Massive MIMO Linear Precoding: A Survey

Nusrat Fatema, Guang Hua, *Member, IEEE*, Yong Xiang, *Senior Member, IEEE*, Dezhong Peng, *Member, IEEE*, and Iynkaran Natgunanathan, *Member, IEEE*

***Abstract*—Recently, massive multiple-input multiple-output (MIMO) system has attracted tremendous research interests in wireless communications. Among various performance aspects and design problems regarding massive MIMO, precoding at the base station (BS) is one of the most important function components to ensure reliable downlink transmission. This paper provides a survey of linear precoding techniques for downlink transmission under both single-cell (SC) and multicell (MC) scenarios. In an SC scenario, a series of classical precoding techniques are reviewed and compared. In an MC scenario, massive MIMO suffers from pilot contamination problems, where orthogonal pilot sequences have to be reused among cells, leading to estimation errors of channel state information at the transmitter. Precoding techniques with and without the cooperation among BSs are reviewed and the cooperative precoding techniques are further divided into partial and full cooperations. To the best of our knowledge, this is the first work that comprehensively surveys all the existing linear precoding techniques. The comprehensive review reveals current challenges in designing the precoders, where technical merits are usually bounded by practical implementation issues. We show that a viable precoding technique for massive MIMO systems is still unknown to date, and suggest several research potentials that are worthy of future research efforts.**

***Index Terms*—Downlink, massive multiple-input multiple-output (MIMO), multicell (MC), precoding techniques, single-cell (SC), wireless communications.**

I. Introduction

OVER the past decade, the number of digital wireless devices has increased dramatically and this trend will continue to thrive. The growing traffic in digital communication systems requires more efficient utilization of available equipment and network resources. To achieve this, technology has advanced in each generation of communication architecture, aiming to improve network capacity, efficiency, and reliability. As one of the most promising wireless communication solutions, multiple-input multiple-output (MIMO) technology has been widely used in IEEE 802.11ac/n (WiFi), IEEE 802.16e (WiMAX), LTE/LTE-A (4G), and other protocols and systems [1]. The development of MIMO wireless communication technology originated from the idea of point-to-point MIMO communication where the two communicating devices have multiple antennas and the idea of multi-user MIMO (MU-MIMO) communication where a base station (BS) equipped with multiple antennas serves single-antenna end users [2]. MU-MIMO communication is more advantageous than point-to-point MIMO communication in terms of enabling cheap single-antenna terminals and being less sensitive to propagation environment. However, a conventional MU-MIMO link suffers from interuser interference, which considerably limits the performance of the system. Although techniques such as dirty paper coding (DPC) can reduce the effect of interuser interference, system implementation complexity could be sacrificed as a tradeoff [3].

In contrast, if the BS uses many antennas, the interuser interference can be largely eliminated because of the asymptotic orthogonality among the MIMO channels, which will yield huge performance improvement [4], [5]. Meanwhile, a large antenna array at the BS also brings the benefit of huge capacity gain [6]. In the context of massive MIMO, it is normally assumed: 1) The number of the large antenna array elements at the BS is greater than the number of single antenna users being served in the same frequency band. 2) The system is operated in a time-division duplexing (TDD) mode [7]. Other benefits of massive MIMO include improving spectral efficiency (SE) without BS densification, increasing energy efficiency (EE) by extensive use of inexpensive low-power components, enlarging the capability of focusing transmission signal energy on smaller regions, and so on [8], [9].

Due to the above-mentioned advantages of massive MIMO over conventional point-to-point MIMO and MU-MIMO, extensive research and development work has been carried out in both academia and industry with the objectives from establishing a new theoretical foundation to building practically deployed systems [5]. Subdomain research topics of massive MIMO are indeed diverse, including but not limited to precoding, resource allocation, channel estimation, signal detection, antenna array design, performance analysis, SE and EE control, pilot decontamination, etc. This paper is dedicated to a comprehensive review of precoding techniques in massive MIMO systems. Precoding is an important signal processing procedure in downlinks, which utilizes the channel state information (CSI) at the transmitter. With reliable CSI, precoding can be used to maximize link performance.[1]

Manuscript received February 8, 2017; revised June 25, 2017 and October 4, 2017; accepted November 18, 2017. Date of publication December 7, 2017; date of current version November 22, 2018. *(Corresponding author: Yong Xiang.)*

N. Fatema, Y. Xiang, and I. Natgunanathan are with the School of Information Technology, Deakin University, Burwood, VIC 3125, Australia (e-mail: nfatem@deakin.edu.au; yxiang@deakin.edu.au; iynkaran.natgunananthan@deakin.edu.au).

G. Hua is with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: ghua@ieee.org).

D. Peng is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: pengdz@scu.edu.cn).

Digital Object Identifier 10.1109/JSYST.2017.2776401

[1]Note that transmit signal processing without the knowledge of CSI is termed as space-time processing [12]. Instead of relying on CSI, space-time processing explores transmit diversity to improve link performance.

CSI is estimated by linear channel estimators through uplink pilot transmissions, which leverages the channel reciprocity. In TDD operation, the time required to acquire CSI is dependent on the number of users but independent of the number of BS antennas. Therefore, TDD is a more preferable operating mode for massive MIMO systems than the frequency-division duplexing (FDD) mode, where the amount of time required by FDD to obtain CSI relies on the excessive number of BS antennas [11]. However, the uncontrollable wireless propagation environment usually make it difficult to obtain reliable CSI [12]. Although perfect CSI is usually unavailable at the transmitter, the performance of downlink transmission largely depends on CSI and the corresponding precoding technique employed. Given the estimated (not perfect) CSI, precoding can still subdue interference and upsurge the achievable sum rates. In other words, massive MIMO employing suitable precoding techniques can yield large gains in SE and EE as compared to conventional MIMO systems, as the effects of noise and interference are negligible when the number of antennas approaches infinity [13]. The price to pay for the advantages of massive MIMO is the potentially high computational complexity of the precoding process [14].

A massive MIMO wireless system could use a single-cell (SC) or a multicell (MC) structure. In an SC system, a single BS equipped with a large number of antennas serves multiple single-antenna users, while an MC system contains multiple linked SC systems. Furthermore, according to whether the BSs in different cells collaborate or not, MC systems can be classified as noncooperative and cooperative MC systems. In MC massive MIMO, because the maximum number of orthogonal pilot sequences is constrained by the duration of coherence interval and coherent bandwidth, the pilot sequences among terminals in different cells may no longer be orthogonal, leading to pilot contamination problem [15]. Pilot contamination is a unique problem in MC massive MIMO systems [1].

This paper reviews the existing linear precoding techniques for massive MIMO wireless systems according to different cell scenarios. It serves as the first comprehensive summary and tutorial of precoding techniques for researchers and practitioners in this research field. Specifically, precoding techniques in each cell scenario, with and without BS collaborations, are reviewed and compared. Note that a series of nonlinear precoding techniques have also been proposed in the literature, including DPC [16], vector perturbation [17], and lattice-aided methods [18]. However, due to the usage of huge number of antennas at each BS in massive MIMO, linear precoding techniques are more advantageous over nonlinear precoding ones, in terms of achieving better performance and higher computational efficiency. Thus, nonlinear precoders are not considered in this paper.

Recently, several survey papers related to massive MIMO systems have been published [1], [5], [19]–[22]. While these survey papers review a number of aspects of massive MIMO, none of them extensively discuss the important downlink linear precoding techniques. More specifically, compared with the survey papers [1], [5], and [19]–[22], the unique contributions of this paper are as follows.

1) It presents a comprehensive overview, comparison, and discussion of the existing linear precoding techniques for downlink transmission under both SC and MC scenarios.
2) In the SC scenario, the popular precoding techniques are discussed.
3) In the MC scenario, the important pilot contamination problems are reviewed.
4) The partial and full cooperative precoding techniques are provided and discussed.
5) The challenges and research potentials in massive MIMO linear precoding are discussed in detail.

The rest of the paper is organized as follows. In Section II, we introduce the downlink channel model of SC massive MIMO systems, where the conventional linear precoders are studied and the corresponding achievable data rates are compared. Then, a thorough review of noncooperative precoding techniques in the MC scenario is provided in Section III, followed by the discussions of cooperative precoding techniques in Section IV. The current challenges and future research potentials about precoding techniques for massive MIMO systems are addressed in Section V. Section VI concludes the paper. Symbols and notations used in this paper are shown in Table I.

TABLE I
SYMBOLS AND NOTATIONS

Symbols	Notations
$(\cdot)^*$	Complex conjugate operator
$(\cdot)^T$	Transpose operator
$(\cdot)^H$	Hermitian transpose operator
$(\cdot)^+$	Moore Penrose pseudo-inverse operation
$(\cdot)^{-1}$	Inverse operator
$\Vert\cdot\Vert$	Euclidean norm operator
$\triangleq$	Equal by defination
$\mathrm{diag}(\cdot)$	Diagonal elements of a matrix
$\mathrm{tr}(\cdot)$	Trace function
$\mathbf{I}_K$	$K \times K$ identity matrix
$\mathcal{CN}(0,\sigma)$	Circular symmetric complex Gaussian distribution
$\det(\cdot)$	Determinant of a matrix
$\circ$	Khatri–Rao product
$\Re(\cdot)$	Real part of a complex number
$E(\cdot)$	Expectation operator

II. Single-Cell Precoding Techniques

In this section, we consider an SC downlink massive MIMO system consists of a central BS equipped with massive number of antennas transmitting to multiple users simultaneously over the same spectrum [23]. Let the number of antennas at the BS be M and the number of single-antenna users be K. Then, the attainable antenna array gain is proportional to M and the multiplexing gain is K. In a practical situation, the achievable throughput and gains depend on the associated precoding measures, which are the focus of the following sections.

