphere search process within the perturbation vectors. Once the threshold is achieved, the searching process should be terminated, and hence, considerable complexity is obtained. The sum-rate achieved by the proposed algorithm is approximately 90% of the conventional VP sum-rate, with at lower than 50% of the computational complexity.

In [172], a novel VP algorithm with limited feedback is proposed for MU-MIMO systems. The proposed algorithm avoids the extensive high complexity of the sphere searching in the conventional VP algorithm by employing a Min-Max

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



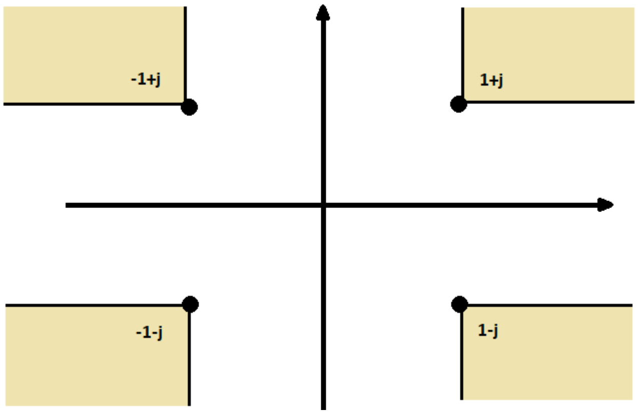


FIGURE 8: The search areas of CVP algorithm for 4-QAM.

optimization. The proposed algorithm dodges the scaling process at the receiver by constraining the searching area of perturbing vectors to the symbol constellation areas, as denoted in Fig. 8 by the shaded areas, which are constructive to the data symbol. Therefore, this algorithm is known as a constructive vector perturbation (CVP). Dodging a scaling process at the receiver is the main advantage in the CVP algorithm, especially in the limited feedback scenarios.

In [170], a novel hybrid TH and VP algorithm (TH-VP) is proposed. The proposed algorithm employs the TH algorithm to remove the IUI. It also employs the VP algorithm in equalizing each diversified spatial stream of the user. In the TH-VP algorithm, the two non-linear algorithms can be combined in one optimization process to obtain a low computational complexity and a satisfactory performance in comparison with other competitive algorithms such as the VP, ZF-VP and BD-VB algorithms. The TH-VP performance overcomes the ZF-VP and BD-VB algorithms, and is close to the performance of the DPC algorithm.

In [171], a robust VP algorithm is proposed, which can jointly deals with the CSI imperfections and inaccurate power-scaled factors under the criterion of the MMSE algorithm. The proposed VP algorithm is less sensitive to CSI imperfections and inaccurate power-scaled factors in contrast to the conventional VP algorithm.

4) Lattice Reduction Aided (LR) Precoding:

Lattice theory is firstly exploited in detection and precoding techniques for MIMO systems in [179]. The main concept of the LR algorithm is exploiting the discrete behavior of the digital information and deal with $\mathbf{H}$ as a basis of a point lattice [180]. In the LR algorithm, there are typically three steps [180]:

- Reducing the basis for the lattice by employing a uni-modular matrix.
- Solving precoding problem related to a reduced basis.
- Transforming back the solution into the original domain by the uni-modular matrix.

There are sundry definitions of the LR algorithm with corresponding reduction criteria, like the LLL reduction [181], [182], the Seysen reduction (SR) [183], the Brun reduction (BR) [184], the Korkine-Zolotareff reduction (KZ) [185], [186], the Minkowski reduction (MR) [187], and the Gauss reduction (GR) [188]. The selecting of the optimal perturbed vector brings about the problem of searching the closest point within a lattice [189], which can be done by sphere-searching [167]. This optimal solution is computationally too costly. Solving precoding problems related to the reduced basis displays advantages in performance and complexity. In [168], [190]–[192], some of sub-optimum precoding techniques can offer a full diversity when preceded by the LLL lattice reduction. However, employing the LR algorithm in the precoding process is unlike the LR-aided data detection, where it can avoid the shaping problem, e.g., the relaxation from a finite to an infinite lattice [193].

Assuming $N \leqslant M$, ZF-based precoding, and $\mathbf{O}$ is a right pseudo-inverse matrix which equals to

$$\mathbf{O} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathbf{T}}\mathbf{H}^{*}\right)^{-1} \tag{43}$$

the corresponding reduced matrix $\tilde{\mathbf{O}}$ is given by

$$\tilde{\mathbf{O}} = \mathbf{O}\mathbf{T}, \tag{44}$$

where $\mathbf{T}$ is an uni-modular transformation matrix. Thus, the (42) can be reformulated in terms of $\mathbf{O}$ and $\mathbf{T}$ as

$$\begin{aligned}\left\|\mathbf{O}\left(\mathbf{a}+\tau\mathbf{l}^{*}\right)\right\|^{2} &= \left\|\mathbf{O}\,\mathbf{T}\,\mathbf{T}^{-1}\left(\mathbf{a}+\tau\mathbf{l}^{*}\right)\right\|^{2}\\ &= \left\|\tilde{\mathbf{O}}\left(\tilde{\mathbf{a}}+\tau\tilde{\mathbf{l}}^{*}\right)\right\|^{2},\end{aligned} \tag{45}$$

where $\tilde{\mathbf{a}} = \mathbf{T}^{-1}\mathbf{a}$, and $\tilde{\mathbf{l}}^{*} = \mathbf{T}^{-1}\mathbf{l}^{*}$ [190], [194], [195]. The LR algorithm can be also exploited with the conventional linear precoding algorithm. In [196], the jointly LR-SR aided the ZF or MMSE linear precoding algorithms are proposed (see Fig. 9). Utilization of the SR technique offers a more orthogonal basis than that offered by the LLL technique. The SR technique achieves 0.5 dB in BER performance at $10^{-5}$ over the LLL performance. The computational complexity of the SR technique is about 92% of the LLL technique for $4{\times}4$ MIMO system with 4 QAM [197], [198]. In [199], the LR-aided linear precoding algorithm is employed in the BD algorithm in lieu of the second SVD process to parallelize each stream of users. Also, a complex LLL (CLLL) algorithm is employed in the LR algorithm to reduce the computational complexity of the LLL algorithm by approximately 50% without any effect in performance [200].

In [201], the employing of the LR algorithm in the VP algorithm based on a multi-branch (MB) strategy (MB-LR-VP) for MU-MIMO systems is proposed. The MB builds a collection of branches for transmitting information streams depending on a pre-infectious ordering scheme. Besides that, a development of an efficient scheme to construct the transmit ordering patterns is also proposed to find an optimal selection mechanism. The MB-LR-VP algorithm offers a better BER performance than the conventional VP precoding algorithm.

In [202], a joint LR and VP algorithm is proposed for massive MU-MIMO systems to offer reduced BER at all

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



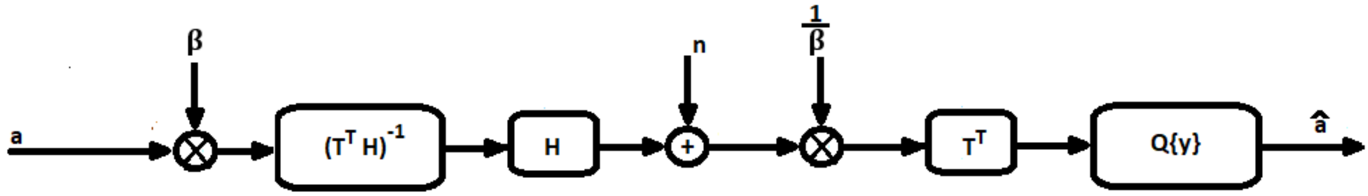


FIGURE 9: The block diagram of LR aided linear precoding using SR algorithm.

TABLE 5: Computational Complexity of Some Non-linear Precoding Algorithms.

| Algorithm | Computational complexity | Notes |
|---|---|---|
| TH | $8\left(\frac{N}{M}+\frac{2}{3}\right)N^3+2\left(2T\left(1+2\frac{N}{M}\right)-1\right)N^2+2\left(T\left(2-\frac{N}{M}\right)+2\right)N-8T$ [166]. | Where $T$ is the channel coherence time. |
| HL-TH | $4\left(4\frac{N}{MG}+\frac{10}{3G^2}\right)N^3+\left(2\frac{N}{M}(4T-1)+4\frac{(T-1)}{G}\right)N^2+\left(4-\frac{N}{M}\right)N-8TG$ [166] | Where $T$ is the channel coherence time, and $G$ is the number of user groups. |
| VP-sphere searching | $\mathcal{O}(N^6)$ [169] | - |
| ZF-VP | $4\frac{M^3}{3}+4M^2+2M$ [170] | - |
| BD-VP | $\frac{4}{3}M^3+\sum_{k=1}^{N}\mathcal{O}(n_k^6)$ [169] | This complexity is for MU-MIMO system where each user has $n_k$ antennas. On the other hand, the complexity for MU-MIMO system where each user has a single antenna can be easily obtained as $\frac{4}{3}M^3+\mathcal{O}(N^6)$. |
| UG-VP | $\frac{4}{3}M^3+\sum_{g=1}^{G}\mathcal{O}(n_g^6)$ [169] | Where $G$ is the number of user groups, and $n_g$ is the number of user in each group. |
| JVP | $\frac{4}{3}M^3+\mathcal{O}(M^6)+2M$ [169] | - |
| Low complexity VP | $\frac{4}{3}M^3+\sum_{k=1}^{N}\left[n_k^4\mathcal{O}(\log(\omega))\right]$ [176] | • This complexity is for an MU-MIMO system where each user has $n_k$ antennas. On the other hand, the complexity for MU-MIMO system where each user has single antenna can be easily obtained as $\frac{4}{3}M^3+\left[N^4\mathcal{O}(\log(\omega))\right]$.<br>• $\omega$ is the norm of the longest basis of **D** matrix of the BD algorithm. |
| VP-RTS | $\mathcal{O}(N^2M)$ [177] | - |
| LR-LLL | $\mathcal{O}(N^4)$ [201] | - |
| LR-CLLL | $8\left(4M^2N+8MN^2+9N^3\right)$ [204] | - |

SNRs with no lowering in capacity. Two types of the LR-VP algorithm have been proposed, namely LR-VP-ZF and LR-VP-MMSE algorithms. The LR-VP algorithms have better BER and higher capacity than the BD, TH and VP precoding algorithms for massive MU-MIMO systems.

In [203], a comparison between the LR and TH algorithms is presented. The authors deduced that the TH algorithm always overcomes the LR algorithm in a well-conditioned channel, whilst the LR algorithm is outshined in an ill-conditioned channel. The optimization process for stream ordering makes the TH algorithm triumph in a broad range number of channels, essentially in the large-scale MIMO systems.

For smooth readability and comparison, the computational complexity of the non-linear precoding algorithms is presented in Table 5 and their pros and cons are reviewed in Table 6.

### C. PEAK-TO-AVERAGE POWER RATIO (PAPR) PRECODING

It is clear that a MIMO system with a large number of antennas will considerably boost hardware cost and power consumption when costly linear power amplifiers are employed [212]–[215]. Thus, it was necessary to find a practical way to implement massive MIMO systems. An efficient non-linear power amplifiers can be exploited to offer a practical implementation of massive MIMO systems [212]–[215]. Thus, the PAPR should be minimized to relieve the effect of amplifier non-linearities. In this sub-section, the precoding algorithms, which aim to reduce the PAPR, will be reviewed.

1) Constant Envelope (CE) Precoding:

The CE precoding algorithm for massive MU-MIMO sys-

TABLE 6: Pros and Cons of the Non-linear Precoding Algorithms.

| Algorithm | Pros | Cons |
|---|---|---|
| DPC | • It offers the sum-rate capacity, where its capacity is the same as the capacity when there is no interference [154], [155].<br>• It offers the best performance than all other precoding algorithms [154].<br>• It eliminates the known-interference at the transmitter and obtains free-interference output [154]. | • It demands a high computational complexity and that leads to unfavorable level of complexity especially in massive MIMO systems [127], [155]. |
| TH | • It has a close-to-capacity performance [166].<br>• It can efficaciously eschew the noise amplification [205].<br>• It can recompense the interference through multiple antennas and multiple users in the MIMO systems [205].<br>• It employs the modulo operation to offer ISI canceling [205].<br>• It employs the modulo operation to limit the transmitted power at the transmitter, and to void the non-linear distortion at the receiver [206].<br>• It acts as spatial equalization [207].<br>• It has a promising performance, by employing SIC of the data streams at the transmitter, where it overcomes other linear precoding algorithms performance [208], [209].<br>• It has a practical implementation in contrast with the DPC algorithm [98]. | • It is more expensive and complex compared to linear precoding algorithms [205].<br>• Its computational complexity may also be too high for a medium or large ratio of $(M/N)$ [166].<br>• It has a sensitivity to the CSI inaccuracies [209].<br>• It may increase the transmitted power at the transmitter due to employing the SIC process [98].<br>• The use of the modulo operation leads to some losses of performance as shaping, modulo, and power losses [163], [164].<br>• It sequentially chooses its scalar integer offset, and it does not execute nearly as well as the jointly VP algorithm vector selection [167].<br>• It suffers some diversity penalty [195]. |
| VP | • It improves the channel inversion performance and offers near-capacity performance [169].<br>• It has a simple encoding technique without explicating the dirty-paper techniques [167].<br>• It minimizes transmit power by perturbing the transmit signal vector by another vector [175].<br>• It finds the optimal perturbation vector by solving a minimum distance type problem [175].<br>• It much improves transmit scaling factors and enhanced the received SNRs in contrast to linear precoding [167].<br>• It offers a full diversity order with much lower complexity compared with the DPC algorithm [168].<br>• It mitigates the IUI [167]. | • It needs to a computationally terrible sphere search out of different candidate perturbation vectors to reduce the precoded signal norm [174].<br>• Its power-scaled factors in the transmitter are data-dependent, which leads to a considerable overhead in transmission to feedforward the instantaneous power-scaled factors to the receiver [174].<br>• It has a dynamic range of received signals related to the power-scaled factor [174].<br>• It can not employ the adaptive modulation because to the constant modulo operation base $\tau$ [169].<br>• Its performance degrades significantly in the limited feedback scenarios [173].<br>• It suffers from the CSI imperfections and inaccurate power-scaled factors [171]. |
| LR | • It significantly diminishes the searching complexity of the benchmark sphere encoding [203].<br>• It outperforms the performance of the TH algorithm in the ill-conditioned channels [203].<br>• It is vastly implemented in practical designs [201].<br>• It diminishes the average transmitted power by amending the region which contains the constellation points [168].<br>• It can achieve the optimum asymptotic slope of symbol error rate [168].<br>• It can offer a full diversity when preceded by the LLL lattice reduction technique [190], [191], [192].<br>• It can offer high performance with a low computational complexity [180].<br>• It does not experience the shaping problem [180].<br>• It orthogonalizes the columns of H and reduces its size [210]. | • It needs a perfect CSI at the transmitter [211].<br>• It has relatively high complexity with corresponding reduction criteria, like the LLL technique [199].<br>• Its degree of H columns orthogonlization depends on the corresponding reduction criteria [196]. |

tems was proposed by Mohammed and Larsson in [212], [213], [216] to minimize the PAPR of the transmit signal. The CE algorithm has inexpensive and highly power-efficient amplifiers. For a given sum-rate and a large value of $M$, it can reduce the total transmit power by about 4 dB in contrast to an algorithm that employs highly linear power-inefficient amplifiers. In the CE algorithm, the array power gain is still offered on certain mild channel conditions, and the sum capacity of the average-only total transmit power-constrained channel can be offered [213].