A. Single-Cell Downlink System Model

The transmitted signal vector for the K users during the downlink transmission, where $M > K$, can be expressed as

$$\mathbf{x} = \sqrt{\rho}\mathbf{W}\mathbf{s} \tag{1}$$

where $\mathbf{W} \in \mathbb{C}^{M\times K}$ is the linear precoding matrix, $\mathbf{s} \in \mathbb{C}^{K\times 1}$ is the transmitted source information before precoding, and ρ is the average transmit power at the BS. Here, M and K are both large and their ratio is assumed to be a constant [24]. The precoding matrix $\mathbf{W}$ is a function of the channel matrix denoted by $\mathbf{H} \in \mathbb{C}^{M\times K}$. The power of the transmitted source signal is normalized, i.e., $\|\mathbf{s}\|^2 = 1$. To satisfy the power constraint at the BS, $\mathbf{W}$ is chosen such that $\mathrm{tr}(\mathbf{W}\mathbf{W}^H) = 1$.

In the TDD mode, the downlink channel is simply the transpose of the channel matrix $\mathbf{H}$ [1]; hence, the set of signals received at K terminals are

$$\mathbf{y} = \mathbf{H}^T\mathbf{x} + \mathbf{n} = \sqrt{\rho}\mathbf{H}^T\mathbf{W}\mathbf{s} + \mathbf{n} \tag{2}$$

where $\mathbf{n} \in \mathbb{C}^{K\times 1}$ represents noise and interference, and $\forall k \in \{0, 1, \ldots, K-1\}$, each element follows $n_k \sim \mathcal{CN}(0, \sigma)$. It can be seen from (2) that the precoding processing plays an important role in determining the downlink performance. We further assume a Rayleigh block fading channel, where each channel vector in $\mathbf{H}$, i.e., $\mathbf{h}_k$ follows $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_{M\times 1}, \mathbf{\Phi}/K)$, with $\mathbf{\Phi} \in \mathbb{C}^{M\times M}$ being the channel covariance matrix with bounded norm as $M \to \infty$. Thus, we have $\mathrm{tr}(\mathbf{\Phi}) \propto M/K$.

B. Precoding Techniques

1) Matched Filter (MF): The matched filter (MF) precoder is simply the conjugate transpose of the downlink channel matrix, i.e.,

$$\mathbf{W}_{\text{MF}} = \sqrt{\alpha}\mathbf{H}^* \tag{3}$$

where α is a scaling factor to normalize signal power. The received signal vector is then given by

$$\mathbf{y}_{\text{MF}} = \sqrt{\rho\alpha}\mathbf{H}^T\mathbf{H}^*\mathbf{s} + \mathbf{n}. \tag{4}$$

MF precoder is also known as maximum ratio transmission (MRT), which maximizes signal gain at the intended user [25], [26]. It is the counterpart of the maximal-ratio combining receiver for uplink. The key performance parameters for SC MF precoding, i.e., achievable sum rate and the total downlink transmit power, are discussed in [3] and [25]. In terms of EE, it is shown in [27] that good EE can be achieved with $M = 81$ and $K = 77.7$. However, since the values of M and K are close, it results in a degenerated massive MIMO case and the asymptotic massive MIMO property does not hold. In this situation, the system suffers from a strong interuser interference and the rate per user is small.

With the increase of BS antennas, the channel vectors in $\mathbf{H}$ become closer to mutually orthogonal. As a result, the term $\mathbf{H}^{\mathbf{T}}\mathbf{H}^*$ approaches a diagonal matrix [13], leading to the optimal solution. Consequently, MF precoding is near-optimal, as long as the number of BS antennas is much greater than the number of terminal users.

2) Zero Forcing (ZF): Zero forcing (ZF) precoding is another type of basic precoding technique, which eliminates the interference by transmitting the signal toward the intended user while nulling in the directions of other users. The ZF precoder is obtained by

$$\mathbf{W}_{\text{ZF}} = \sqrt{\alpha}\mathbf{H}^*(\mathbf{H}^T\mathbf{H}^*)^{-1} \tag{5}$$

and the corresponding received signal vector is

$$\mathbf{y}_{\text{ZF}} = \sqrt{\rho\alpha}\mathbf{H}^T\mathbf{H}^*(\mathbf{H}^T\mathbf{H}^*)^{-1}\mathbf{s} + \mathbf{n}. \tag{6}$$

The term $\mathbf{H}^T\mathbf{H}^*$ forms a Gram matrix whose diagonal elements denote power imbalance among the channels, while the off-diagonal elements characterize mutual correlations between the channels. When highly correlated channels exist, ZF precoding decorrelates the channels at the price of losing channel capacity [24]. It is an optimal precoding scheme in the absence of additive noise [28]. When additive noise is present, this precoding technique could amplify the noise effect.

3) Regularized Zero Forcing (RZF): Regularized zero forcing (RZF) precoder has been considered as the state-of-the-art linear precoder for MIMO wireless communications systems for its capability of trading off the advantages of MF and ZF precoders [14], [26], [29], [30]. Its popularity is also reflected by its alternative names such as eigenvalue-based beamforming [31], virtual signal-to-interference-noise ratio (SINR) maximizing beamforming [32], transmit Wiener filter [33], [34], and signal-to-leakage-and-noise ratio (SLNR) maximizing beamforming [35]. According to [26] and [36], the RZF precoding matrix is given by

$$\mathbf{W}_{\text{RZF}} = \sqrt{\alpha}\mathbf{H}^*(\mathbf{H}^T\mathbf{H}^* + \mathbf{X} + \lambda\mathbf{I}_K)^{-1} \tag{7}$$

which is a ZF precoder regularized by a Hermitian non-negative matrix $\mathbf{X}$ and a regularization factor λ. The corresponding received signal vector is

$$\mathbf{y}_{\text{RZF}} = \sqrt{\rho\alpha}\mathbf{H}^T\mathbf{H}^*(\mathbf{H}^T\mathbf{H}^* + \mathbf{X} + \lambda\mathbf{I}_K)^{-1}\mathbf{s} + \mathbf{n}. \tag{8}$$

The choices of $\mathbf{X}$ and λ are discussed in [36]. If $\mathbf{X} = \mathbf{0}$, then (7) becomes an MF precoder when $\lambda \to \infty$ and a ZF precoder when $\lambda \to 0$. The extensions of the RZF technique with arbitrary user priorities are discussed in [37], where the RZF precoding matrix is modified to achieve optimality when the BS has the knowledge of statistical information of the user positions. This technique is referred to as position-aware RZF (PA-RZF) precoder. It turns out that PA-RZF is equivalent to the optimal linear precoder when the same SINR constraint and different path loss are imposed for all users. On the other hand, the commonly used RZF precoder becomes optimal only when the ratio between the SINR requirement and the average channel attenuation is the same for all users. Generally, the RZF precoder is obtained via minimizing the mean square error (MSE) between the transmitted and received symbols, which is, thus, also termed as minimum MSE (MMSE) precoder [1], [24], [38], [39]. The computation of the RZF precoding matrix involves the inversion of a matrix with very large dimension, especially for large values of M and K [40].

4) Truncated Polynomial Expansion (TPE): Truncated polynomial expansion (TPE)-based precoding is a recently proposed technique to reduce the computational complexity of RZF precoder while maintaining similar performance [14], [40], [41]. This technique replaces the matrix inversion in RZF precoder by a polynomial expansion, which is then truncated. The TPE precoding matrix is given by

$$\mathbf{W}_{\text{TPE}} = \sum_{j=0}^{J-1} w_j (\mathbf{H}^T\mathbf{H}^*)^j \mathbf{H}^T \tag{9}$$

where w_j is the coefficient of the precoder polynomial of order J. A proper selection of J enables a smooth transition

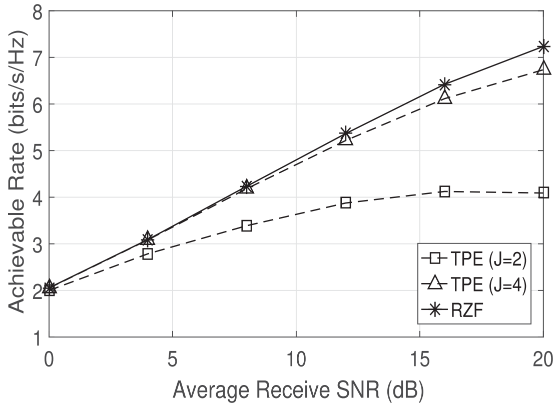

Fig. 1. Achievable sum rates of TPE ($J = 2, 4$) and RZF precoding techniques ($K = 128$ and $M = 512$).

between the traditional low-complexity MF precoding and the high-complexity RZF precoding. Specifically, if $J = 1$, then $\mathbf{W}_{\text{TPE}} = w_0 \mathbf{H}^T$, which is equivalent to the MF precoder. On the other hand, the approximation of RZF precoder using the TPE technique is established by the inversion formula of a positive definite Hermitian matrix, and the details are provided in [41]. Matrix polynomial based precoding scheme with power allocation for downlink transmission is provided in [42], where the precoding and power allocation matrices are optimized to maximize the minimum user rate as well as the sum rate. It has been shown that even for a small number of matrix polynomial terms, TPE-based schemes closely approach the sum rate and the minimum user rate of the optimal RZF precoder. Fig. 1 compares the achievable sum rates of TPE and RZF precoding techniques against average receive SNR. We can see that RZF performs better than TPE. However, the achievable rate of TPE approaches that of RZF with increasing J value. In terms of computational complexity, by using an appropriate implementation technique, e.g., Horners scheme, significant computation reduction can be achieved as compared to the RZF precoder.

The advantages of TPE in terms of system efficiency are summarized as follows.

1) The computational delay is minimized as computations are pipelined over multiple processing cores and operations are spread out uniformly over time [43].
2) The precoding computation is divided into a number of simple matrix–vector multiplications, which can be highly parallelized and can be implemented using a multitude of simple application-specific circuits.
3) The parameter J can be tailored directly to the available hardware.

5) Phased ZF (PZF): Another low-complexity ZF-based precoding technique, termed as phased ZF (PZF), is proposed in [44], which takes advantage of the low-dimensional radio frequency (RF) chain. Specifically, the precoder matrix is decomposed into RF and baseband processing, with the corresponding precoding submatrices denoted as $\mathbf{R} \in \mathbb{C}^{M\times K}$ and $\mathbf{B} \in \mathbb{C}^{K\times K}$, respectively. The received signal vector in (2) is then modified as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}^T\mathbf{R}\mathbf{B}\mathbf{s} + \mathbf{n} \tag{10}$$

where $\mathbf{R}$ is determined according to $\mathbf{H}$. Therefore, the precoding design for the above-mentioned model lies in the design of dimension-reduced square matrix $\mathbf{B}$. Specifically, applying ZF processing to the model, we can have [44]

$$\mathbf{W}_{\text{PZF}} \triangleq \mathbf{B} = \tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)^{-1}\mathbf{\Lambda} \tag{11}$$

where $\tilde{\mathbf{H}} \triangleq \mathbf{H}^T\mathbf{R}$ and $\mathbf{\Lambda} \in \mathbb{R}_+^{K\times K}$ is a positive diagonal matrix for column power normalization. It can be seen that $\mathbf{W}_{\text{PZF}} \in \mathbb{C}^{K\times K}$, whose number of rows is dramatically reduced from M to K. Furthermore, it has also been shown in [44] that the SE of a PZF precoder is tightly upper bounded by that of a ZF precoder. A brief summary of the advantages and disadvantages of the MF, ZF, RZF, TPE, and PZF is shown in Table II.

C. Achievable Rates

The DPC technique has been shown to achieve the optimal downlink sum rate in massive MIMO wireless systems [24], which is

$$C_{\text{DPC}} = \max_{\mathbf{P}} \log_2 \det\left(\mathbf{I}_M + \rho\mathbf{H}\mathbf{P}\mathbf{H}^H\right) \quad \text{bits/s/Hz} \tag{12}$$

where $\mathbf{P} \in \mathbb{R}_+^{K\times K}$ is a diagonal power allocation matrix, and $\sum \text{diag}(\mathbf{P}) = 1$. This optimal rate has been considered as a performance benchmark. However, this nonlinear scheme entails high computational complexity [40], making it difficult to be implemented in practice [24]. The work in [24] has compared the basic precoding schemes, both the ZF and RZF precoders, with the optimal DPC precoder, exploring how the low complexity precoders perform in relation to the optimal DPC scheme. It is shown that with linear precoding, a sum rate as high as 98% of that of the DPC scheme can be achieved for two single antenna users served by 20 BS antennas.

The achievable data rates of ZF and MF precoders have been discussed in [25], with the assumption that the total downlink power is fixed and equally divided among all users. From Shannon's theorem, the achievable rate over additive white Gaussian noise (AWGN) is obtained as a function of the signal-to-noise ratio (SNR) [25]

$$C = \log_2(1 + \text{SNR}) \quad \text{bits/s/Hz.} \tag{13}$$

The achievable data rate per user in an SC downlink massive MIMO system with perfect CSI is [25]

$$R_k = \log_2(1 + \text{SINR}_k) \quad \text{bits/s/Hz} \tag{14}$$

where SINR_k is the SINR at the kth user. For large values of M and K, the achievable sum rates for ZF and MF precoders with perfect CSI are [25]

$$R_{\text{sum}}^{\text{ZF}} = K\log_2\left[1 + \rho\left(\frac{M-K}{K}\right)\right] \tag{15}$$

and

$$R_{\text{sum}}^{\text{MF}} = K\log_2\left[1 + \frac{\rho M}{K(\rho+1)}\right] \tag{16}$$

where the second terms in logarithm are the expressions of per-user SINR for MF and ZF precoders, respectively. When the number of users K and the number of BS antennas M grow

TABLE II
ADVANTAGES AND DISADVANTAGES OF THE MF, ZF, RZF, TPE, AND PZF PRECODING TECHNIQUES

SC precoding techniques	Advantages	Disadvantages
MF	-Near optimal performance if there are more BS antennas than users -Low computational complexity -Better performance at lower SNRs	-Unable to achieve full diversity at high spectral efficiency -Suffering from error floors for all positive multiplexing gains -Having lower achievable rate in the case of less BS antennas -Not robust against interuser interference
ZF	-Low computational complexity -Higher power efficient -Better performance at high SNRs -Decoupling a multi-user channel into independent single-user channels -Achieving a large portion of dirty paper coding capacity	-Noise amplification and power penalty if the channel is highly correlated -Unable to support too many users -Medium complexity
RZF	-Guaranteed optimality if the ratio between the SINR requirement and the average channel attenuation is the same for all users	-Requiring matrix inverse calculation, leading to high complexity -Suffering from error floors for all positive multiplexing gains
TPE	-Scalable computational complexity, thus computationally efficient -Balancing precoding complexity and system throughput -More suitable for real-time hardware implementation	-Lower order techniques having limited performance -Higher performance requiring considerably more hardware
PZF	-Highly desirable closed-form performanc -Low complexity -Facilitating multistream processing -Large power gain	-Spectral efficiency being tightly upper bounded by the ZF precoder

large but with fixed ratio M/K, the SINRs for ZF and MF under imperfect CSI are [5]

$$\text{SINR}_{\text{ZF}} = \frac{\epsilon^2\rho(M-K)}{(1-\epsilon^2)\rho K+1} \quad \text{and} \quad \text{SINR}_{\text{MF}} = \frac{\epsilon^2\rho M}{(\rho+1)K} \tag{17}$$

where ϵ is the reliability of the channel estimate.

The SINR of RZF precoding is shown in [55, eq. (17)]. Then, substituting the SINR of RZF into (14), one can obtain the achievable data rate of RZF precoding. Since the expression in [55, eq. (17)] and the expressions of its associated parameters are complex, they are not shown here. Readers can refer to [55] for more details. Compared with ZF precoding and MF precoding, RZF precoding has better performance at both low and high transmission power.

Similar performance analysis on TPE precoding can be found in [14] and [41], and the ergodic achievable data rate is given as [14]

$$R_k = E\left[\log_2\left(1+\frac{\mathbf{w}^H\mathbf{A}_k\mathbf{w}}{\mathbf{w}^H\mathbf{B}_k\mathbf{w}-\mathbf{w}^H\mathbf{A}_k\mathbf{w}+\sigma^2}\right)\right] \text{ bits/s/Hz} \tag{18}$$

where $\mathbf{w}$ is the precoding vector and the matrices $\mathbf{A}_k, \mathbf{B}_k \epsilon \mathbb{C}^{J\times J}$ have entries given as $[\mathbf{A}_k]_{j,m} = \mathbf{h}_k^H(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^j \hat{\mathbf{H}}\mathbf{e}_k\mathbf{e}_k^H\hat{\mathbf{H}}^T(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^m\mathbf{h}_k$ and $[\mathbf{B}_k]_{j,m} = \mathbf{h}_k^H(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^{j+m+1}\mathbf{h}_k$. Here, $\mathbf{e}_k$ is the kth column of the identity matrix and $\mathbf{h}_k \epsilon \mathbb{C}^{M\times 1}$ represents the random channel vector between the BS and the kth user.

Regarding PZF precoding, its achievable data rate is [44]

$$R_k = E\left[\log_2\left(1+\frac{\rho}{K}\left|\mathbf{h}_k^T\mathbf{r}_k\right|^2\right)\right] \tag{19}$$

where $\mathbf{r}_k$ denotes the kth column of the RF precoder $\mathbf{R}$. The RZF precoder has similar performance compared to the ZF precoder but is more efficient in computation.