Figure 10 shows the transmission process of data signals **a** by the CE constraint [212], [217], [218]. Where, the transmitted signal from each antenna is

$$x_m = \sqrt{\frac{P}{M}} e^{j\theta_m} \quad \text{for } m = 0, 1, ..., M, \tag{46}$$

where $P$ is the total transmitted power, and $\theta_m$ is the phase angles of transmitted signals and exemplifies the precoding phase of the CE precoder signals. Then, the received signal at the $n^{\text{th}}$ user can be represented as [217]

$$y_n = \sum_{m=1}^{M} h_{n,m} \sqrt{\frac{P}{M}} e^{j\theta_m} + n_n \quad \text{for } n = 0, 1, ..., N, \tag{47}$$

where $h_{n,m}$ is the channel coefficient between the $m^{\text{th}}$ antenna and the $n^{\text{th}}$ user, and $n_n$ is the zero mean AWGN vector of the $n^{\text{th}}$ user. For simplicity, the (47) can be rewritten as [217]

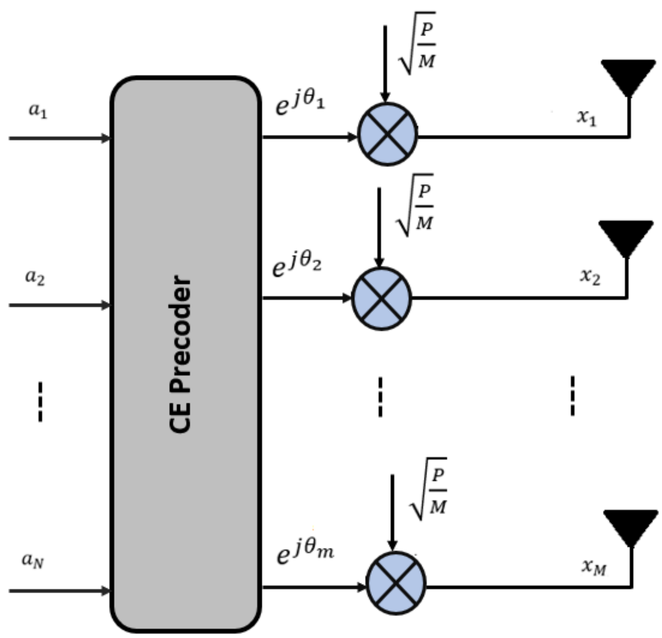


FIGURE 10: The transmission strategy of data signals by the CE algorithm for the DL massive MU-MIMO systems.

$$y_n = a_n + t_n + n_n \quad \text{for } n = 0, 1, ..., N, \tag{48}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



where $a_n = g_n e^{j\Phi_m}$ is the PSK data symbol for the $n^{\text{th}}$ user, and $t_n$ is the interfering signal for the $n^{\text{th}}$ user

$$t_n = \left( \sum_{m=1}^{M} h_{n,m} \sqrt{\frac{P}{M}} e^{j\theta_m} - g_n e^{j\Phi_n} \right). \quad (49)$$

The total MUI energy can be represented as [213], [217]

$$E_{MUI} = \sum_{n=1}^{N} \left| \left( \sum_{m=1}^{M} h_{n,m} \sqrt{\frac{P}{M}} e^{j\theta_m} - g_n e^{j\Phi_n} \right) \right|^2. \quad (50)$$

The phase angles of transmitted signals $\{\theta_1, \theta_2, ..., \theta_M\}$ are optimized to minimize (50) [212], [219]. Therefore, the CE precoding algorithm can be designed as follows [213], [217]

$$\mathfrak{p}_{CE}: \ \min_{\theta} \left\{ \sum_{n=1}^{N} \left| \left( \sum_{m=1}^{M} h_{n,m} \sqrt{\frac{P}{M}} e^{j\theta_m} - g_n e^{j\Phi_n} \right) \right|^2 \right\}, \\ \text{subject to} \quad |\theta_m| \leqslant \pi, \ \forall m \in \{1, 2, ..., M\}. \quad (51)$$

In the CE algorithm, the transmitted signals are restricted by a constant amplitude, and the MUI at all users are minimized by optimizing the non-linear least squares (NLS) problem of the transmit signals phase angles (51) [212], [213], [217]. Unfortunately, the conventional CE precoding algorithm depends on gradient descent (GD) method to solve this non-convex NLS problem which converges to a local minimum and suffers from a slow convergence rate [213], [214], [219], [220].

The authors in [213] deal with the NLS problem in the CE algorithm by proposing the sequential gradient descent (SGD) search algorithm, but the SGD algorithm is often stuck in the local minima when the number of $M$ is not large enough to suppress the MUI [214]. In [221], the annulus constrained (AC) precoding algorithm is proposed to relax the lower-bound amplitude constraints of the CE algorithm. The AC algorithm offers an additional degree of freedom by allowing amplitudes to alter within a pre-specified interval to further enhance the performance of the MUI suppression, but it increases the transmit signal PAPR. The AC algorithm employs the SGD algorithm to search for optimal precoding weights. However, the AC precoding algorithm is higher computational complexity than the SGD algorithm because it has a larger searching space to find optimal precoding weights.

In [220], a novel algorithm is proposed to enhance the MUI suppression ability of the CE algorithm by employing a cross-entropy optimization (CEO) algorithm. However, this algorithm has a higher computational complexity than the SGD algorithm. In [222], another novel algorithm known as Riemannian conjugate gradient (RCG) algorithm, is proposed to enhance the MUI suppression ability of the CE algorithm by dealing with the feasible region of NLS problem in the CE algorithm as a manifold of a complex circle. However, the RCG algorithm has a remarkable lower computational complexity than the SGD and CEO algorithms.

In [223], the impact of phase-angle constraints at the BS is investigated. Where more restrictions on the transmitted phases are utilized at different symbol times. These restrictions can be led to an increase in the system power efficiency. While the above CE algorithms describe only the interference minimization problem, prior researches on linear precoding [167], [224], [225] prove that the interference minimization does not certainly offer the preferable performance in a MIMO system. In general, as the interference depends on data, the transmitter has the capability to prophesy the MUI at the receiver and can exploit that knowledge to impact it and profit from it [226]. In precocious researches [227], [228], precoders are proposed to reduce the negative results of interference while conserving its positive components. Whereas these interference effects are defined depending on the correlation between the sub-streams of a phase shift keying MIMO system.

In [226],the authors proposed a novel CE precoding algorithm for massive MIMO systems with the concept of constructive interference (CIN) which offers a remarkable performance enhancement compared to interference reduction algorithms [229]–[231]. In [219], a manifold-based algorithm to solve the CIN problem of the CE precoding, by using the RCG algorithm to find a local minimizer, is proposed. The proposed CE-CIN-RCG precoding algorithm views the feasible region of NLS problem in the CE algorithm as an oblique manifold (OM). The precoded symbols by the CE-CIN-RCG algorithm is perfectly constant envelopes, in contrast to the relaxed convex problem in the algorithm proposed in [226]. Furthermore, the CE-CIN-RCG algorithm has superior symbol error rate (SER) performance and lower computational complexity than the CE-CIN algorithm proposed in [226].

In [232], the continuous-time CE (CTCE) precoding algorithm is proposed for the DL massive MU-MIMO systems. The CTCE transmits signals are transmitted to different users from arbitrary constellations simultaneously. The CTCE algorithm needs about 3 dB more radiated power above the traditional linear precoders at low sum-rates. The CTCE algorithm has less consumed power than the traditional linear precoders due to more power efficient hardware designs, where it dos not require the hardware linearity of the BS and power amplifiers operated at maximum efficiency.

In [233], the performance of the CE algorithm is compared with the corresponding performance of the ZF algorithm for a large scale MU-MIMO system. The achievable SINR of the CE algorithm overcomes the ZF algorithm by about 5-6 dB when the power amplifier works in the saturation region.

2) Approximate Message Passing (AMP) Precoding:

Solving the non-convex NLS problem of the CE algorithms is computationally intractable. Inspired by the amazing performance of the AMP algorithm in a rapid inference in the related of compressed sensing topic [235], [235]–[237], the AMP precoding algorithm for massive MU-MIMO systems is proposed in [214] to offer a practical solution for the CE precoding problem without the need to find a global optimal solution of non-convex problem in a computationally difficult method. In addition, the AMP algorithm can offer a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



TABLE 7: Computational Complexity of the PAPR Precoding Algorithms.

| Algorithm | Computational complexity | Notes |
|---|---|---|
| CE-GD | $\mathcal{O}(M^2N)$ [213], [219]. | This complexity is for each iteration. |
| CE-SGD | $\mathcal{O}(MN)$ [213] | This complexity is for each iteration. |
| CE-AC | $\mathcal{O}(2M^2N)$ [214] | This complexity is for each iteration. |
| CE-CEO | $\mathcal{O}(kMN)$ [219] | • This complexity is for each iteration.<br>• $k$ is the number of random samples in each iteration.<br>• $k$ is usually quit larger than $M$ and $N$. |
| CE-RCG | $\mathcal{O}(MN)$ [219] | This complexity is for each iteration. |
| CE-CIN-RCG | $\mathcal{O}(MN)$ [219] | This complexity is for each iteration. |
| AMP | $7MN + 9M + 2N$ [214] | This complexity is for each iteration. |
| One bit precoder | $\mathcal{O}(4^M)$ [155] | - |
| One bit-SDR | $\mathcal{O}((4M+1)^{4.5})$ [234] | - |
| One bit-SQUID | $\mathcal{O}(2(k_1M^3 + k_2M^2))$ [234] | $k_1$ is the iteration numbers of the first loop, and $k_2$ is the iteration numbers of the second loop |
| One bit-ADMM | $\mathcal{O}(M^2)$ [234] | - |

trade-off between computational complexity and achievable performance.

The AMP algorithm converts the precoding problem of massive MU-MIMO systems to a probabilistic inference problem by recasting the non-convex NLS problem of the CE algorithms into an estimation problem. To design the AMP algorithm which can estimate the transmitted vector $\mathbf{x}$ based on the received vector $\mathbf{a}$, the following virtual model is introduced [214], [218]

$$\mathbf{a} = \mathbf{H}^{\mathbf{T}}\mathbf{x}. \tag{52}$$

The conditional probability distribution function (pdf) of an anonymous signal $\mathbf{x}$ given the knowledge of matrix $\mathbf{H}^{\mathbf{T}}$ and the observations of $\mathbf{a}$ can be acquired by employing the Bayes' rule [238] and depending on (52) as

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{H}^{\mathbf{T}}, \mathbf{a}) &\propto p(\mathbf{a} \mid \mathbf{H}^{\mathbf{T}}, \mathbf{x})p(\mathbf{x}) \\ &\propto \underbrace{\left[\prod_{n=1}^{N} p(a_n \mid \mathbf{H}^{\mathbf{T}}, \mathbf{x})\right]}_{\text{the likelihood function}} \underbrace{\left[\prod_{m=1}^{M} p(x_m)\right]}_{\text{the prior pdf}}, \end{aligned} \tag{53}$$

where $\propto$ indicates the identity after normalization to unity. The (53) can be represented by using the factor graph [239], as shown in Fig. 11, where the likelihood function consists of a product of $N$ factors relative to the constraint over each $a_n$ and the prior pdf consists of a product of $M$ factors relative to what is expected of each $x_m$ [214], [218]. The factor graph consists of $M + N$ factor nodes and $M$ variable nodes. The factor graph takes apart the estimation problem into a number of simple mutually local problems. The local problems are solved by constrained local nodes in an interactively and parallel manner [218]. Then the signal components which contain the pdf of the signal component are described based on a belief propagation (BP) [218].

The BP consists of two kinds of messages: first the messages from variable nodes to factor nodes, $msg_{m\to n}(x_m)$, and second the messages from factor nodes to variable nodes, $msg_{n\to m}(x_m)$ [214], [218]. The messages are updated through the local nodes by the AMP algorithm to iteratively optimize the precoding design for massive MU-MIMO systems. The AMP algorithm overcomes the CE-GD and CE-AC algorithms in relation to the desired complex multiplication numbers and the required time to converge. Furthermore, the AMP algorithm can offer a trade-off between the PAPR reduction and MUI suppression with favorable results [214], [218]. In [218], a comparative study between the AMP algorithm, the CE algorithm, and the linear precoding algorithms is introduced. The AMP algorithm has a significant improvement in the BER performance than the CE algorithm and has comparable BER performance with the linear precoding performance.

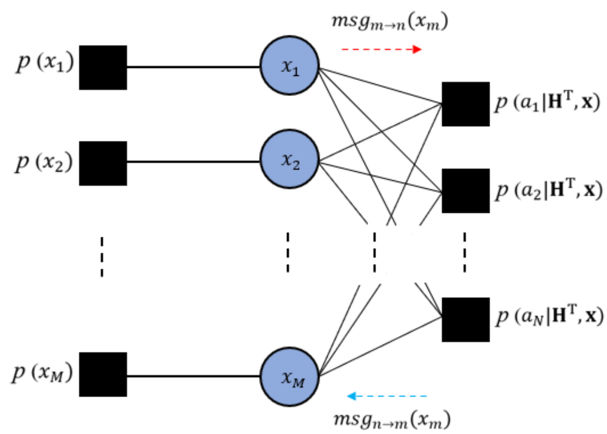


FIGURE 11: Factor graph representing the probabilistic model in (53), where the filled rectangles denotes to factor nodes, and the filled circles denotes to variable nodes.

3) Quantized Precoding (QP):

In order to exploit the favorable features of massive MIMO systems, one needs to take into account the related negative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



TABLE 8: Pros and Cons of the PAPR Precoding Algorithms.

| Algorithm | Pros | Cons |
|---|---|---|
| CE | • It can offer an $\mathcal{O}(M)$ array power gain on certain mild channel conditions [213].<br>• It can provide beamforming and spatial multiplexing gains which leads to improve the spectral and energy efficiencies [233].<br>• It allows reducing the overall transmitted power linearly with increasing the fixed data rate at each user [213].<br>• It minimizes the MUI [213].<br>• It offers near-optimal performance [213].<br>• It achieves a close to sum capacity of the MRT algorithm [212], [213].<br>• It eases the using of non-linear and high power-efficient RF components [213], [220].<br>• Its high tolerance to the non-linearities of power amplifier leads to basically a milder in-band distortion [233].<br>• It needs smaller transmitted power than the MRT algorithm by about 2 dB for sufficient large ratio of $M/N$ [213].<br>• It needs smaller transmitted power than a system employs highly linear power-inefficient amplifiers by about 4 dB for a sufficient large ratio of $M/N$ [213].<br>• It can overcome the ZF algorithm by about 5-6 dB [233]. | • It can not be exploited when the MUI is significantly high [213].<br>• It needs a sufficiently large ratio of $M/N$ [213], [220].<br>• It needs to solve an non-convex NLS problem which leads to a high complexity and slow convergence rate [213], [220]. |
| CE-GD | • It is exploited to solve the non-convex NLS problem of the CE algorithm [213]. | • It has a slow convergence rate [213].<br>• It has a high computational complexity [213].<br>• It describes only the interference minimization problem [219].<br>• It suffers from set in advance optimization because it strongly depends on the constellation energy [219].<br>• It has a lower performance than CI algorithms [219]. |
| CE-SGD | • It improves the performance of MUI suppression for the CE algorithm by using sequential GD method [213], [219]. | • It is often stuck in the local minima when the ratio of $M/N$ is not large enough [213], [220].<br>• It strongly depends on the selection of the initial guess [213], [220].<br>• It describes only the interference minimization problem [219].<br>• It suffers from set in advance optimization because it strongly depends on the constellation energy [219].<br>• It has a lower performance than the CIN algorithms [219]. |
| CE-AC | • It improves the performance of the MUI suppression for the CE algorithm by relaxing the amplitudes constraints lower band [214].<br>• It offers an additional degree of freedom by allowing amplitudes to change within the pre-determined interval [214]. | • It has a higher computational complexity than the SGD algorithm [214].<br>• It has a faster convergence rate than the CE-GD algorithm [214].<br>• It increases the transmitted signal PAPR [214].<br>• It describes only the interference minimization problem [219].<br>• It suffers from the set in advance optimization because it strongly depends on the constellation energy [219].<br>• It has a lower performance than the CIN algorithms [219]. |
| CE-CEO | • It further improves the performance of the MUI suppression for the CE algorithm by employing CEO method [213], [220].<br>• It offers a better MUI suppression performance than the SGD algorithm [214], [220]. | • It has a higher computational complexity than the SGD algorithm [220].<br>• It is not suitable for practical implementation [214], [220].<br>• It describes only the interference minimization problem [219].<br>• It suffers from set in advance optimization because it strongly depends on the constellation energy [219].<br>• It has a lower performance than the CIN algorithms [219]. |
| CE-RCG | • It further improves the performance of the MUI suppression for the CE algorithm by viewing the feasible region of the NLS problem as a manifold of complex circle [219].<br>• It has the fastest convergence rate of the CE algorithms [219].<br>• It has much lower complexity than both the CE-GD and CE-CEO algorithms. | • It describes only the interference minimization problem [219].<br>• It suffers from set in advance optimization because it strongly depends on the constellation energy [219].<br>• It has a lower performance than the CIN algorithms [219]. |
| CE-CIN-RCG | • It offers near-optimal performance [219].<br>• It offers a remarkable enhancement compared to the RCG, CEO, AC, SGD, and GD based algorithms [219].<br>• It enhances the CE algorithm by viewing the feasible region of the NLS problem as an oblique manifold [219].<br>• It has perfectly constant envelops in contrast to the relax-convex problem of CE-CIN algorithm, which proposed in [219], [226].<br>• It has a higher SER performance and lower computational complexity than the CE-CIN algorithm, which proposed in [219], [226]. | • It has a lower convergence rate than the RCG algorithm [219]. |
| AMP | • It offers a practical solution for the CE problem without the need to find globally optimal solution of non-convex problem [214].<br>• It has a practical implementation for massive MU-MIMO systems [214], [218].<br>• Its computational complexity increases linearly with the number of antennas in the BS, and that supports it in massive MU-MIMO systems [218].<br>• It offers a trade-off between PAPR reduction and MUI suppression with favorable results [214], [218].<br>• It has a parallel nature [214].<br>• It offers a faster convergence rate and fewer complex multiplication process than the CE-GD and CE-AC algorithms [214].<br>• It has a remarkable higher BER performance than the CE algorithm and has comparable BER performance with the linear precoding performance [218]. | • It can slightly overcome the AC precoding algorithm at the initial stage only, but eventually the CE-AC algorithm overcomes it [214]. |
| One-bit precoder [155] | • It greatly diminishes the power consumption.<br>• It facilitates the industrial design of the QP algorithm.<br>• It is vastly implemented in practical designs.<br>• It has a significant enhancement in performance over the linear QP algorithm. | • It needs a huge number of antennas in the BS to reach the performance of ideal infinite-resolution performance.<br>• It suffers from the coarse quantization which leads to heavy performance loss especially with a high order of modulation.<br>• It has a very high computational complexity. |
| One-bit-SDR | • It offers near-optimal solution for the QP problem [155].<br>• Its BER performance is robust in small and large-size of MU-MIMO systems for a wide SNR range [155].<br>• It considers as a benchmark for the QP performance [234].<br>• It avoids the non-convex constraint in the original precoding problem by solving the relaxed versions of the non-convex problem [234].<br>• It is robust against the CSI imperfections [155]. | • It has a high computational complexity [234].<br>• It is not practical in the massive MIMO systems, with hundreds number of antennas [234].<br>• Its run-time is higher than the run-time of the proposed precoder in [234], which is based on the ADMM framework, by approximately 380 times, and than the run-time of the SQUID precoder by approximately 10 times. |
| One-bit-SQUID | • It has a comparable performance with the SDR precoder in large-size MU-MIMO systems [203].<br>• It has a lower computational complexity than the SDR precoder [234].<br>• It can be carefully employed in the massive MIMO systems [234].<br>• It is based on the ADMM framework [155], [234].<br>• It is robust against the CSI imperfections [155]. | • Its iteration procedure has double loops, where the inner loop is necessary to solve $\ell_\infty$-norm operator [234].<br>• Its run-time is higher than the run-time of the proposed precoder in [234], which is based on the ADMM framework, by approximately 3.9 times. |
| Proposed precoder in [234] | • It offers state-of-the-art BER performance comparable to the SDR precoder while preserving the advantages of low complexity.<br>• It can be employed in the massive MIMO systems.<br>• It is faster than the SDR algorithm by 300 times.<br>• It is more efficient than the SQUID precoder, where it has a single loop in its iteration procedure.<br>• It has a global convergent.<br>• It has a significant reduction in the run-time compared to the SDR and SQUID precoders.<br>• It is based on the ADMM framework.<br>• It can serve a high order of modulation system, e,g, QPSK, 16-QAM, and 64-QAM. | • It has a significant higher computational complexity in contrast to the linear QP algorithms.<br>• It still needs more antennas in high order modulations, e.g. 64-QAM. |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



effects caused by employing a large number of antennas in the BS. A high number of RF chains at the BS leads to a considerable increase in the complexity of hardware, costs, and power consumption [155]. One of the main causes of power consumption in massive MIMO systems is the data converters at the transmitter [155]. In the DL transmission, each RF chain generates the transmit baseband signal by using a pair of digital-to-analog converters (DACs) [155]. These DACs experience a power consumption that increase with the resolution (in bits) exponentially and with the bandwidth linearly [240], [241]. In conventional MIMO systems, RF chains employ high resolution DACs ($\leqslant$ 10 bits) [155], and consume approximately 40-50 % of the overall operational power consumption [213]. Subsequently, the resolution of DACs must be restricted to save power consumption within reasonable levels. In [242], a linear QP based on the MMSE algorithm by paying attention to the DACs distortion is proposed. This proposed algorithm exploits DACs with 4 to 6 bits resolution and overcomes the traditional linear QP algorithms for moderate-size MIMO systems at high SNR.

The more restricted one-bit DAC precoding is mainly dependent on the well-known Bussgang theorem [243] and it is a special case of constant-envelope (CE) algorithm, where the transmitted signal phase is restricted to only four diverse values [213], [216].

In [244], a quantized MRT precoding algorithm with one-bit DACs, which leads to manageable distortion levels, is described for Massive MU-MIMO systems. Furthermore, in [245], the performance of the quantized ZF precoding algorithm with one-bit DACs is analyzed on a Rayleigh-fading channel. The authors in [245] show that the intense distortion for each antenna resulted from one-bit DACs can be averaged out when a large number of transmit antennas are employed. Moreover in [246], a comparative study of using ideal DACs and one-bit DACs in massive MU-MIMO systems. The one-bit DACs performance loss can be compensated by deducting about 2.5 times more antennas of the BS.

In [155], the issue of DL precoding for massive MU-MIMO systems on frequency-flat channels with low-resolution DACs at the BS is investigated. The authors in [155] considered both quantized linear precoder, where a precoder is succeeded by a finite resolution DAC, and non-linear precoder where the outputs of DAC are directly generating by jointly using the data vector with the CSI. The performance of the MRT and ZF linear precoders according to a coarse quantization is also analyzed. The performance of infinite resolution DACs can be achieved by using 3 to 4 bits DACs. The authors in [155] also proposed a new non-linear precoder with one-bit DACs which overcomes linear precoders but with a cost of computational complexity. The performance of the proposed non-linear precoder is less than the performance in the infinite resolution case by 3 dB for $10^{-3}$ uncoded BER, with 128 BS antennas and 16 single-antenna users. Where the performance of linear precoders is less than the performance in the infinite resolution case by 8 dB. Figure 12 previews the QP for the DL massive MU-MIMO systems with low-resolution DACs. Where $\mathbf{x} \in \mathcal{X}^M$ is the precoded vector and $\mathcal{X}$ is the set of complex numbers $\mathbb{C}$ when DACs has infinite resolution [155].

In practical MIMO architectures with finite-resolution DACs, the quantization labels can be defined as [155]

$$\mathcal{L} = \{\ell_0, \ell_1, ..., \ell_L\}, \tag{54}$$

where

$$\ell_i = \alpha\triangle\left(i - \frac{L-1}{2}\right), \qquad i = 0, 1, 2, ..., L-1, \tag{55}$$

where $\triangle$ is a step size of symmetric uniform quantizers, $\alpha$ is a scaled power factor, and $L$ is the number of quantization levels.

The quantizer-mapping function of the one-bit DACs can be presented as [155]

$$\mathbf{x} = \mathcal{Q}\left(\mathbf{Pa}\right) = \sqrt{\frac{P}{2M}}\left(\operatorname{sgn}\left(\Re\left\{\mathbf{Pa}\right\}\right) + j\operatorname{sgn}\left(\Im\left\{\mathbf{Pa}\right\}\right)\right), \tag{56}$$

where $P$ is the average power, $\mathbf{P}$ is the precoding matrix, and $\mathcal{Q}(.) : \mathbb{C}^M \to \mathcal{X}^M$ is the non-linear quantizer-mapping function which characterizes the joint process of the $2M$ DACs at the transmitter.