Fig. 2 shows the achievable sum rates of MF, ZF, and PZF precoding techniques under the same available downlink transmit power and a fixed number of users. One can see that as the number of transmit antennas increases with $M > K$, the achievable sum rates of MF, ZF, and PZF precoding techniques increase. Also as expected, the PZF precoding technique achieves higher data rate than MF and ZF precoding techniques, while the ZF precoding technique outperforms the MF counterpart.

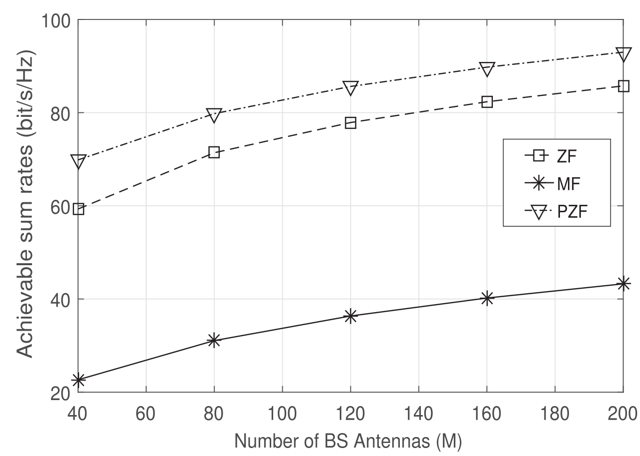

Fig. 2. Achievable sum rates of MF, ZF, and PZF precoding techniques ($K = 10$ and total power = 20 dB).

III. NONCOOPERATIVE MULTICELL PRECODING TECHNIQUES

An MC massive MIMO system involves multiple BSs with massive number of antennas serving different users in different cells [23]. The advantages of an MC network are as follows.

1) Given seamless intercell handover, it can increase coverage area and enhance user mobility.
2) One or more users can communicate their messages through multiple surrounding BSs to provide diversity against thermal and Rayleigh fading.
3) Substantial gain in capacity and increased fairness across cell locations could be accrued.

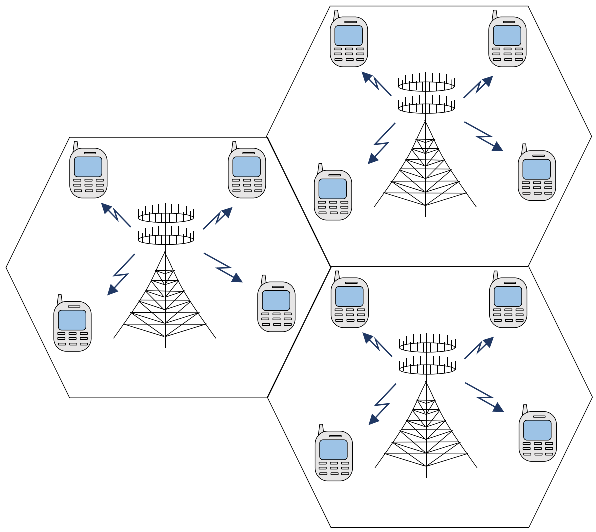

Fig. 3. Noncooperative three-cell downlink massive MIMO system.

4) It can significantly improve SE by reusing frequency bands in multiple cells.

However, MC systems also encounter some challenges [45], e.g., the pilot contamination problem caused by intercell interference. A possible way of tackling pilot contamination is precoding. Depending on whether the BSs in an MC system cooperate or not, MC precoding techniques can be classified as noncooperative precoding and cooperative precoding [1]. The noncooperative MC precoding techniques are reviewed in this section, while the cooperative counterparts are discussed in Section IV.

A. Multicell Downlink System Model

For illustration purpose, Fig. 3 shows a noncooperative three-cell downlink massive MIMO system. More generally, we consider the downlink of an MC system with $L > 1$ cells. The BS in each cell has M antennas and serves K users. Similar to the SC scenario, the TDD protocol is used, where the BSs acquire instantaneous CSI from uplink by exploiting channel reciprocity. We assume that the pilot signaling and data transmission take place simultaneously in all cells. The received complex baseband signal $y_{j,m} \in \mathbb{C}$ at the mth user in the jth cell is [26]

$$y_{j,m} = \sum_{l=1}^{L} \mathbf{h}_{l,j,m}^T \mathbf{x}_l + v_{j,m} \tag{20}$$

where $\mathbf{x}_l \in \mathbb{C}^{M\times 1}$ is the transmit signal vector from the lth BS, $\mathbf{h}_{l,j,m} \in \mathbb{C}^{M\times 1}$ is the channel vector from the lth BS to the mth user in the jth cell, and $v_{j,m} \sim \mathcal{CN}(0, \sigma^2)$ is the zero-mean AWGN with variance σ^2. The small-scale channel fading and Rayleigh block-fading conditions can be found in [26]. All BSs use Gaussian codebooks and linear precoding. The precoding of the transmit data symbols in the model (20) is then given by

$$\mathbf{x}_l = \mathbf{W}_l \mathbf{s}_l \tag{21}$$

where $\mathbf{W}_l \in \mathbb{C}^{M\times K}$ is the precoding matrix for the lth cell, and $\mathbf{s}_l \in \mathbb{C}^{K\times 1}$ is the transmitted symbol vector. Note that $\mathbf{W}_l$ is controlled by the transmit power constraint

$$\frac{1}{K}\,\mathrm{tr}(\mathbf{W}_l \mathbf{W}_l^H) = P_l \tag{22}$$

where P_l is the average transmit power per user in the lth cell. Using (20) and (21), the received signal vector at the jth cell can be expressed as

$$\begin{aligned}\mathbf{y}_j &= \sum_{l=1}^{L} \mathbf{H}_{l,j}^T \mathbf{W}_l \mathbf{s}_l + \mathbf{v}_j \\ &= \underbrace{\mathbf{H}_{j,j}^T \mathbf{W}_j \mathbf{s}_j + \mathbf{v}_j}_{\text{Equivalent SC}} + \underbrace{\sum_{l=1, l\neq j}^{L} \mathbf{H}_{l,j}^T \mathbf{W}_l \mathbf{s}_l}_{\text{Intercell interference}}\end{aligned} \tag{23}$$

where $\mathbf{H}_{l,j} = [\mathbf{h}_{l,j,1}, \mathbf{h}_{l,j,2}, \ldots, \mathbf{h}_{l,j,K}] \in \mathbb{C}^{M\times K}$. In (23), the left portion is an SC model equivalent to (2), and the right portion is the intercell interference.

B. Basic Precoding Techniques

The precoding techniques discussed in Section II for the SC scenario, i.e., MF (or MRT), ZF, RZF (or MMSE), TPE, and PZF, can also be applied to MC scenarios. In the noncooperative case, it is equivalent to applying SC linear precoding techniques to (20) and (21), by ignoring the intercell interference term in (23) at the receiver side. A lot of efforts have been devoted to improving precoder design without resort to cooperation among cells. An MMSE-based precoding method is developed in [46] to mitigate the problem of pilot contamination. A full-pilot ZF scheme is proposed in [10] to suppress intra- and intercell interference, where each BS only needs the local CSI to perform precoding. This scheme yields a closed-form solution, which takes a similar form to the RZF precoder, involving matrix inversion and regularization. The impact of pilot contamination in massive MIMO systems is studied in [47] and [48], where the performance limits and asymptotic SINR are derived. Results show that the RZF precoder can obtain an interference suppression gain, leading to higher asymptotic performance. The power allocation for RZF precoder is investigated in [24], and it is shown that the power is only transmitted to the stronger user channel when channel correlation is greater than 0.7. Hence, as the correlation gets higher, RZF precoder eventually approaches the single-user scenario. Another RZF-based precoder for MC scenario is proposed in [29], which is termed as interference-aware RZF precoder. Such a design establishes a tradeoff between achievable sum rate and interference rejection. The TPE precoder for MC case is discussed in [49] to alleviate the computational complexity of matrix inversion existing in RZF precoder. On the contrary to the SC case, TPE precoding outperforms the RZF precoding in terms of both complexity and throughput in MC massive MIMO regime [49]. In addition to its low computational complexity, TPE achieves better throughput in the MC scenario due to tractable offline optimization of the polynomial coefficients, which allows an enhanced coordination of intracell and intercell interference. This cannot be achieved through RZF preceding in the MC scenario.

C. Advanced Precoding Techniques

Several advanced precoding techniques for MC massive MIMO systems are reviewed in the following sections.

1) H-Infinity (H-inf) Precoding: An optimization-based precoding method, called H-infinity (H-inf) precoding, is proposed in [50]. The core idea of this method is to find a precoding matrix $\mathbf{W}_j$ so that the ratio between the precoding estimate error in the jth cell and the whole interference from all other cells is less than a prescribed bound s_p. The objective function is expressed as [50]

$$\sup_{\mathbf{H}_{l,j}^T \mathbf{W}_l + v_j} \frac{\left\| \mathbf{H}_{j,j}^T \mathbf{W}_j \mathbf{s}_j - \mathbf{s}_j \right\|_w^2}{\left\| \sum_{l=1, l\neq j}^{L} \mathbf{H}_{l,j}^T \mathbf{W}_l \mathbf{s}_l + \mathbf{v}_j \right\|^2} < s_p. \tag{24}$$

Using the obtained precoding matrix $\mathbf{W}_j$, the achievable average rate per user can be substantially improved over ZF and RZF methods.

2) Max Signal-to-Interference-and-Noise Ratio (Max-SINR): In [51] and [52], the Max-SINR algorithm is proposed for an MC massive MIMO system, aiming to increase the SINR of the system by maximizing the ratio of signal to intercell interference and noise. Assume that P is the maximum power per BS and ϕ is the utilization rate of the transmission power, where $\phi \in [0, 1]$. Under the condition that $\mathrm{SINR}_{j,m}$ of user m in cell j is greater than the desired threshold $\gamma_{j,m}$, the downlink precoding should maximize the utilization rate of the transmission power and minimize the total power (i.e., ϕP) consumption. The objective function is [52]

$$\begin{aligned} &\min_{\phi, \mathbf{w}_{j,m}} \ \phi P \\ &\text{s.t.} \ \sum_{m=1}^{K} |\mathbf{w}_{j,m}|^2 \leq \phi P \ \text{and} \ \mathrm{SINR}_{j,m} \geq \gamma_{j,m}. \end{aligned} \tag{25}$$

Numerical results show that the Max-SINR precoding algorithm outperforms the traditional massive MIMO SC-MF precoding algorithm. Under severe pilot contamination with large M, the SE of the proposed algorithm is 1.9 times higher than the MF algorithm.

3) Multilayer (ML) Approach: A multilayer (ML) precoding approach is proposed in [53] with BSs equipped with two-dimensional (2-D) antenna arrays. Let the numbers of horizontal and vertical antennas be M_H and M_V, respectively. Then, the precoding matrix in the lth cell is expressed as

$$\mathbf{W}_l = \mathbf{W}_l^A \circ \mathbf{W}_l^E \in \mathbb{C}^{M_H M_V \times K} \tag{26}$$

where $\mathbf{W}_l^A \in \mathbb{C}^{M_H \times K}$ is the Azimuth precoding component, $\mathbf{W}_l^E \in \mathbb{C}^{M_V \times K}$ is the elevation precoding component, and the operation $\circ$ denotes the Khatri–Rao product. The scope of this section is to apply the idea of multilayer precoding in the elevation direction in order to leverage the low-rank property of the elevation covariance matrix. The elevation precoding component can be further decomposed into multiple layers. A three-layer decomposition is provided in [53] as follows:

$$\mathbf{W}_l^E = \mathbf{W}_l^{E(1)} \mathbf{W}_l^{E(2)} \mathbf{W}_l^{E(3)}. \tag{27}$$

In this design, $\mathbf{W}_l^{E(1)} \in \mathbb{C}^{M_V \times K}$ is used to suppress intercell interference, $\mathbf{W}_l^{E(2)} \in \mathbb{C}^{K \times K}$ is used as a beamforming matrix to maximize signal power, and $\mathbf{W}_l^{E(3)} \in \mathbb{C}^{K \times K}$ is used to suppress multi-user interference. Analysis results have demonstrated that the multilayer precoding technique has better coverage gain over conventional solutions [53].

4) Hybrid Analog–Digital Approach: Hybrid analog–digital precoding architecture has been studied in [54], which contains a low-dimensional digital precoder $\mathbf{W}_D$ followed by an RF precoder $\mathbf{W}_{\mathrm{RF}}$ implemented using analog phase shifters. The problem of interest is to maximize the overall SE under total transmit power constraint, assuming perfect knowledge of the channel. Specially, it is aimed to find the optimal hybrid precoders at the BS and the optimal hybrid combiners for each user by [54]

$$\begin{aligned} &\max_{\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_D, \mathbf{V}_{\mathrm{RF}}, \mathbf{V}_D} \sum_{m=1}^{K} \beta_m R_m \\ \text{s.t.} \quad & \mathrm{tr}\left(\mathbf{W}_{\mathrm{RF}} \mathbf{W}_D \mathbf{W}_{\mathrm{RF}}^H \mathbf{W}_D^H \right) \leqslant P_t, \\ & |\mathbf{W}_{\mathrm{RF}}(c,d)|^2 = 1 \quad \forall c, d, \\ & |\mathbf{V}_{\mathrm{RF}_m}(c,d)|^2 = 1 \quad \forall c, d, m \end{aligned} \tag{28}$$

where P_t is the total power budget at the BS, and β_m and R_m represent the priority and overall rate of user m, respectively. The user m first processes the received signals using an RF combiner $\mathbf{V}_{\mathrm{RF}}$ implemented via phase shifters and then down-converts the signals to the baseband RF chains. Finally, using a low-dimensional digital combiner $\mathbf{V}_D$, the final processed signals are obtained. It is demonstrated that under much fewer number of RF chains, this precoding architecture can approach the performance of a fully digital scheme. Furthermore, with even less RF chains, some heuristic algorithms are proposed to tackle the problem of overall SE maximization for point-to-point and multiple user systems.

Table III summarizes the features of the advanced MC precoding techniques. Since the advanced precoding techniques are designed with complicated environment and conditions, it is difficult to compare them in a unified framework. Therefore, we describe only the features of each advanced MC technique.

TABLE III
FEATURES OF ADVANCED MC PRECODING TECHNIQUES

Advanced MC precoding techniques	Features
H-infinity	Guaranteed average rate over ZF and RF, high complexity due to pseudo-inverse operations
Max-SLNR	Guaranteed energy efficiency, medium complexity
Multilayer	Able to handle intercell and multi-user interference, low complexity
Hybrid analog–digital	Guaranteed rate, low complexity

D. Achievable Rates

By decomposing the received signal vector in (20), the received downlink symbol at the mth user in the jth cell can be written as

$$\begin{aligned} y_{j,m} &= \sum_{l=1}^{L}\sum_{k=1}^{K} \mathbf{h}_{l,j,m}^{T}\mathbf{w}_{l,k}s_{l,k} + v_{j,m} \\ &= E\left(\mathbf{h}_{j,j,m}^{T}\mathbf{w}_{j,m}\right)s_{j,m} \\ &\quad + \left[\mathbf{h}_{j,j,m}^{T}\mathbf{w}_{j,m} - E\left(\mathbf{h}_{j,j,m}^{T}\mathbf{w}_{j,m}\right)\right]s_{j,m} \\ &\quad + \sum_{(l,k)\neq(j,m)} \mathbf{h}_{l,j,m}^{T}\mathbf{w}_{l,k}s_{l,k} + v_{j,m} \end{aligned} \tag{29}$$

where $\mathbf{w}_{l,k}$ is the kth column in $\mathbf{W}_l = [\mathbf{w}_{l,1}, \mathbf{w}_{l,2}, \ldots, \mathbf{w}_{l,K}]$, and $s_{l,k}$ is the kth symbol in $\mathbf{s}_l = [\mathbf{s}_{l,1}, \mathbf{s}_{l,2}, \ldots, \mathbf{s}_{l,K}]$. The ergodic achievable rate is then given by [26]

$$R_{j,m} = \log_2(1 + \mathrm{SINR}_{j,m}) \tag{30}$$

where (after some manipulations [26])

$$\mathrm{SINR}_{j,m} = \frac{\left|E\left(\mathbf{h}_{j,j,m}^{T}\mathbf{w}_{j,m}\right)\right|^2}{\sum\limits_{l,k} E\left[\left|\mathbf{h}_{l,j,m}^{T}\mathbf{w}_{l,k}\right|^2\right] - \left|E\left(\mathbf{h}_{j,j,m}^{T}\mathbf{w}_{j,m}\right)\right|^2 + \sigma^2}. \tag{31}$$

In [55], the asymptotic approximations of the achievable rates for MF and RZF precoders are derived under the conditions $M \to \infty$ and $K/M \to 0$. The results indicate that with RZF, the number of RF antennas can be reduced by one order of magnitude to achieve the same performance as the simple MF scheme. Similar studies have been reported in [43], where the deterministic analytical results are easily computable for practical values of M and K. It has also been noted that ZF precoder has a similar performance to RZF precoder under delayed CSI. The approximations of the achievable rate for TPE precoder are studied in [49]. The average rate increases with the increase of antennas. Another feature of the asymptotic SINR for TPE precoder is that it only depends on the statistics of the channel but not on the instantaneous CSI. Thus, it can be optimized in an offline manner, which is an advantage over other techniques.

IV. Cooperative Multicell Precoding Techniques

In MC MIMO systems, precoding through BS cooperation is an effective approach to reduce the intercell interference [56]–[58], at the cost of cooperation overhead and increasing the complexity of precoder design and implementation. The BS cooperation is usually supported by a backhaul network [59]. Fig. 4 shows a cooperative three-cell downlink massive MIMO system. There are two representative BS cooperation schemes: partial cooperation and full cooperation. In partial cooperation, BSs share the CSI of active users by offering a balance between ensuring a reasonable load on the backhaul and attaining the performance gain using cooperation. In full cooperation, BSs exchange a great amount of information, including CSI and transmit and precoding data. A qualitative comparison of the achievable capacity of partially cooperative and fully cooperative massive MIMO systems is shown in [60].