The finite quantization outputs of $\mathcal{X}$ are [155]

$$\mathcal{X} = \left\{\sqrt{\frac{P}{2M}}\left(\pm 1 \pm j\right)\right\}. \tag{57}$$

The one-bit QR problem can be formulated as follows [155]

$$\begin{aligned} &\text{minimize}_{\mathbf{x}\in\mathcal{X}^M,\ \beta\in\mathbb{R}} \left\|\mathbf{a} - \beta\mathbf{H}^{\mathbf{T}}\mathbf{x}\right\|_2^2 + \beta^2 N\sigma_n^2, \\ &\text{subject to} \quad \beta > 0, \end{aligned} \tag{58}$$

where $\beta$ is a precoding factor. To solve the QP problem in (58) at a fixed value of $\beta$, the evaluation of $|\mathcal{X}|^M = 4^M$ vectors is required, where the computational complexity grows exponentially with $M$. In recent literature review, there is a variety of developing low-complexity non-linear precoders which offer near-optimal performance for the one-bit QP problem, such as semidefinite relaxation (SDR) [155], [247]–[249], squared-infinity norm Douglas-Rachford splitting (SQUID) [155], [250], [251], adaptation sphere precoding (ASP) [155], [252]. Where the relaxed versions of the QP problem can be solved by these non-linear algorithms.

In [234], a highly efficient non-linear precoding algorithm, for massive MU-MIMO systems with one-bit DACs based on an alternative direction method of multipliers (ADMM) framework is proposed. The ADMM based algorithm solves the original non-convex precoding problem directly instead of solving relaxed versions problem. The ADMM based algorithm offers a performance of the SDR algorithm, and it is faster than the SDR algorithm by 300 times.

For smooth readability and comparison, the computational complexity of the PAPR precoding algorithms is presented in Table 7 and their pros and cons are comprehensively reviewed in Table 8.

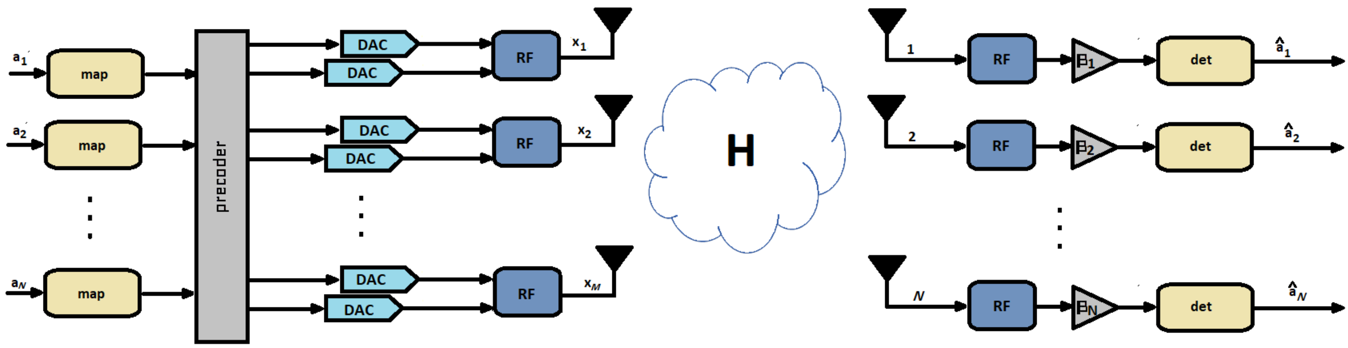


FIGURE 12: Previewing of the QP for the DL massive MU-MIMO systems with low resolution DACs [155].

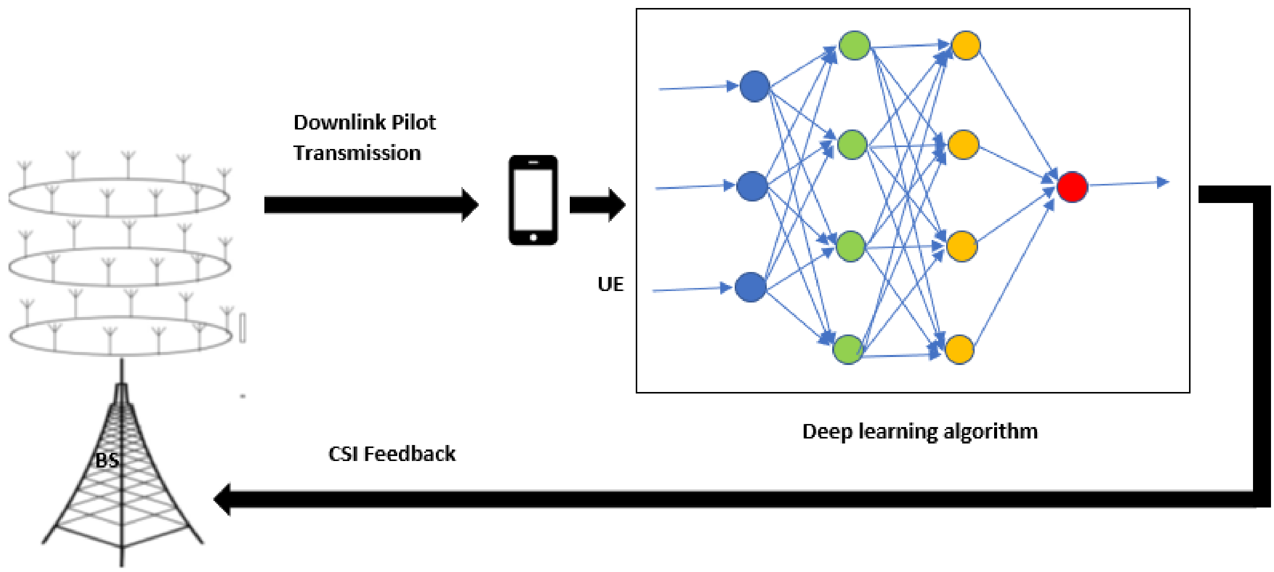


FIGURE 13: Channel estimation for precoded massive MIMO with FDD utilizing machine learning in reducing the feedback overhead.

### D. MACHINE LEARNING PRECODING

Machine learning algorithms have lately shown a great prospective in treating complex optimization problems in the emerging wireless communications applications and settings [253]. For instance, there is a plethora of detection techniques for massive MIMO based on machine learning [254]–[256]. There is also an extensive literature that is dealing with deep learning utilization precoded mmWave MIMO systems [257]. For example, the authors in [258] have introduced a precoded massive MIMO system that is using FDD. As mention earlier, the CSI should be known at the transmitter for the precoding process. Unlike, TDD which takes advantage of channel reciprocity in CSI estimation, FDD needs feedback to the BS with the CSI. The multiuser channel estimation and feedback problem have been considered as a distributed source coding (DSC) problem which is solved using generalized deep neural network (DNN) architecture (see Fig. 13).

The overhead problem of FDD CSI is also tackled in [259], [260] for precoded massive MIMO in conventional frequency bands. A two-stage decoder is a common solution to reduce the size of the feedback information [261]. An outer decoder reduces the large channel dimensions to be ready to be used by the inner decoder to control the IUI similar to multiuser MIMO precoding. To this end, a DNN architecture is developed in the outer decoder to optimize the channel dimensions. The DNN machine-learning model is evaluated and proved to enhance the average sum-rate and achieves near-optimal performance.

The deep learning algorithms are also used to enhance the problem of SIC used in massive MIMO-NOMA systems [259]. SIC in massive MIMO suffers from imperfections especially when multiple user's real-world scenarios is considered. A joint optimization for both MIMO-NOMA precoding and SIC is done by minimizing the total mean square error of the users' signals. The superior performance and effective-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



ness of the proposed scheme are demonstrated through the numerical results. In [262], deep learning tools are exploited to optimize the biConvex 1-bit precoding algorithm where per-iteration parameters are introduced. The algorithm is tested in different channel models and shows satisfactory results in vastly changing propagation conditions.

## VI. PRECODING IN PROMISING ANTENNA ARCHITECTURES FOR B5G

The 5G cellular technology was implemented by several mobile carriers. It reduces the data connections latency and increases the data rate. However, inter-cell interference and handover issues are remaining to limit the cell-edge performance. In addition, a large number of antennas at the massive MIMO transceiver causes an extra computational and implementation complexity. The design of energy-efficient and sustainable communication systems is also still an issue to be handled in B5G. Therefore, three promising multiple antenna technologies/architectures are flashing up in B5G networks: the CF-M-MIMO, beamspace massive MIMO, and the IRSs [263].

### A. CELL-FREE MASSIVE MIMO

In CF-M-MIMO, the concepts of distributed MIMO and massive MIMO are combined with no cell-boundaries, and hence, intercell interference is mitigated [264]. The UEs are simultaneously served by a large number of service antennas (access points (APs)) that distributed over a wide geographic area. In CF-M-MIMO, there is a central processing units (CPU), but the information exchange between the CPU and the APs is very limited to the payload data. The ZF precoding scheme is utilized in many CF-M-MIMO systems. However, it requires an exchange of the instantaneous CSI among all APs which complicates the processing when a large number of APs is utilized [265]. In [266], [267], a conjugate beamforming, ZF precoding scheme, and max-min power control were utilized with the DL CF-M-MIMO to guarantee a good service at a high spectral efficiency. Centralized ZF precoding was implemented in part of APs, while the maximum-ratio-transmission (MRT) was applied in other APs.

In [268], it has shown that increasing the number of antennas per AP results in a stronger hardening effect. However, this paper has used the MRT without taking into account neither the pilot contamination nor the imperfect CSI. It also omits the effect of power allocation strategies on the channel hardening. In [269], it is shown that the degree of hardening is highly affected by the power allocation coefficients, and hence, it affects the specific precoding scheme. The work in [269] has been analyzed in [270] with accounting for different realistic channel model assumptions and different system configurations where ZF precoding scheme was exploited.

In [271], the high power consumption in CF-M-MIMO was tackled by a low complexity power control technique with ZF precoding scheme to maximize the energy efficiency of CF-M-MIMO when imperfect CSI is used. In order to avoid instantaneous CSI exchange, local partial ZF, and local protective partial ZF schemes were proposed to provide an interference cancellation gain and improve the spectral efficiency of the CF-M-MIMO system. Compared to the traditional ZF and MRT schemes, local partial ZF and local protective partial ZF schemes have significantly enhanced the spectral efficiency of the CF-M-MIMO system [272]. Unfortunately, a large number of antennas at each AP is required in local partial ZF schemes which is more challenging for UL design [273]. In [274], a partial MMSE is proposed as a precoding scheme. Although it is nearly optimal, the power allocation for distributed operation was not investigated.

### B. BEAMSPACE MASSIVE MIMO

A low complexity realization can be achieved by exploiting the spatial structure of the channels and transceiver hardware can be utilized without sacrificing the operational flexibility or the performance. By employing designed discrete lens array (DLA), the conventional channel in the spatial domain can be converted to the beamspace channel [275]. The more antennas that are used in a M-MIMO transceiver, there is a need to rethinking the signal processing and linear precoding where beamspace MIMO formulation is one of the most popular approach. It is expected that the number of antennas will continue to increase in sub-6 GHz communication networks. Hence, the dimensionality of M-MIMO arrays with hybrid digital-analog, tiled arrays and sub-arraying is going to be impractical. Therefore, utilization of subspace approach based on effective channels will benefit the massive MIMO processing. In case of high frequencies (i.e. millimeter wave (mmWave)), beamspace will be mandatory [263]. Beamspace MIMO can significantly reduces the number of power RF chains in mmWave communications [276].

The beamspace channel is sparse, hence, the dimension of the massive MIMO can be reduced by selecting a small number of powerful beams [277]. In [278], a precoding scheme in beamspace was proposed where the spatial channel sparsity was exploited. The beamspace techniques have a great impact in accomplishing a satisfactory precoding with a low complexity. In [276], optimal hybrid cross-entropy (HCE) based hybrid precoding scheme and lens array architecture were utilized to propose a feasible precoding scheme. The proposed architecture has achieve a satisfactory performance and a high energy efficiency. In [277], an optimal HCE based hybrid precoding with machine learning was proposed. The probability distribution of the hybrid precoder is updated by minimizing the cross-entropy. Numerical results show that the proposed scheme can achieve a satisfactory performance with high energy efficiency. Figure 14 illustrates the concepts of beamspace MIMO and hybrid precoding architecture. In [279], the beamspace channel sparsity was exploited for the training of the deep neural network. A deep learning compressed sensing channel estimation and hybrid precoding were considered and the network was trained offline to predict the beamspace channel amplitude. Then, a deep learning quantized phase hybrid precoding method was developed to

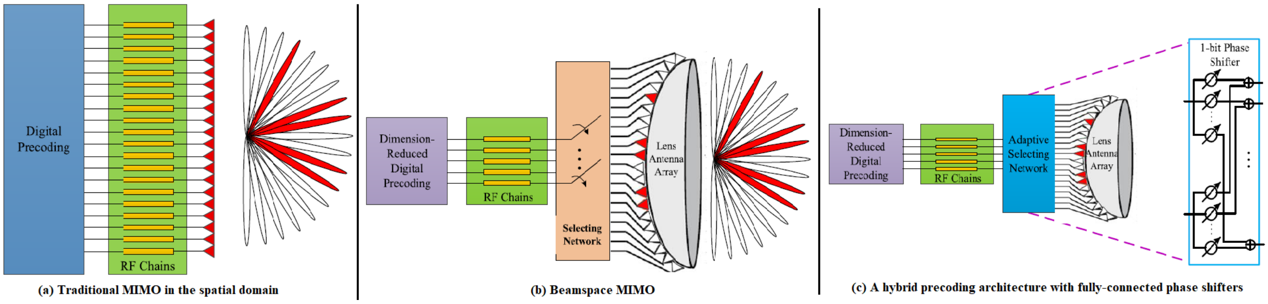


FIGURE 14: Traditional MIMO in the spatial domain, beamspace MIMO, and hybrid precoding architecture [277].