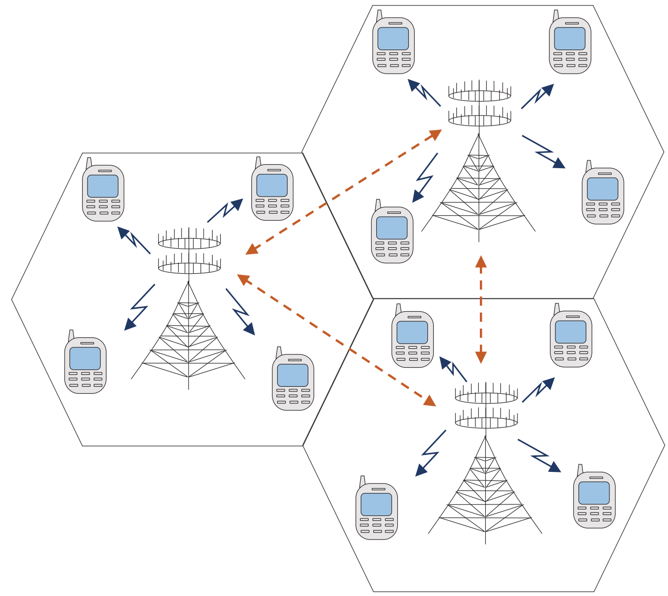

Fig. 4. Cooperative three-cell downlink massive MIMO system.

A. Partial Cooperation

A partial cooperation technique with only CSI sharing, termed as "interference coordination level" of BS collaboration, is proposed in [45]. The partial cooperation among BSs can leverage the shared CSI to design joint precoding matrices for exclusive downlink transmission [60]. Precoding matrices for all BSs can be designed as the minimization of the MSEs of all users in the system [61]. The expression of the expected sum rate under Rician CSI, which has been derived in [62], can be used for centralized precoding optimization. In practice, the computational complexity and backhaul signaling often make it unaffordable to derive optimal precoding for the system. But luckily, for partial BS cooperation, heuristic precoding strategies like the layered virtual SINR (LV-SINR) maximizing precoding technique show very good performance. The LV-SINR maximizing beamformers are obtained by, for all j, m, solving [62]

$$\mathbf{w}_{j,m}^{\text{LV-SINR}} = \arg\max_{\|\mathbf{w}\|} \frac{E(|\mathbf{h}_{j,m}^{H}\mathbf{w}|^2)}{\frac{\sigma^2}{\rho_{j,m}} + \sum_{\bar{m}\neq m} E(|\mathbf{h}_{j,\bar{m}}^{H}\mathbf{w}|^2)} \tag{32}$$

where $\|\mathbf{w}\| = 1$ and the power allocation coefficients are $\rho_{j,m}$.

Based on BS partial cooperation, a two-stage precoding scheme is proposed in [63]. The first stage includes a spaced partitioning codebook whose beam is focused on a certain coverage area, acting as an outer precoder. The second stage acts as an inner precoder exploiting dimension-reduced instantaneous CSI in order to improve the performance of the user end. For the hierarchical method proposed in [64], the outer precoder uses the spatial DoF to vanish the intercell interference by exploiting the average CSI only. Meanwhile, the inner precoder helps the transmission by using the time-variant CSI. In [65], the multi-user downlink beamforming in MC massive MIMO is addressed, with the objective of having high SE for cell-edge users. The impact of cooperation level on the performance with limited local CSI is investigated based on the vector perturbation technique. Another well-known cooperative precoding method is based on a nested structure [66], [67]. Such nested precoders provide a tradeoff between decreasing the intercell interference and furnishing a high SINR for different users.

A dissimilar method for limited coordination is to group the BS clusters of small sizes to coordinate and perform joint processing. In this situation, the overall backhaul signaling is reduced by sharing user data signals among a small number of BSs.

B. Full Cooperation

As shown in [61], [62], and [68], achieving full collaboration among BSs increases the complexity of the system as the BSs need to exchange much information, e.g., CSI and transmit and precoding data. Nevertheless, it has various benefits as described in [60]. While the constraint of total transmit power is used in [60], a more realistic per BS power constraint is considered in [69]. In this system, the BSs fully cooperate to ensure that both data streams and CSI are available at all BSs. Consequently, the BSs pool their antennas together to jointly serve the users. That is, precoding for multi-user downlink is employed at each BS. Specifically, the ZF precoder is utilized in [69] for the sake of simplicity and the case of two cooperative BSs is considered. The total power minimization with sum power constraint ($\min_{p\geq 0} \sum_{l=1}^{2} \sum_{m=1}^{K} p_{l,m}$) is formulated as [69]

$$\min_{p\geq 0} \sum_{l=1}^{2} \sum_{m=1}^{K} p_{l,m}, \ \text{s.t.} \left\| \sum_{l=1}^{2} \frac{\sqrt{p_{l,m}}}{\|\mathbf{f}_{l,m}\|} \right\|^2 \geq \gamma_m, \quad 1 \leq m \leq K \tag{33}$$

where $p_{l,m}$ is the transmission power at the lth BS for the mth user, γ_m is the target SINR, and $\mathbf{f}_{l,m}$ is the normalized version of the transmit precoding $\mathbf{w}_{l,m}$. Similarly, with per BS power constraint, the power minimization problem is formulated as [69]

$$\begin{aligned} &\min_{p\geq 0} \sum_{l=1}^{2} \sum_{m=1}^{K} p_{l,m} \\ &\text{s.t.} \left\| \sum_{l=1}^{2} \frac{\sqrt{p_{l,m}}}{\|\mathbf{f}_{l,m}\|} \right\|^2 \geq \gamma_m, \quad 1 \leq m \leq K \\ &\sum_{m=1}^{K} p_{l,m} \leq P_{\text{av}}, \quad 1 \leq l \leq 2 \end{aligned} \tag{34}$$

where P_{av} is the available BS power.

Extensive discussions on how BSs could cooperate to deal with pilot contamination and interference are provided in [70], by considering a series of aspects such as BS clustering, ZF precoder with intercluster interference mitigation, and frequency reuse. For the sake of simplicity, a one-dimensional (1-D) cell layout is used to illustrate the essential idea behind. A method used to tailor the precoding matrix in a full cooperation scenario is to impose additional constraints on the ZF precoder so that the intercluster interference from neighboring clusters could be mitigated. Suppose two neighboring BSs are grouped as a cluster, then the cluster-wise channel matrix (as to distinguish with BS-wise channel matrix $\mathbf{H}_{l,j}$) is denoted by $\bar{\mathbf{H}}_{c_1,c_2}(f) \in \mathbb{C}^{2M\times K}$, where c_1 and c_2 are the indices of clusters, and f is the frequency reuse index. Assume the users in cluster c require the downlink communication to mitigate the neighboring intercluster interference from cluster $c-1$ and $c+1$, then the cooperative precoding matrix is given by

$$\mathbf{W}_{c,c}(f) = \left[\dddot{\mathbf{H}} \left(\dddot{\mathbf{H}}^T \dddot{\mathbf{H}} \right)^{-1} \right]_1^K \tag{35}$$

followed by column norm normalization, where the composite (cooperative) matrix $\dddot{\mathbf{H}}$ is constructed as

$$\dddot{\mathbf{H}} = \left[\underbrace{\hat{\mathbf{H}}_{c,c}(f)}_{2M\times K} \ \underbrace{\left[\hat{\mathbf{H}}_{c-1,c}(f)\right]_1^{K/2}}_{2M\times \frac{K}{2}} \ \underbrace{\left[\hat{\mathbf{H}}_{c+1,c}(f)\right]_1^{K/2}}_{2M\times \frac{K}{2}} \right]_1^K \tag{36}$$

where $[\cdot]_1^K$ extracts the first K columns in a matrix, and $\hat{\mathbf{H}}$ is the estimate of $\bar{\mathbf{H}}$ in uplink training by exploiting channel reciprocity. Note that (36) is the simplest case in which only the 1-D neighboring clusters are taken into consideration. More generally, it could be extended for more interfering clusters or even all of the rest clusters, as well as 2-D layout scenarios.

A big challenge of full BS cooperation is the promised gain, which depends on the assumptions of perfect CSI and ideal backhaul links, while decreasing the overhead of cooperation [71]. In practice, the backhaul network is bandwidth limited due to the costs of establishing high capacity links. It restricts the amount of information that can be exchanged among BSs, which determines the level of BS cooperation [69].

The work in [71] derives the achievable rate bounds for ZF and MF precoders with vector or matrix normalization for fully cooperative massive MIMO. It is shown that while normalization is suitable for both MF and ZF precoding, vector normalization is better for ZF precoding and matrix normalization is better for MF precoding. In [60], the performance of massive MIMO systems with fully cooperative BSs and centralized BSs is compared, where ZF precoders with 20 transmit antennas are used. The results show that the fully cooperative BS deployment clearly outperforms the centralized one. Moreover, ZF precoder power allocation under full cooperation scheme is addressed in [72]. It is shown that the proposed ZF precoding algorithm gives near optimal performance whereas satisfying the per BS SINR constraints. However, due to limited feedback, it cannot be applied to large-scale array. Therefore, implementing massive MIMO in fully cooperative BS system is complex.