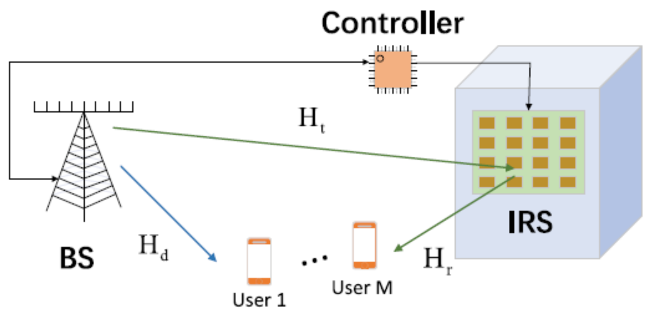


FIGURE 15: A communication system aided with the IRS [280].

obtain a satisfactory spectral efficiency.

### C. INTELLIGENT REFLECTING SURFACES

In last few years, IRS has gained a great attention in the research community and considered as a promising solution to provide new degrees of freedom and hence, improve the spectral and energy efficiency (bit/Joule) of B5G communication networks with low hardware cost. In the IRS, a large set of low-cost elements are composed in a metasurface (a planar array) to diffusely reflects incoming signals in a smart controllable manner. In other words, with a pre-programmed phase shifts and/or reflecting amplitudes, each element is able to independently reflect the incoming signal. It is similar to the concept of reconfigurable reflect arrays with real-time control and reconfigurability [263]. Figure 15 shows an IRS-aided multi-user multi-input-single-output (MU-MISO) communication system.

In [281], [282], a symbo-level precoding (SLP) with the assistance of IRS was proposed. In order to minimize the worst-case symbol error probability at the BS, phase shift at the IRS is proposed. Numerical results have shown that the performance can be significantly improved by incorporating the IRS. In [280], a precoding design of IRS-aided communication system was proposed where the multi-user interference (MUI) was exploited to improve the performance. Alternating optimization (AO) algorithm was utilized to obtain the precoding matrix and phase shift matrix. In [283], the precoding matrix of the BS and the reflection coefficients of the IRS were optimized by the block coordinate descent (BCD). In addition, the minorization-maximization (MM) algorithm was utilized to reduce the computational complexity.In [284], the Riemannian conjugate gradient, branch-and-bound, and direct quantization were used to attain a low-resolution SLE precoder for single-antenna users. For multi-antenna receivers, decomposition of the large scale optimization problem was applied to decompose the original problem into several sub-problems.

### D. OTHER MASSIVE MIMO PRECODERS

Full-duplex (FD) radios were incorporated into advanced technology such as the massive MIMO [285]. In addition, beam domain transmission has been introduced in a massive MIMO system to further enhance the spectral efficiency. In [286], the precoding for FD massive MIMO systems was comprehensively discussed. A beam-domain FD (BDFD) based on the basis expansion model is proposed to make the co-time co-frequency UL and DL transmission possible. Intelligent scheduling of beam-domain distribution is utilized to mitigate the self-interference (SI) and improve the transmission efficiency. This method has used the time-frequency efficiently and hence, the spectral efficiency gain is improved. Although numerical results show the superiority of the BDFD scheme over the TDD/FDD massive MIMO, the beamforming complexity should be taken into consideration, particularly in high speed railways (HSR) scenarios [287].

In [288], a beam extraction method is utilized to eliminate the pilot contamination using a secure beam-domain transmission scheme. In [289], a beam domain hybrid time switching (TS) and power splitting (PS) simultaneous wireless information and power transfer (SWIPT) system for FD massive MIMO is proposed for energy harvesting and channel estimation. The SI is eliminated without using instantaneous SI CSI. Numerical results show that the proposed hybrid system can achieve a considerable transmission efficiency gain. In [287], a hybrid beamforming and an angle-domain (AD) channel tracking schemes were proposed for HSR scenarios where the channel has been decomposed into spatial angular information and beam gain. In order to reduce the computational complexity, a beam-domain precoding scheme is utilized in the hybrid beamforming.

In [290], the optimal jamming precoding with a power solution in DL massive MIMO was investigated where ap-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access



proximate orthogonal beam domain channels of the BS were considered. The correlation of the beam domain channels was also investigated. The ZF precoding has achieved an optimal configuration for jamming defense. It is noteworthy that the power consumption of the precoding in the defense against a jammer was reduced by a proper increase in the channel approximation error [288].

## VII. OPEN RESEARCH AREA AND CONCLUSION

### A. OPEN RESEARCH AREA

Most research efforts carried out on massive MIMO have so far focused mainly on linear precoding algorithms. As it has been illustrated throughout this survey, there are trials to find low complexity versions of the non-linear precoders such as the DPC and TH. However, there are great potentials in finding more efficient and high performance non-linear precoders with comparable complexity with linear precoders.

Moreover in [259], [260] and [261], the DNN is utilized in massive MIMO systems to solve the problem of multiuser channel estimation and feedback, tackle the overhead problem of FDD CSI, and enhance the SIC problem. Furthermore, there is an extensive literature that is dealing with the precoding problem of mmWave MIMO systems by employing the deep learning [257]. Nevertheless, the literature has a remarkable lack of employing the AI technology, in general, in the conventional sub-6 GHz massive MIMO systems. However, there are huge opportunities to employ the AI technologies such as machine learning and DNN to design a high-performance and low-complexity precoder. For example, the AI technologies can be used to find the optimal perturbation vector of the non-linear VP precoder instead of using the sphere search algorithms. In addition, it can be used to optimize the MUI in the CE precoding algorithm. Furthermore, it can be used to minimize the one-bit problem, and to do various optimization processes in the precoder algorithms. Also the machine learning can be exploited to choose the best algorithm to be applied instead of the best data estimation. Although the learning stage could severely increase the computational complexity, it can be performed off-line to obtain the optimal precoding algorithm. In addition, the employment of the virtual channel model (VCM) is a potential direction of new innovation for precoding algorithms in massive MIMO systems. Although most existing precoding techniques are proposed for centralized massive MIMO networks, they can be exploited in the CF-M-MIMO.

The research in a high altitude platform (HAP) massive MIMO to obtain an efficient RF precoder and a baseband precoder with limited RF chains is also still in its infancy where AI can be exploited to provide high efficiency.

### B. CONCLUSION

Massive MIMO provides a great improvement in user experience and mobile services. It will stay a competitive candidate in the next decade. However, significant research dedicated to the transmitter's design is proposed. This paper has surveyed the linear and non-linear precoding schemes that pertain to massive MIMO systems. Although linear precoders suffer from performance deterioration under certain scenarios, they still play a crucial role in the transmitter design due to their relative simplicity. In this paper, a comparison between different linear precoders is provided. In addition, an in-depth discussion on non-linear precoders with their performance-complexity profile is presented. It is shown that the non-linear precoders have a high computational complexity but they are promising to obtain a satisfactory performance. This paper also reviewed the potential of machine learning role in precoding algorithms. Moreover, this paper has reviewed the precoding schemes in CF-M-MIMO, beamspace massive MIMO, and the IRS technologies. Besides that, channel estimation, collection of CSI in TDD and FD, impact of the condition number, and energy efficiency have been discussed.

## REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.
[2] D. C. Araujo, T. Maksymyuk, A. L. de Almeida, T. Maciel, J. C. Mota, and M. Jo, "Massive MIMO: Survey and future research topics," *IET Commun.*, vol. 10, no. 15, pp. 1938–1946, Oct 2016.
[3] D. Mercer, "Global connected and IoT device forecast update," Strategy Analytic research services, Tech. Rep., May 2019.
[4] "Ericsson mobility report : On the pulse of the networked society," Ericsson, Tech. Rep. EAB-16:006659, July 2016.
[5] M. A. Albreem, "5G wireless communication systems: Vision and challenges," in *Proc. Int. Conf. on Computers, Commun., and Contr. Technol.* IEEE, 2015, pp. 493–497.
[6] S. L. Mohammed, M. H. Alsharif, S. K. Gharghan, I. Khan, and M. Albreem, "Robust hybrid beamforming scheme for millimeter-wave massive-MIMO 5G wireless networks," *Symmetry*, vol. 11, no. 11, p. 1424, 2019.
[7] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Commun. Inform. Th.*, vol. 11, no. 3-4, pp. 154–655, 2017.
[8] T. L. Marzetta and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.
[9] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11 727–11 758, Jun 2017.
[10] R. Ahmad, M. A. Asim, S. Z. Khan, and B. Singh, "Green IOT—issues and challenges," 2019.
[11] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-frequency mmWave massive MIMO channel measurements and characterization for 5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1591–1605, Apr 2017.
[12] S. Mumtaz, J. Rodriguez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. Academic Press, 2016.
[13] J. Yuan, H. Shan, A. Huang, T. Q. Quek, and Y.-D. Yao, "Massive machine-to-machine communications in cellular network: Distributed queueing random access meets MIMO," *IEEE Access*, vol. 5, pp. 2981–2993, Feb 2017.
[14] Y. Mehmood, N. Haider, W. Afzal, U. Younas, I. Rashid, and M. Imran, "Impact of massive MIMO systems on future M2M communication," in *Proc. IEEE Malaysia Int. Conf. Commun.* IEEE, Nov 2013, pp. 534–537.
[15] L. Foschini, T. Taleb, A. Corradi, and D. Bottazzi, "M2M-based metropolitan platform for IMS-enabled road traffic management in IOT," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 50–57, Nov 2011.
[16] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 2019.
[17] M. Pappa, C. Ramesh, and M. N. Kumar, "Performance comparison of massive MIMO and conventional MIMO using channel parameters," in

*Proc. Int. Conf. on Wireless Commun., Signal Process. and Networking*. IEEE, Mar 2017, pp. 1808–1812.

[18] M. A. Albreem, W. Salah, A. Kumar, M. H. Alsharif, A. H. Rambe, M. Jusoh, and A. N. Uwaechia, "Low complexity linear detectors for massive MIMO: A comparative study," *IEEE Access*, vol. 9, pp. 45 740–45 753, 2021.

[19] Q. Hu, M. Zhang, and R. Gao, "Key technologies in massive MIMO," in *ITM Web of Conf.*, vol. 17. EDP Sciences, Feb 2018, p. 01017.

[20] S. Shahabuddin, "MIMO detection and precoding architectures," Ph.D. dissertation, University of Oulu, June 2019.

[21] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, Dec 2018.

[22] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO—5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, Nov 2015.

[23] A. Acheampong, N. Martey, and D. A. Kumah, "A comprehensive study of optimal linear pre-coding schemes for a massive MU-MIMO downlink system: A survey," *Int. Joint Conf. on Neur. Net.*, vol. 32, no. 1, pp. 21–33, 2019.

[24] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, Dec 2017.

[25] P. Zhang, J. Chen, X. Yang, N. Ma, and Z. Zhang, "Recent research on massive MIMO propagation channels: A survey," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 22–29, Dec 2018.

[26] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep 2017.

[27] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62 367–62 414, Mar 2020.

[28] F. Wen, H. Wymeersch, B. Peng, W. P. Tay, H. C. So, and D. Yang, "A survey on 5G massive MIMO localization," *Digital Signal Process., Elsevier*, vol. 94, pp. 21–28, Nov 2019.

[29] L. Li and W. Meng, "Detection for uplink massive MIMO system: A survey," in *Int. Conf. Advanc. Hybrid Infor. Process.* Springer, Sep 2019, pp. 287–299.

[30] B. Rajarajeswarie and R. Sandanalakshmi, "A short survey: Applications of artificial intelligence in massive MIMO." IEEE, July 2020, pp. 1096–1100.

[31] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digital Signal Process., Elsevier*, vol. 94, pp. 3–20, June 2019.

[32] E. Ali, M. Ismail, R. Nordin, and N. F. Abdulah, "Beamforming techniques for massive MIMO systems in 5G: Overview, classification, and trends for future research," *Front. Informat. Tech. Elect. Eng., Springer*, vol. 18, no. 6, pp. 753–772, June 2017.

[33] Y. Hei, C. Zhang, W. Song, and Y. Kou, "Energy and spectral efficiency tradeoff in massive MIMO systems with multi-objective adaptive genetic algorithm," *Soft Comput., Springer*, vol. 23, no. 16, pp. 7163–7179, July 2019.

[34] R. Chataut and R. Akl, "Massive MIMO systems for 5G and beyond networks—overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, May 2020.

[35] J. Chen, H. Chen, H. Zhang, and F. Zhao, "Spectral-energy efficiency tradeoff in relay-aided massive MIMO cellular networks with pilot contamination," *IEEE Access*, vol. 4, pp. 5234–5242, July 2016.

[36] Q. He, L. Xiao, X. Zhong, and S. Zhou, "Increasing the sum-throughput of cells with a sectorization method for massive MIMO," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1827–1830, Oct 2014.

[37] M. Matalatala, M. Deruyck, E. Tanghe, L. Martens, and W. Joseph, "Optimal low-power design of a multicell multiuser massive MIMO system at 3.7 GHz for 5G wireless networks," *Wireless Commun. and Mobile Comput., Wiley*, vol. 2018, Oct.

[38] H. A. J. Alshamary, "Coherent and non-coherent data detection algorithms in massive MIMO," *Doctoral dissertation, University of Iowa*, May 2017.