V. Discussions

The inherent difficulties in massive MIMO systems are caused by pilot contamination and intercell interference. Normally, exploring advanced precoding techniques could lead to possible solutions but obtaining an efficient one is not an easy task. In this section, we first discuss some existing challenges, which are related to the design of precoding mechanisms and practical implementations. Then, we point out some research potentials for precoding in massive MIMO.

A. Challenges

1) Imperfect CSI: The performance of all precoding strategies is crucially dependent on the accuracy of CSI. In TDD-based practical systems, a BS has to attain the CSI of the downlink channel by leveraging channel reciprocity. Meanwhile, the channel coherence time is limited. Thus, the information of the prompt channel state is fundamentally incomplete, which is the main barrier of the precoding strategies. So far, only few research works have been carried out to discover the impact of imperfect CSI on the system behavior [30]. In [40], a dual-structured linear precoding approach is presented, where a preprocessing based on the spatial correlation and a subsequent linear precoding based on the short-term CSI are concatenated to reduce the feedback overhead. Its performance under imperfect CSI is asymptotically analyzed based on random matrix theory. In [73], the impact of multi-user interference on rate performance is discussed in the situation of imperfect CSI. However, so far, the study on precoding performance under imperfect CSI has been preliminary and novel precoding techniques based on partial CSI need to be developed.

2) Computational Complexity: The computational complexity of most state-of-the-art linear precoding schemes is excessively high in the massive MIMO regime [26]. As shown in [36], ZF and RZF precoders can attain the performance level of MF precoder with an order of magnitude fewer antennas. Nevertheless, the number of antennas is still large with respect to the user population, in addition to the concern that conducting precoding requires rapid and flexible connections between local baseband processors [36], [74]. Besides, the RZF precoder is undesirable as its complexity scales in polynomial fashion, according to the cube power of the number of antennas [36]. As for the ZF precoder, it needs to compute the pseudo-inverse of very large matrices, which results in high computational complexity.

3) Number of Served Users: Another surprising challenge is that fewer users could be served for the ZF precoder. To enable the capacity of serving more users, grouping strategies have been introduced in [51] but reliable strategies still have not been well established. In addition, when the MF precoder is used in a scenario with excessive users, it does not actively suppress residual interuser interference [49]. Therefore, the optimal linear precoding is still unknown in massive MIMO regime.

4) Application-Related Challenges: a) advanced precoding techniques can effectively mitigate the pilot contamination problem in the MC scenario, practical implementation aspects have not been well addressed yet [75], [76], which is likely to be a worthy research topic. b) Another practical concern lies in the question of how many antennas do we exactly need, although theory has been established given the number of antennas approaching infinity. In practice, a limit must be imposed on the size of the antenna array [74]. Note that antennas are not in themselves costly, but the very large antenna arrays with hundreds or thousands of RF front-ends, A/D, and D/A converters are massive and energy-hungry [36]. c) Under different channel conditions, the attainable performance is different. In this paper, the independent identically distributed Rayleigh channel model is assumed for different precoding strategies. The consequence of hiring more antennas has different effects on different channel models, which is an application-related challenge. d) When BSs collaborate, the joining backhaul links are not cost-effective, which throws several challenges among the levels of BS cooperation. These challenges include who is providing the backhaul, how big should the cluster size be for cooperating BSs, and what information should be shared among the BSs? All these important questions need to be answered. e) Finally, because massive MIMO is one of the main candidates for 5G cellular communications, system development is further needed to have real-time test beds in order to explore limitations and possibilities of massive MIMO technology. There are some existing test beds for massive MIMO, as shown in Table IV. A practical study based on a 128-antenna test bed is provided in [77], yet more practical testings of real-time hierarchical hardware architecture of massive MIMO systems need to be conducted. The main challenge to do so is the hardware bottleneck and complicated real-time requirement in both uplink and downlink hardware phases.

TABLE IV
EXISTING TEST BEDS FOR MASSIVE MIMO SYSTEMS

Institution	Band	Hardware	Users	BS size
Lund	2.6	RUSK channel sounder	6	128
Samsung	1–28	Proprietary	N/A	64
Rice	2.4	WARP, Power PC	15	64

B. Research Potentials

Most research works on massive MIMO have so far considered a single-antenna user scenario. The extension to multi-antenna users is needed too. Moreover, many results are derived from a Rayleigh block fading channel model for simplified analysis, but other channel models, e.g., jointly correlated Rician fading channel model, have not been rigorously considered yet. Therefore, the generalization of communication scenarios and channel conditions in the massive MIMO system analysis is worth of further research attention.

Another issue, system security, could also be considered in system designs. Some works on massive MIMO security have been seen in [78]–[80]. Moreover, in [13], [81], and [82], the artificial noise-aided jamming is incorporated to address the security issue with the use of a ZF precoder. It is also important to note that the ongoing research works on massive MIMO mainly focus on the physical layer and little work has been done on higher layers. For instance, when a new user is added to the network, the user needs to assign a new pilot and the nursery handshaking task needs to be done in order for the BS to utilize the full selectivity of the antenna array [2].

Furthermore, an efficient precoder has to maintain a good tradeoff between performance and complexity. Unfortunately, low complexity precoders do not show good performance compared with more complicated ones. However, a complicated precoder design is more difficult to implement practically. Actually, a viable design for the precoding technique is unknown to date.

VI. Conclusion

This paper provided a survey of the linear precoding techniques for massive MIMO systems under both SC and MC scenarios. To the authors' knowledge, this is the first survey paper that comprehensively studies downlink precoding techniques. In the SC scenario, a series of techniques have offered high flexibility in trading-off system efficiency and computational complexity. In the MC scenario, the inevitable reuse of pilot signal introduces pilot contamination problem, which degrades the estimation of CSI. The solutions based on noncooperative precoding techniques can somewhat mitigate this effect. In comparison, the cooperative techniques (partial or full cooperation) can yield better performance at the price of incurring more computational complexity. Existing challenges and future research potentials are also discussed in this paper.

References

[1] 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.

[2] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, Mar. 2015.

[3] V. P. Selvan, M. S. Iqbal, and H. S. Al-Raweshidy, "Performance analysis of linear precoding schemes for very large multi-user MIMO downlink system," in *Proc. 4th Int. Conf. Innov. Comput. Technol.*, Luton, U.K., 2014, pp. 219–224.

[4] P. Xingdong, H. Wei, Y. Tianyang, and L. Linsheng, "Design and implementation of an active multibeam antenna system with 64 RF channels and 256 antenna elements for massive MIMO application in 5G wireless communications," *IEEE Trans. Commun.*, vol. 11, no. 11, pp. 16–23, Nov. 2014.

[5] F. Rusek *et al.*, "Scaling up MIMO—Opportunities and challenges," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. 51th Annu. Allerton Conf.*, Monticello, IL, USA, 2013, pp. 293–298.

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

[9] L. Zhao, K. Zheng, H. Long, and H. Zhao, "Performance analysis for downlink massive MIMO system with ZF precoding," *Trans. Emerging Telecommun. Technol.*, vol. 25, no. 12, pp. 1219–1230, Dec. 2014.

[10] E. Bjornson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016, doi: 10.1109/TWC.2015.2488634.

[11] C. Sun, X. Gao, S. Jin, and M. Matthaiou, "Beam division multiple access for massive MIMO downlink transmission," in *Proc. 2015 IEEE Int. Conf. Commun.*, London, U.K., 2015, pp. 1970–1975.

[12] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.

[13] J. Zhu, R. Schober, and V. K. Bhargava, "Secure downlink transmission in massive MIMO system with zero-forcing precoding," in *Proc. 20th Eur. Wireless Conf.*, Barcelona, Spain, 2014, pp. 1–6.

[14] A. Muller, A. Kammounz, E. Bjornson, and M. Debbahx, "Efficient linear precoding for massive MIMO systems using truncated polynomial expansion," in *Proc. 8th Sens. Array Multichannel Signal Process. Workshop*, A Coruña, Spain, 2014, pp. 273–276.

[15] L. Zhao, K. Zheng, H. Long, H. Zhao, and W. Wang, "Performance analysis for downlink massive multiple-input multiple-output system with channel state information delay under maximum ratio transmission precoding," *IET Commun.*, vol. 8, no. 3, pp. 390–398, Feb. 2014.

[16] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[17] 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, Mar. 2005.

[18] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec. 2004.

[19] S. Buzzi, C. L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[20] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[21] D. C. Arajo, T. Maksymyuk, A. L. F. de Almeida, T. Maciel, J. C. M. Mota, and M. Jo, "Massive MIMO: Survey and future research topics," *IET Commun.*, vol. 10, no. 15, pp. 1938–1946, Oct. 2016.

[22] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surv. Tuts.*, vol. 17, no. 4, pp. 1941–1988, Nov. 2015.