[39] P. Popovski, Č. Stefanović, J. J. Nielsen, E. De Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug 2019.

[40] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[41] T. T. Do, E. Björnson, E. G. Larsson, and S. M. Razavizadeh, "Jamming-resistant receivers for the massive MIMO uplink," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 210–223, Jan 2017.

[42] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb 2013.

[43] S. Jin, X. Wang, Z. Li, K.-K. Wong, Y. Huang, and X. Tang, "On massive MIMO zero-forcing transceiver using time-shifted pilots," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 59–74, Jan 2016.

[44] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[45] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[46] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. European Sign. Proc. Conf.* IEEE, 2014, pp. 76–80.

[47] X. Wu, N. C. Beaulieu, and D. Liu, "On favorable propagation in massive MIMO systems and different antenna configurations," *IEEE Access*, vol. 5, pp. 5578–5593, 2017.

[48] L. You, J. Xiong, X. Yi, J. Wang, W. Wang, and X. Gao, "Energy efficiency optimization for downlink massive MIMO with statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2684–2698, 2020.

[49] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct 2014.

[50] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov 2010.

[51] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug 2011.

[52] T. Younas, J. Li, M. M. Tulu, H. M. Munir, and M. Liaqat, "Study of single-cell massive MIMO systems with channel aging and prediction," *Wireless Networks, Springer*, pp. 1–10, May 2019.

[53] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Cellular systems with many antennas: Large system analysis under pilot contamination," in *Proc. Annual Allerton Conf. Commun., Contr., Computing*. IEEE, Oct 2012, pp. 1220–1224.

[54] M. A. M. Moqbel, W. Wangdong, and A.-m. Z. Ali, "MIMO channel estimation using the LS and MMSE algorithm," *IOSR Journal of Electronics and Communication Engineering*, vol. 12, no. 1, pp. 13–22, Feb 2017.

[55] M. Medra, Y. Huang, W. Ma, and T. N. Davidson, "Low-complexity robust MISO downlink precoder design under imperfect CSI," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3237–3249, 2016.

[56] J. Nielsen, A. Karstensen, P. C. F. Eggers, E. De Carvalho, G. Steinböck, and M. Alm, "Precoding for TDD and FDD in measured massive mimo channels," *IEEE Access*, vol. 8, pp. 193 644–193 654, 2020.

[57] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, 2018.

[58] H. Q. Ngo, *Massive MIMO: Fundamentals and system designs*. Linköping University Electronic Press, 2015, vol. 1642.

[59] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, 2016, pp. 974–978.

[60] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[61] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access




[62] T. Chang, W. Ma, and C. Chi, "Worst-case robust multiuser transmit beamforming using semidefinite relaxation: duality and implications," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, 2011, pp. 1579–1583.

[63] A. Morsali and B. Champagne, "Robust hybrid analog/digital beamforming for uplink massive-mimo with imperfect CSI," in *Proc. IEEE Wireless Commun. and Networking Conf.*, 2019, pp. 1–6.

[64] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, 2016.

[65] M. Hanif, H. Yang, G. Boudreau, E. Sich, and H. Seyedmehdi, "Practical hybrid precoding for multi-user massive MIMO systems," in *Proc. Canadian Conf. Elect. and Comp. Eng..*, 2016, pp. 1–4.

[66] M. Rihan, T. Abed Soliman, C. Xu, L. Huang, and M. I. Dessouky, "Taxonomy and performance evaluation of hybrid beamforming for 5G and beyond systems," *IEEE Access*, vol. 8, pp. 74 605–74 626, 2020.

[67] F. Hu, K. Wang, H. Wu, and L. Jin, "An adaptive energy consumption optimization for massive MIMO systems with imperfect CSI," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2018, pp. 377–381.

[68] F. Rosas and C. Oberli, "Impact of the channel state information on the energy-efficiency of MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4156–4169, 2015.

[69] A. Zappone, P. Cao, and E. A. Jorswieck, "Energy efficiency optimization in relay-assisted mimo systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, 2014.

[70] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in ofdma systems with large numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3292–3304, 2012.

[71] R. S. Prabhu and B. Daneshrad, "Energy-efficient power loading for a mimo-svd system and its performance in flat fading," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–5.

[72] H. Ye, G. Lim, L. J. Cimini, and Z. Tan, "Energy-efficient scheduling and resource allocation in uplink OFDMA systems," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 439–442, 2015.

[73] A. Akbari, R. Hoshyar, and R. Tafazolli, "Energy-efficient resource allocation in wireless OFDMA systems," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2010, pp. 1731–1735.

[74] H. Ye, G. Lim, L. J. Cimini, and Z. Tan, "Energy-efficient resource allocation in uplink OFDMA systems under QoS constraints," in *Proc. IEEE Military Commun. Conf.*, 2013, pp. 424–428.

[75] K. Illanko, M. Naeem, A. Anpalagan, and D. Androutsos, "Frequency and power allocation for energy efficient OFDMA systems with proportional rate constraints," *IEEE Wireless Commun. Lett.*, vol. 3, no. 3, pp. 313–316, 2014.

[76] X. Xiao, X. Tao, Y. Jia, and J. Lu, "An energy-efficient hybrid structure with resource allocation in OFDMA networks," in *Proc. IEEE Wireless Commun. and Networking Conf.*, 2011, pp. 1466–1470.

[77] Z. Chu, W. Hao, P. Xiao, F. Zhou, D. Mi, Z. Zhu, and V. C. M. Leung, "Energy efficient hybrid precoding in heterogeneous networks with limited wireless backhaul capacity," in *Proc. IEEE Global Telecommun. Conf.*, 2018, pp. 1–7.

[78] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, 2010.

[79] Z. Hasan, G. Bansal, E. Hossain, and V. K. Bhargava, "Energy-efficient power allocation in ofdm-based cognitive radio systems: A risk-return model," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6078–6088, 2009.

[80] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink OFDMA," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 112–120, 2012.

[81] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE Global Telecommun. Conf.*, 2010, pp. 1–6.

[82] A.-A. Lu, X. Gao, Y. R. Zheng, and C. Xiao, "Low complexity polynomial expansion detector with deterministic equivalents of the moments of channel Gram matrix for massive MIMO uplink," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 586–600, Dec 2015.

[83] A. Elghariani and M. Zoltowski, "Successive interference cancellation for large-scale MIMO OFDM," in *IEEE Int. Conf. Elect. Informat. Tech.* IEEE, Oct 2015, pp. 657–661.

[84] Y.-S. Jeon, N. Lee, S.-N. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, July 2018.

[85] M. A. Albreem, "An efficient lattice sphere decoding technique for multi-carrier systems," *Wireless Personal Communications*, vol. 82, no. 3, pp. 1825–1831, 2015.

[86] M. A. Albreem and M. F. M. Salleh, "Regularized lattice sphere decoding for block data transmission systems," *Wireless Personal Communications*, vol. 82, no. 3, pp. 1833–1850, 2015.

[87] Y.-C. Liang, G. Pan, and Z. Bai, "Asymptotic performance of MMSE receivers for large systems using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4173–4190, Nov 2007.

[88] Y. Kim, J. H. Seo, H. M. Kim, and S. Kim, "Soft linear MMSE detection for coded MIMO systems," in *IEEE Asia-Pacific Conf. Commun.* IEEE, Mar 2013, pp. 657–660.

[89] X. Gao, L. Dai, Y. Hu, Z. Wang, and Z. Wang, "Matrix inversion-less signal detection using SOR method for uplink large-scale MIMO systems," *arXiv preprint arXiv:1507.04588*, Mar 2015.

[90] F. Rosario, F. A. Monteiro, and A. Rodrigues, "Fast matrix inversion updates for massive MIMO detection and precoding," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 75–79, Jan 2015.

[91] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "Conjugate gradient-based soft-output detection and precoding in massive MIMO systems," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Feb 2015.

[92] B. Yin, M. Wu, J. R. Cavallaro, and C. Studer, "VLSI design of large-scale soft-output MIMO detection using conjugate gradients," in *Proc. IEEE Int. Symp. on Circuits and Systems*, 2015, pp. 1498–1501.

[93] Z. Wu, C. Zhang, Y. Xue, S. Xu, and X. You, "Efficient architecture for soft-output massive MIMO detection with Gauss-Seidel method," in *Proc. IEEE Int. Symp. on Circuits and Systems*. IEEE, May 2016, pp. 1886–1889.

[94] R. Chataut, R. Akl, and U. K. Dey, "Least square regressor selection based detection for uplink 5G massive MIMO systems," in *IEEE Wireless Microwave Tech. Conf.* IEEE, Apr 2019, pp. 1–6.

[95] V.-K. Dinh, M.-T. Le, V.-D. Ngo, and C.-H. Ta, "PCA-aided linear precoding in massive MIMO systems with imperfect CSI," *Wireless Commun. and Mobile Comput., Wiley*, vol. 2020, Feb.

[96] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, June 2014.

[97] O. Simeone, U. Spagnolini, and Y. Bar-Ness, "Linear and non-linear precoding/decoding for MIMO systems using the fading correlation at the transmitter," in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.* IEEE, June 2003, pp. 6–10.

[98] Y. Chen, "Low complexity precoding schemes for massive MIMO systems," Ph.D. dissertation, Newcastle University, June 2019.

[99] Y. Liu, J. Liu, Q. Wu, Y. Zhang, and M. Jin, "A near-optimal iterative linear precoding with low complexity for massive MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 1105–1108, Apr 2019.

[100] X. Qiang, Y. Liu, Q. Feng, J. Liu, X. Ren, and M. Jin, "Approximative matrix inversion based linear precoding for massive MIMO systems," in *Int. Conf. Comput. Net. Commun.* IEEE, Feb 2020, pp. 950–955.

[101] B. M. Lee, "Simplified antenna group determination of RS overhead reduced massive MIMO for wireless sensor networks," *Sensors*, vol. 18, no. 1, p. 84, Jan 2018.

[102] Y.-G. Lim, C.-B. Chae, and G. Caire, "Performance analysis of massive MIMO for cell-boundary users," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6827–6842, July 2015.

[103] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. Annual Allerton Conf. Commun., Contr., Computing*. IEEE, Oct 2013, pp. 293–298.

[104] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, July 2005.

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

[106] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[107] L. Liu, G. Peng, and S. Wei, *Massive MIMO Detection Algorithm and VLSI Architecture*. Springer, Feb 2019.

[108] J. Speidel, *Introduction to Digital Communications*. Springer, Sep 2018.

[109] C. Zhang, Z. Li, L. Shen, F. Yan, M. Wu, and X. Wang, "A low-complexity massive MIMO precoding algorithm based on Chebyshev iteration," *IEEE Access*, vol. 5, pp. 22 545–22 551, Oct 2017.

[110] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 916–929, Mar 2014.

[111] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *Proc. IEEE Wireless Commun. and Networking Conf.* IEEE, Apr 2013, pp. 2710–2715.

[112] I. Al-Nahhal, M. Alghoniemy, O. Muta, and A. B. Abd El-Rahman, "Reduced complexity -best sphere decoding algorithms for ill-conditioned MIMO channels," in *IEEE Annual Cons. Commun. Netw. Conf.* IEEE, Jan, pp. 183–187.

[113] K. K.-C. Lee, Y.-H. Yang, and J.-W. Li, "A low-complexity AEPDF-assisted precoding scheme for massive MIMO systems with transmit antenna correlation," *J. Signal Process., Springer*, pp. 1–11, Jan 2020.

[114] A. Müller, "Random matrix analysis of future multi cell MU-MIMO networks.(analyse des réseaux multi-cellulaires multi-utilisateurs futurs par la théorie des matrices aléatoires)." Ph.D. dissertation, Supélec, Gif-sur-Yvette, Rennes, Metz, France, Nov 2014.

[115] H. Prabhu, O. Edfors, J. Rodrigues, L. Liu, and F. Rusek, "Hardware efficient approximative matrix inversion for linear pre-coding in massive MIMO," in *Proc. IEEE Int. Symp. on Circuits and Systems*. IEEE, June 2014, pp. 1700–1703.

[116] Y. Bai, Z. Liang, C. Zhai, Y. Xin, and W. Li, "Joint precoding using successive over-relaxation matrix inversion and Newton iteration for massive MIMO systems," in *Proc. Int. Conf. on Wireless Commun. Signl Process.* IEEE, Oct 2019, pp. 1–5.

[117] M. A. Albreem, A. A. El-Saleh, and M. Juntti, "Linear massive MIMO uplink detector based on joint Jacobi and Gauss-Seidel methods," in *IEEE Int. Conf. Design Reliable Commun. Net.* IEEE, Mar 2020, pp. 1–4.

[118] L. Shao and Y. Zu, "Joint Newton iteration and Neumann series method of convergence-accelerating matrix inversion approximation in linear precoding for massive MIMO systems," *Mathematical Problems in Engineering*, vol. 2016, May 2016.

[119] B. Kang, J.-H. Yoon, and J. Park, "Low-complexity massive MIMO detectors based on richardson method," *ETRI Journal*, vol. 39, no. 3, pp. 326–335, June 2017.

[120] B. Nagy, M. Elsabrouty, and S. Elramly, "Fast converging weighted Neumann series precoding for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 154–157, Oct 2017.

[121] Q. Deng, X. Liang, X. Wang, M. Huang, C. Dong, and Y. Zhang, "Fast converging iterative precoding for massive MIMO systems: An accelerated weighted Neumann series-steepest descent approach," *IEEE Access*, vol. 8, pp. 50 244–50 255, Mar 2020.

[122] C. Tang, C. Liu, L. Yuan, and Z. Xing, "High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 490–493, Jan 2016.

[123] Y. Lee and S.-I. Sou, "On improving Gauss-Seidel iteration for signal detection in uplink multiuser massive MIMO systems," in *Proc. IEEE Int. Conf. Comput. Commun. Sys.* IEEE, Apr 2018, pp. 268–272.

[124] J. Zeng, J. Lin, and Z. Wang, "An improved Gauss-Seidel algorithm and its efficient architecture for massive MIMO systems," *IEEE Trans. Circuits Syst. II*, vol. 65, no. 9, pp. 1194–1198, Feb 2018.

[125] L. Shao and Y. Zu, "Approaches of approximating matrix inversion for zero-forcing pre-coding in downlink massive MIMO systems," *Wireless Networks, Springer*, vol. 24, no. 7, pp. 2699–2704, Oct 2018.

[126] X. Gao, L. Dai, Y. Hu, Z. Wang, and Z. Wang, "Matrix inversion-less signal detection using SOR method for uplink large-scale MIMO systems," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Dec 2014, pp. 3291–3295.

[127] X. Qiao, Y. Zhang, and L. Yang, "Conjugate gradient method based linear precoding with low-complexity for massive MIMO systems," in *Proc. IEEE Int. Conf. Comput. Commun.* IEEE, Dec 2018, pp. 420–424.

[128] Z. Zhang, J. Wu, X. Ma, Y. Dong, Y. Wang, S. Chen, and X. Dai, "Reviews of recent progress on low-complexity linear detection via iterative algorithms for massive MIMO systems," in *Proc. IEEE Int. Conf. Comput. Commun.* IEEE, July 2016, pp. 1–6.

[129] T. Xie, L. Dai, X. Gao, X. Dai, and Y. Zhao, "Low-complexity SSOR-based precoding for massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 744–747, Feb 2016.

[130] D. Liu and W. Zhou, "A low-complexity precoding algorithm based on improved SOR method for massive MIMO systems," in *Proc. Int. Conf. on Wireless Commun. Signl Process.* IEEE, Oct 2019, pp. 1–6.

[131] S. Hashima and O. Muta, "Fast matrix inversion methods based on Chebyshev and Newton iterations for zero forcing precoding in massive MIMO systems," *EURASIP J. Wireless Comm. and Netw.*, vol. 2020, no. 1, pp. 1–12, Feb 2020.

[132] D. Subitha, J. Mathana, J. Jasmine, and R. Vani, "Modified conjugate gradient algorithms for Gram matrix inversion of massive MIMO downlink linear precoding," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2S11, pp. 2277–3878, Sep 2019.

[133] X. Qin, Z. Yan, and G. He, "A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 276–279, Dec 2015.

[134] W. Song, X. Chen, L. Wang, and X. Lu, "Joint conjugate gradient and Jacobi iteration based low complexity precoding for massive MIMO systems," in *Proc. IEEE Custom Integrated Circuits*. IEEE, July 2016, pp. 1–5.

[135] S. Rahaman, S. Shahabuddin, M. B. Hossain, and S. Shahabuddin, "Complexity analysis of matrix decomposition algorithms for linear MIMO detection," in *Int. Conf. Informatics Electr. Vision*. IEEE, May 2016, pp. 927–932.

[136] S. Shahabuddin, M. H. Islam, M. S. Shahabuddin, M. A. Albreem, and M. Juntti, "Matrix decomposition for massive MIMO detection," *arXiv preprint arXiv:2009.11172v2*, Sep 2020.

[137] M. Wu, B. Yin, K. Li, C. Dick, J. R. Cavallaro, and C. Studer, "Implicit vs. explicit approximate matrix inversion for wideband massive MU-MIMO data detection," *J. Signal Process., Springer*, vol. 90, no. 10, pp. 1311–1328, Oct 2018.

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

[139] J. Wu, S. Fang, L. Li, and Y. Yang, "Qr decomposition and Gram Schmidt orthogonalization based low-complexity multi-user MIMO precoding," pp. 61–64, 2014.

[140] J. Singh and D. Kedia, "Improved precoding algorithm design for downlink large scale MU-MIMO system," *Int. J. of Intelligent Eng. Syst.*, vol. 13, no. 3, Feb 2020.

[141] J. Singh and D. Kedia, "Spectral efficient precoding algorithm for large scale MU–MIMO communication system," in *Proc. Int. Conf. Image Inf. Process.*, 2019, pp. 391–396.

[142] Y. Xu, W. Zou, and L. Du, "A fast and low-complexity matrix inversion scheme based on CSM method for massive MIMO systems," *EURASIP J. Wireless Comm. and Netw.*, vol. 2016, no. 1, pp. 1–6, Dec 2016.

[143] S. Pramono, E. Triyono, and B. B. Subagio, "Performance of leakage based precoding scheme for minimizing interference," *J. Commun., Wiley*, vol. 15, no. 2, Feb 2020.

[144] X. Gao, Z. Lu, Y. Han, and J. Ning, "Near-optimal signal detection with low complexity based on Gauss-Seidel method for uplink large-scale MIMO systems," in *IEEE Int. Symp. Broadband Multim. Sys. Broadcasting*. IEEE, June 2014, pp. 1–4.

[145] X. Wei, L. Xiang, L. Cottatellucci, T. Jiang, and R. Schober, "Cache-aided massive MIMO: Linear precoding design and performance analysis," in *Proc. IEEE Int. Conf. Commun.* IEEE, May 2019, pp. 1–7.

[146] J. Minango and C. de Almeida, "A low-complexity linear precoding algorithm based on Jacobi method for massive MIMO systems," in *Proc. IEEE Veh. Technol. Conf.* IEEE, June 2018, pp. 1–5.

[147] A. Thanos and V. Paliouras, "Hardware trade-offs for massive MIMO uplink detection based on Newton iteration method," in *Int. Conf. Modern Cir. Syst. Techno.* IEEE, May 2017, pp. 1–4.

[148] L. Dai, X. Gao, S. Han, Z. Wang, *et al.*, "Near-optimal linear precoding with low complexity for massive MIMO," *arXiv preprint arXiv:1411.4141*, Nov 2014.

[149] Z. Wang, "Massive MIMO detection algorithms based on MMSE-SIC, ZF-MIC, Neumann series expansion, Gauss-Seidel, and Jacobi method," in *Journal of Phys.*, vol. 1438, no. 1. IOP Publishing, Jan 2020, p. 012006.

[150] J. Minango and C. de Almeida, "Low-complexity MMSE detector based on refinement Gauss-Seidel method for massive MIMO systems," in *IEEE Latin-American Conf. Commun.* IEEE, Nov 2017, pp. 1–5.

[151] Å. Björck, *Numerical methods in matrix computations*. Springer, 2015, vol. 59.

[152] X. Gao, L. Dai, Y. Hu, Y. Zhang, and Z. Wang, "Low-complexity signal detection for large-scale MIMO in optical wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 9, pp. 1903–1912, July 2015.

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

[154] V. L. Babu, L. Mathews, and S. S. Pillai, "Performance analysis of linear and nonlinear precoding in MIMO systems," *Int. J. Advanced Res. Computer Commun. Eng.*, vol. 4, no. 6, pp. 373–376, June 2015.
[155] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, July 2017.
[156] A. D. Dabbagh and D. J. Love, "Precoding for multiple antenna Gaussian broadcast channels with successive zero-forcing," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3837–3850, June 2007.
[157] P. Thakor and R. Sathvara, "Performance of Tomlinson-Harashima precoding and dirty paper coding for broadcast channels in MU-MIMO," *Int. Res. J. of Eng. and Techno.*, vol. 3, no. 4, pp. 2458–2462, Apr 2016.
[158] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 953–961, Mar 2008.
[159] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, Aug 1972.
[160] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *IEE Electron. Lett.*, vol. 7, no. 5, pp. 138–139, Mar 1971.
[161] H. Yan, T. Tian, L. Chen, and J. Qiu, "A physical layer solution for Tomlinson-Harashima precoding in the framework of LTE-advanced," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Dec 2012, pp. 291–296.
[162] S. Begashaw, X. Shao, E. Visotsky, F. Vook, and A. Ghosh, "Evaluation of Tomlinson-Harashima precoding for 5G massive MU-MIMO," in *2018 IEEE 5G World Forum (5GWF)*. IEEE, July 2018, pp. 77–82.
[163] E. C. Peh and Y.-C. Liang, "Power and modulo loss tradeoff with expanded soft demapper for LDPC coded GMD-THP MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 714–724, Feb 2009.
[164] W. Yu, D. P. Varodayan, and J. M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1220–1230, July 2005.
[165] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Efficient Tomlinson-Harashima precoding for spatial multiplexing on flat MIMO channel," in *Proc. IEEE Int. Conf. Commun.*, vol. 3. IEEE, May 2005, pp. 2021–2025.
[166] S. Zarei, W. Gerstacker, and R. Schober, "Low-complexity hybrid linear/Tomlinson-Harashima precoding for downlink large-scale MU-MIMO systems," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Dec 2016, pp. 1–7.
[167] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Apr 2005.
[168] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4567–4582, Dec 2007.
[169] A. Li and C. Masouros, "A constellation scaling approach to vector perturbation for adaptive modulation in MU-MIMO," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 289–292, Mar 2015.
[170] R. Chen, M. Moretti, and X. Wang, "Hybrid TH-VP precoding for multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11 399–11 403, July 2017.
[171] A.-A. Lu, X. Gao, W. Zhong, C. Xiao, and X. Meng, "Robust transmission for massive MIMO downlink with imperfect CSI," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5362–5376, Apr 2019.
[172] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Limited feedback vector perturbation precoding by MinMax optimization," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Dec 2014.
[173] L. Du, L. Li, P. Zhang, D. Miao, and Z. Liu, "Vector perturbation precoding under imperfect CSI and inaccurate power scaling factors," *IEEE Access*, vol. 7, pp. 89 162–89 171, July 2019.
[174] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Computationally efficient vector perturbation precoding using thresholded optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1880–1890, Mar 2013.
[175] C.-B. Chae, S. Shim, and R. W. Heath, "Block diagonalized vector perturbation for multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4051–4057, Dec 2008.
[176] R. Chen, C. Li, J. Li, and Y. Zhang, "Low complexity user grouping vector perturbation," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 189–192, Mar 2012.
[177] W. Ding, T. Lv, and Y. Lu, "A low-complexity vector perturbation precoding approach based on reactive tabu search for large multiuser MIMO systems," in *Int. Conf. Trustworthy Comput. Serv.* Springer, May 2012, pp. 317–324.
[178] C. Masouros, M. Sellahurai, and T. Ratnarajah, "Bridging he Gap between linear and non-linear precoding in small-and large-scale MIMO downlinks," in *Proc. IEEE Int. Conf. Commun.* IEEE, Juhe 2014, pp. 4483–4487.
[179] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Telecommun. Conf.*, vol. 1. IEEE, Nov 2002, pp. 424–428.
[180] D. Wübben, D. Seethaler, J. Jalden, and G. Matz, "Lattice reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, Apr 2011.
[181] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische annalen*, vol. 261, pp. 515–534, July 1982.
[182] P. Q. Nguyen and B. Vallée, *The LLL algorithm*. Springer, 2010.
[183] M. Seysen, "Simultaneous reduction of a lattice basis and its reciprocal basis," *Combinatorica*, vol. 13, no. 3, pp. 363–376, Sep 1993.
[184] V. Brun, "En generalisation av kjedebrøken II," *Skr. Vidensk. Selsk. Kristiana, Mat. Nat. Klasse*, vol. 6, pp. 1–24, 1919.
[185] C. Hermite, "Extraits de lettres de M. Ch. hermite à m. Jacobi sur différents objects de la théorie des nombres." *Journal für die reine und angewandte Mathematik*, vol. 1850, no. 40, pp. 261–278, July 1850.
[186] A. Korkine and G. Zolotareff, "Sur les formes quadratiques," *Mathematische Annalen*, vol. 6, no. 3, pp. 366–389, Sep 1873.
[187] H. Minkowski, "Ueber positive quadratische Formen." *Journal für die reine und angewandte Mathematik*, vol. 1886, no. 99, pp. 1–9, Jan 1886.
[188] C. F. Gauss, *Untersuchungen über höhere Arithmetik*. American Mathematical Soc., 2006, vol. 191.
[189] X. Ma, W. Zhang, and A. Swami, "Lattice-reduction aided equalization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1608–1613, May 2009.
[190] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Lll reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec 2007.
[191] X. Ma and W. Zhang, "Performance analysis for MIMO systems with lattice-reduction aided linear equalization," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 309–318, Feb 2008.
[192] J. Jaldén and P. Elia, "Lr-aided MMSE lattice decoding is DMT optimal for all approximately universal codes," in *Proc. IEEE Int. Symp. Inform. Theory*. IEEE, June 2009, pp. 1263–1267.
[193] C. Studer, D. Seethaler, and H. Bolcskei, "Finite lattice-size effects in MIMO detection," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.* IEEE, Oct 2008, pp. 2032–2037.
[194] C. Windpassinger and R. F. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proceedings 2003 IEEE Information Theory Workshop (Cat. No. 03EX674)*. IEEE, Mar 2003, pp. 345–348.
[195] C. Windpassinger, R. F. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec 2004.
[196] H. An, M. Mohaisen, and K. Chang, "Lattice reduction aided precoding for multiuser MIMO using Seysen's algorithm," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.* IEEE, Sep 2009, pp. 2479–2483.
[197] B. A. LaMacchia, "Basis reduction algorithms and subset sum problems," 1991.
[198] D. Seethaler, G. Matz, and F. Hlawatsch, "Low-complexity MIMO data detection using Seysen's lattice reduction algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc.*, vol. 3. IEEE, Apr 2007, pp. III–53.
[199] K. Zu, R. C. de Lamare, and M. Haardt, "Low-complexity lattice reduction-aided channel inversion methods for large-dimensional multi-user MIMO systems," *arXiv preprint arXiv:1304.6470*, Apr 2013.
[200] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2701–2710, Feb 2009.
[201] L. Zhang, Y. Cai, R. C. de Lamare, and M. Zhao, "Multi-branch vector perturbation precoding design using lattice reduction for MU-MIMO systems," in *Proc. IEEE Veh. Technol. Conf.* IEEE, May 2016, pp. 1–5.
[202] S. Markkandan and N. Venkateswaran, "A lattice reduction-aided information precoder for multiuser communication system," *Int J. Adv Engg Tech/Vol. VII/Issue I/Jan.-March*, vol. 136, p. 141, Mar 2016.
[203] M. Guenach, "Comparison of lattice-reduction-aided vector perturbation and Tomlinson-Harashima precoding," in *Proc. IEEE Wireless Commun. and Networking Conf.* IEEE, Apr 2019, pp. 1–5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3073325, IEEE Access




[204] S. SOBANA and P. THIRUMURUGAN, "Low complexity lattice reduction aided block diagonalization pre-coding for MU-MIMO," 2018.