[23] R. Chen, "Multiuser MIMO systems in single-cell and multi-cell wireless communication," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Texas at Austin, Austin, TX, USA, 2007.

[24] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. 2011 IEEE Veh. Technol. Conf.*, San Francisco, CA, USA, 2011, pp. 1–5.

[25] T. Parfait, Y. Kuang, and K. Jerry, "Performance analysis and comparison of ZF and MRT based downlink massive MIMO systems," in *Proc. 6th Int. Conf. Ubiquitous Future Netw.*, Shanghai, China, 2014, pp. 383–388.

[26] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 861–875, Oct. 2014.

[27] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 3059–3075, Mar. 2015.

[28] J. He, "Precoding and equalization for MIMO broadcast channels with applications in spread spectrum systems," Ph.D. dissertation, Dept. Elect. Comput. Eng., Northeastern Univ., Boston, MA, USA, 2010.

[29] A. Muller, R. Couillet, E. Bjornson, S. Wagner, and M. Debbah, "Interference-aware RZF precoding for multicell downlink systems," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3959–3973, Aug. 2015.

[30] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Mar. 2012.

[31] R. Stridh, M. Bengtsson, and B. Ottersten, "System evaluation of optimal downlink beamforming with congestion control in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 743–751, Apr. 2006.

[32] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4298–4310, Aug. 2010.

[33] 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, Aug. 2005.

[34] A. H. Mehana and A. Nosratinia, "Diversity of MIMO linear precoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1019–1038, Nov. 2013.

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

[36] J. Hoydis, S. 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.

[37] L. Sanguinetti, E. Bjornson, M. Debbah, and A. L. Moustakas, "Optimal linear precoding in multi-user MIMO systems: A large system analysis," in *Proc. 2014 IEEE Global Commun. Conf.*, Austin, TX, USA, 2014, pp. 3922–3927.

[38] S. Shi and M. Schubert, "MMSE transmit optimization for multi-user multi-antenna systems," in *Proc. 30th IEEE Int. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, USA, 2005, pp. 409–412.

[39] Y. Mehmood, W. Afzal, F. Ahmad, U. Younas, I. Rashid, and I. Mehmood, "Large scaled multi-user MIMO system so called massive MIMO systems for future wireless communication networks," in *Proc. 19th Int. Conf. Autom. Comput.*, London, U.K., 2013, pp. 13–14.

[40] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2532–2547, Jan. 2015.

[41] A. Muller, A. Kammoun, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, 2016, Art. no. 63.

[42] S. Zarei, W. Gerstacker, and R. Schober, "A low-complexity linear precoding and power allocation scheme for downlink massive MIMO systems," in *Proc. 2013 Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 2013, USA, pp. 285–290.

[43] A. Papazafeiropoulos and T. Ratnarajah, "Linear precoding for downlink massive MIMO with delayed CSIT and channel prediction," in *Proc. 2014 Wireless Commun. Netw. Conf.*, Istanbul, Turkey, 2014, pp. 809–914.

[44] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[45] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[46] 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.

[47] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Cellular systems with many antennas: Large system analysis under pilot contamination," in *Proc. 50th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, 2012, pp. 1220–1224.

[48] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4360–4373, Apr. 2014.

[49] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Low-complexity linear precoding for multi-cell massive MIMO systems," in *Proc. 22nd Eur. Signal Process. Conf.*, Lisbon, Portugal, 2014, pp. 2150–2154.

[50] P. Xu, J. Z. Wang, J. Wang, and F. Qi, "Multi-cell H-inf precoding in massive MIMO systems," in *Proc. 2014 IEEE Int. Conf. Commun.*, Sydney, Australia, 2014, pp. 4472–4476.

[51] M. Kurras, L. Raschkowski, M. Talaat, and L. Thiele, "Massive SDMA with large scale antenna systems in a multi-cell environment," in *Proc. 2013 IEEE Africon Conf.*, Pointe-Aux-Piments, Mauritius, 2013, pp. 1–5.

[52] J. Jing and X. Zheng, "A downlink max-SINR precoding for massive MIMO system," *Int. J. Future Gener. Commun. Netw.*, vol. 7, no. 3, pp. 107–116, Jun. 2014.

[53] A. Alkhateeb, G. Leus, and R. W. Heath, "Multi-layer precoding for full-dimensional massive MIMO systems," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2014, pp. 815–819.

[54] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[55] J. Hoydis, S. Brink, and M. Debbah, "Comparison of linear precoding schemes for downlink massive MIMO," in *Proc. 2012 IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, 2012, pp. 2135–2139.

[56] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[57] S. A. Jafar, "Interference alignment–A new look at signal dimensions in a communication network," *Found. Trends Commun. Inf. Theory*, vol. 7, no. 1, pp. 1–134, Jan. 2011.

[58] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[59] R. W. Heath, "Multiple cell MIMO," Univ. Texas at Austin, Austin, TX, USA. [Online]. Available: http://www.profheath.org/mimo-communication/multiple-cell-mimo/

[60] S. Dierks, M. B. Amin, W. Zirwas, M. Haardt, and B. Panzner, "The benefit of cooperation in the context of massive MIMO," in *Proc. 18th Int. OFDM Workshop*, Essen, Germany, 2014, pp. 1–8.

[61] A. Chockalingam and S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[62] E. Bjornson and B. Ottersten, "On the principles of multicell precoding with centralized and distributed cooperation," in *Proc. 2009 Int. Conf. Wireless Commun. Signal Process.*, Nanjing, China, 2009, pp. 1–5.

[63] F. Zhang, Y. Huang, S. Jin, L. Jiang, and G. Wang, "Reduced-backhaul coordinated beamforming for massive MIMO heterogeneous networks," in *Proc. 2015 IEEE Wireless Commun. Netw. Conf.*, New Orleans, LA, USA, 2015, pp. 129–134.

[64] A. Liu and V. Lau, "Hierarchical interference mitigation for massive MIMO cellular networks," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4786–4797, Sep. 2014.

[65] J. Choi, "Multiuser precoding with limited cooperation for large-scale MIMO multicell downlink," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1295–1308, Mar. 2015.

[66] R. Zakhour and S. V. Hanly, "Min-max power allocation in cellular networks with coordinated beamforming," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 287–302, Feb. 2013.

[67] Y. Huang, C. W. Tan, and B. D. Rao, "Joint beamforming and power control in coordinated multicell: Max-min duality, effective network and large system transition," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2730–2742, Jun. 2013.

[68] I. G. Miguelez, E. Avdic, N. Marchetti, I. Macaluso, and L. Doyle, "Cloud-RAN platform for LSA in 5G networks—Tradeoff within the infrastructure," in *Proc. 6th Int. Symp. Commun., Controls Signal Process.*, Athens, GA, USA, 2014, pp. 522–525.

[69] J. Kang, J. Kang, N. Lee, B. M. Lee, and J. Bang, "Minimizing transmit power for cooperative multicell system with massive MIMO," in *Proc. 2013 IEEE Consum. Commun. Netw. Conf.*, Las Vegas, NV, USA, 2013, pp. 438–442.

[70] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

[71] C. Lee, C. B. Chae, T. Kim, S. Choi, and J. Lee, "Network massive MIMO for cell-boundary users: From a precoding normalization perspective," in *Proc. 2012 Int. Workshop Cloud Base-Station Large-Scale Cooperative Commun.*, Anaheim, CA, USA, 2012, pp. 233–237.

[72] E. Bjornson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Found. Trends Commun. Inf. Theory*, vol. 9, no. 3, pp. 113–381, Sep. 2013.

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

[74] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[75] C. Shepard *et al.*, "Argos practical many antenna base stations," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, Istanbul, Turkey, 2012, pp. 53–64.

[76] C. Shepard, H. Yu, and L. Zhong, "ArgosV2 a flexible many antenna research platform," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, New York, NY, USA, 2013, pp. 163–166.

[77] J. Vieira *et al.*, "A flexible 100-antenna testbed for massive MIMO," in *Proc. 2014 IEEE Globecom Workshops*, Austin, CA, USA, 2014, pp. 287–293.

[78] G. Geraci, R. Couillet, J. Yuan, M. Debbah, and I. B. Collings, "Large system analysis of linear precoding in MISO broadcast channels with confidential messages," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1660–1671, Sep. 2013.

[79] G. Geraci, S. Singh, J. G. Andrews, J. Yuan, and I. B. Collings, "Secrecy rates in broadcast channels with confidential messages and external eavesdroppers," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2931–2943, May 2014.

[80] J. Zhang, C. Yuen, C. K. Wen, S. Jin, and X. Gao, "Ergodic secrecy sum-rate for multiuser downlink transmission via regularized channel inversion: Large system analysis," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1627–1630, Sep. 2014.

[81] J. Wang, J. Lee, F. Wang, and T. Q. S. Quek, "Jamming-aided secure communication in massive MIMO Rician channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6854–6868, Dec. 2015.

[82] J. Zhu, R. Schober, and V. K. Bhargava, "Secure transmission in multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4766–4781, Sep. 2014.

Authors' photographs and biographies not available at the time of publication.