[205] C. An, "The advanced progress of precoding technology in 5G system," *MS&E*, vol. 231, no. 1, p. 012059, Sep 2017.

[206] M. Mazrouei-Sebdani and W. A. Krzymien, "Vector perturbation precoding for network MIMO: Sum rate, fair user scheduling, and impact of backhaul delay," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 3946–3957, July 2012.

[207] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," *Itg Fachbericht*, pp. 387–394, Jan 2004.

[208] X. Chen, M. Huang, M. Zhao, S. Zhou, and J. Wang, "Analysis and design of Tomlinson-Harashima precoding for multiuser MIMO systems," *Bizaki H. Kh.(Ed.) MIMO Systems, Theory and Applications*, pp. 237–264, Apr 2011.

[209] D. Darsena, G. Gelli, and F. Verde, "Beamforming and precoding techniques," *Wiley 5G Ref: The Essential 5G Reference Online*, pp. 1–29, Oct 2019.

[210] M. H. A. Khan, J.-G. Chung, and M. H. Lee, "Lattice reduction aided with block diagonalization for multiuser MIMO systems," *EURASIP J. Wireless Comm. and Netw.*, vol. 2015, no. 1, p. 254, Dec 2015.

[211] R. Chen, J. Li, C. Li, and W. Liu, "Lattice-reduction-aided MMSE precoding for correlated MIMO channels and performance analysis," *J. Syst. Eng. Electr.*, vol. 23, no. 1, pp. 16–23, Mar 2012.

[212] S. K. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, Sep 2012.

[213] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, 2013.

[214] J.-C. Chen, C.-J. Wang, K.-K. Wong, and C.-K. Wen, "Low-complexity precoding design for massive multiuser MIMO systems using approximate message passing," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5707–5714, July 2015.

[215] M. Shao, Q. Li, W.-K. Ma, and A. M.-C. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, Aug 2019.

[216] S. K. Mohammed and E. G. Larsson, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 547–550, July 2013.

[217] J.-C. Chen, "Low-complexity constant envelope precoding using finite resolution phase shifters for multiuser MIMO systems with large antenna arrays," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7784–7789, Apr 2018.

[218] M. Meng, X. Li, Y. Liu, and Y. Hei, "Low-PAPR approximate message passing precoding algorithm in massive MIMO systems," *IEICE Trans. Commun.*, vol. 101, no. 4, pp. 1102–1107, Apr 2018.

[219] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An efficient manifold algorithm for constructive interference based constant envelope precoding," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1542–1546, Sep 2017.

[220] J.-C. Chen, C.-K. Wen, and K.-K. Wong, "Improved constant envelope multiuser precoding for massive MIMO systems," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1311–1314, July 2014.

[221] C. Mollén, "Low-PAR precoding for very-large multi-user MIMO systems," 2013.

[222] J.-C. Chen, "Low-PAPR precoding design for massive multiuser MIMO systems via Riemannian manifold optimization," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 945–948, Jan 2017.

[223] S. Mukherjee and S. K. Mohammed, "Constant-envelope precoding with time-variation constraint on the transmitted phase angles," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 221–224, Feb 2015.

[224] V. K. Nguyen and J. S. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," in *Proc. IEEE Global Telecommun. Conf.* IEEE, Dec 2008, pp. 1–4.

[225] C. Masouros, T. Ratnarajah, M. Sellathurai, C. B. Papadias, and A. K. Shukla, "Known interference in the cellular downlink: A performance limiting factor or a source of green signal power?" *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 162–171, Oct 2013.

[226] P. V. Amadori and C. Masouros, "Constant envelope precoding by interference exploitation in phase shift keying-modulated multiuser transmission," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 538–550, Nov 2016.

[227] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1396–1404, Mar 2009.

[228] C. Masouros and E. Alsusa, "Soft linear precoding for the downlink of DS/CDMA communication systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 203–215, 2010.

[229] E. Alsusa and C. Masouros, "Adaptive code allocation for interference management on the downlink of DS-CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2420–2424, July 2008.

[230] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Vector perturbation based on symbol scaling for limited feedback MISO downlinks," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, Jan 2014.

[231] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, May 2015.

[232] C. Mollén and E. G. Larsson, "Multiuser MIMO precoding with per-antenna continuous-time constant-envelope constraints," in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.* IEEE, June 2015, pp. 261–265.

[233] A. Brihuega, L. Anttila, and M. Valkama, "Performance comparison of constant envelope and zero-forcing precoders in multiuser massive MIMO," in *Proc. IEEE Wireless Commun. and Networking Conf.* IEEE, Apr 2018, pp. 1–6.

[234] L. Chu, F. Wen, L. Li, and R. Qiu, "Efficient nonlinear precoding for massive MIMO downlink systems with 1-bit DACs," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4213–4224, June 2019.

[235] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inform. Theory Workshop*, 2010, pp. 1–5.

[236] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: II. analysis and validation," in *Proc. IEEE Inform. Theory Workshop*. IEEE, Jan 2010, pp. 1–5.

[237] A. Mahmood, J. Kang, and H. Lee, "Sparse or dense—message passing (MP) or approximate message passing (AMP) for compressed sensing signal recovery," in *Proc. IEEE Pacific Rim Conf. on Commun., Computers and Signal Process.* IEEE, Aug 2013, pp. 259–264.

[238] H. Bao, J. Fang, Z. Chen, H. Li, and S. Li, "An efficient bayesian PAPR reduction method for OFDM-based massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4183–4195, Mar 2016.

[239] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[240] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr 1999.

[241] B. Murmann *et al.*, "ADC performance survey 1997-2016," *Online] http://www. stanford. edu/murmann/adcsurvey. html*, Aug 2017.

[242] D. W. K. Ng, E. S. Lo, and R. Schober, "Robust beamforming for secure communication in systems with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4599–4615, Apr 2014.

[243] J. J. Bussgang, "Cross correlation functions of amplitude-distorted Gaussian signals," 1952.

[244] J. Guerreiro, R. Dinis, and P. Montezuma, "Use of 1-bit digital-to-analogue converters in massive MIMO systems," *IEE Electron. Lett.*, vol. 52, no. 9, pp. 778–779, Apr 2016.

[245] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *Proc. IEEE Sensor Array and Multichannel Sign. Proc. Workshop*. IEEE, July 2016, pp. 1–5.

[246] Y. Li, C. Tao, A. L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Mar 2017.

[247] M. Bengtsson and B. Ottersten, "Optimal downlink beamformingusing semidefinite optimization," in *Proc. Annual Allerton Conf. Commun., Contr., Computing*, 1999, pp. 987–996.

[248] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, Apr 2010.

[249] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, Mar 2004.

[250] J. Eckstein and D. P. Bertsekas, "On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone oper-

ators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, Apr 1992.

[251] O. Castaneda, T. Goldstein, and C. Studer, "POKEMON: A non-linear beamforming algorithm for 1-bit massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc.* IEEE, Mar 2017, pp. 3464–3468.

[252] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug 2002.

[253] T. Erpek, T. J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications," in *Development and Analysis of Deep Learning Architectures*. Springer, 2020, pp. 223–266.

[254] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *Proc. IEEE Global Telecommun. Conf.* IEEE, 2018, pp. 584–588.

[255] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, 2019.

[256] M. Goutay, F. A. Aoudia, and J. Hoydis, "Deep hypernetwork-based MIMO detection," *arXiv preprint arXiv:2002.02750*, 2020.

[257] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, 2019.

[258] F. Sohrabi, K. M. Attiah, and W. Yu, "Deep learning for distributed channel feedback and multiuser precoding in FDD massive MIMO," *arXiv preprint arXiv:2007.06512*, 2020.

[259] J.-M. Kang, I.-M. Kim, and C.-J. Chun, "Deep learning-based MIMO-NOMA with imperfect SIC decoding," *IEEE Syst. J.*, 2019.

[260] Y.-S. Jeon and M. Min, "Large system analysis of two-stage beamforming with limited feedback in FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4984–4997, 2018.

[261] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, 2015.

[262] A. Balatsoukas-Stimming, O. Castañeda, S. Jacobsson, G. Durisi, and C. Studer, "Neural-network optimized 1-bit precoding for massive MU-MIMO," *CoRR*, vol. abs/1903.03718, 2019. [Online]. Available: http://arxiv.org/abs/1903.03718

[263] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, 2020.

[264] L. Du, L. Li, H. Q. Ngo, T. C. Mai, and M. Matthaiou, "Cell-free massive MIMO: Joint maximum-ratio and zero-forcing precoder with power control," *IEEE Trans. Commun.*, pp. 1–1, 2021.

[265] J. Zhang, Y. Wei, E. Björnson, Y. Han, and S. Jin, "Performance analysis and power control of cell-free massive MIMO systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55 302–55 314, 2018.

[266] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[267] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, 2017.

[268] Z. Chen and E. Bjornson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, 2018.

[269] A. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "New insights on channel hardening in cell-free massive MIMO networks," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–7.

[270] A. Á. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "Channel hardening in cell-free and user-centric massive MIMO networks with spatially correlated ricean fading," *IEEE Access*, vol. 8, pp. 139 827–139 845, 2020.

[271] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, 2017.

[272] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4758–4774, 2020.

[273] M. Matthaiou, O. Yurduseven, H. Q. Ngo, D. Morales-Jimenez, S. L. Cotton, and V. F. Fusco, "The road to 6G: Ten physical layer challenges for communications engineers," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 64–69, 2021.

[274] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, 2020.

[275] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Telecommun. Conf.*, 2013, pp. 3679–3684.

[276] T. Ding, Y. Zhao, L. Li, D. Hu, and L. Zhang, "Energy-efficient hybrid precoding for beamspace MIMO systems with lens array," in *Proc. IEEE Int. Conf. Commun. Technol.*, 2019, pp. 628–632.

[277] T. Ding, Y. Zhao, L. Li, D. Hu, and L. Zhang, "Hybrid precoding for beamspace MIMO systems with sub-connected switches: A machine learning approach," *IEEE Access*, vol. 7, pp. 143 273–143 281, 2019.

[278] M. Abdelghany, U. Madhow, and A. Tölli, "Efficient beamspace downlink precoding for mmWave massive MIMO," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, 2019, pp. 1459–1464.

[279] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2838–2849, 2020.

[280] X. Meng, F. Liu, J. Zhou, and S. Yang, "Interference exploitation precoding for intelligent reflecting surface aided communication system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 126–130, 2021.

[281] S. Wang, Q. Li, and M. Shao, "One-bit symbol-level precoding for MU-MISO downlink with intelligent reflecting surface," *IEEE Signal Process. Lett.*, vol. 27, pp. 1784–1788, 2020.

[282] M. Shao, Q. Li, and W. K. Ma, "Minimum symbol-error probability symbol-level precoding with intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1601–1605, 2020.

[283] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837–851, 2021.

[284] R. Liu, H. Li, M. Li, and Q. Liu, "Symbol-level precoding design for intelligent reflecting surface assisted multi-user MIMO systems," in *Proc. Int. Conf. on Wireless Commun. Signl Process.*, 2019, pp. 1–6.

[285] A. Mohammadian, C. Tellambura, and M. Valkama, "Analysis of self-interference cancellation under phase noise, CFO, and IQ imbalance in GFDM full-duplex transceivers," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 700–713, 2020.

[286] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, 2017.

[287] K. Xu, Z. Shen, Y. Wang, and X. Xia, "Location-aided mMIMO channel tracking and hybrid beamforming for high-speed railway communications: An angle-domain approach," *IEEE Syst. J.*, vol. 14, no. 1, pp. 93–104, 2020.

[288] Z. Shen, K. Xu, X. Xia, W. Xie, and D. Zhang, "Spatial sparsity based secure transmission strategy for massive MIMO systems against simultaneous jamming and eavesdropping," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3760–3774, 2020.

[289] K. Xu, Z. Shen, Y. Wang, X. Xia, and D. Zhang, "Hybrid time-switching and power splitting SWIPT for full-duplex massive MIMO systems: A beam-domain approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7257–7274, 2018.

[290] Z. Shen, K. Xu, and X. Xia, "Beam-domain anti-jamming transmission for downlink massive MIMO systems: A stackelberg game perspective," *IEEE Trans. Inf. Forensics Security*, pp. 1–1, 2021.