# An Overview of Massive MIMO: Benefits and Challenges

Lu Lu, *Student Member, IEEE*, Geoffrey Ye Li, *Fellow, IEEE*, A. Lee Swindlehurst, *Fellow, IEEE*, Alexei Ashikhmin, *Senior Member, IEEE*, and Rui Zhang, *Member, IEEE*

*Abstract*—Massive *multiple-input multiple-output* (MIMO) wireless communications refers to the idea equipping cellular *base stations* (BSs) with a very large number of antennas, and has been shown to potentially allow for orders of magnitude improvement in spectral and energy efficiency using relatively simple (linear) processing. In this paper, we present a comprehensive overview of state-of-the-art research on the topic, which has recently attracted considerable attention. We begin with an information theoretic analysis to illustrate the conjectured advantages of massive MIMO, and then we address implementation issues related to channel estimation, detection and precoding schemes. We particularly focus on the potential impact of pilot contamination caused by the use of non-orthogonal pilot sequences by users in adjacent cells. We also analyze the energy efficiency achieved by massive MIMO systems, and demonstrate how the degrees of freedom provided by massive MIMO systems enable efficient single-carrier transmission. Finally, the challenges and opportunities associated with implementing massive MIMO in future wireless communications systems are discussed.

*Index Terms*—Channel estimation, energy efficiency, massive MIMO systems, orthogonal frequency division multiplexing (OFDM), pilot contamination, precoding and detection, single-carrier transmission, spectral efficiency, time-division duplexing (TDD).

## I. INTRODUCTION

**M**ULTIPLE-INPUT multiple-output (MIMO) technology has been widely studied during the last two decades and applied to many wireless standards since it can significantly improve the capacity and reliability of wireless systems. While initial work on the problem focused on point-to-point MIMO links where two devices with multiple antennas communicate with each other, focus has shifted in recent years to more practical *multi-user MIMO* (MU-MIMO) systems, where typically a
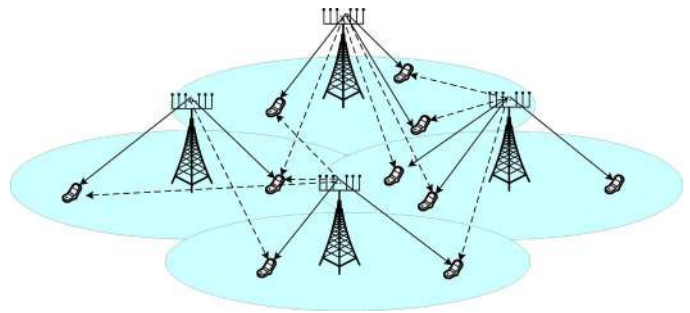
Fig. 1. Illustration of Massive MU-MIMO systems.

*base station* (BS) with multiple antennas simultaneously serves a set of single-antenna users and the multiplexing gain can be shared by all users. In this way, expensive equipment is only needed on the BS end of the link, and the user terminals can be relatively cheap single-antenna devices. Furthermore, due to multi-user diversity, the performance of MU-MIMO systems is generally less sensitive to the propagation environment than in the point-to-point MIMO case. As a result, MU-MIMO has become an integral part of communications standards, such as 802.11 (WiFi), 802.16 (WiMAX), LTE, and is progressively being deployed throughout the world. For most MIMO implementations, the BS typically employs only a few (i.e., fewer than 10) antennas, and the corresponding improvement in spectral efficiency, while important, is still relatively modest.

In a recent effort to achieve more dramatic gains as well as to simplify the required signal processing, massive MIMO systems or *large-scale antenna systems* (LSAS) have been proposed in [1], [2], where each BS is equipped with orders of magnitude more antennas, e.g., 100 or more. A massive MU-MIMO network is depicted in Fig. 1. Asymptotic arguments based on random matrix theory [2] demonstrate that the effects of uncorrelated noise and small-scale fading are eliminated, the number of users per cell are independent of the size of the cell, and the required transmitted energy per bit vanishes as the number of antennas in a MIMO cell grows to infinity. Furthermore, simple linear signal processing approaches, such as *matched-filter* (MF) precoding/detection, can be used in massive MIMO systems to achieve these advantages.

It is shown in [2] that under realistic propagation assumptions, MF-based non-cooperative massive MIMO systems could in principle achieve a data rate of 17 Mb/s for each of 40 users in a 20 MHz channel in both the uplink (reverse link) and downlink (forward link) directions, with an average throughput of 730 Mb/s per cell and an overall spectral efficiency of 26.5 bps/Hz. Since the number of antennas at the BS is typically assumed to

be significantly larger than the number of users, a large number of degrees of freedom are available and can be used to shape the transmitted signals in a hardware-friendly way or to null interference [3]. To make such a system practical, algorithms for massive MIMO systems are required to keep the complexity low.

Another advantage of massive MIMO lies in its potential energy efficiency compared to a corresponding single-antenna system. It is shown in [4] that each single-antenna user in a massive MIMO system can scale down its transmit power proportional to the number of antennas at the BS with perfect *channel state information* (CSI) or to the square root of the number of BS antennas with imperfect CSI, to get the same performance as a corresponding *single-input single-output* (SISO) system. This leads to higher energy efficiency and is very important for future wireless networks where excessive energy consumption is a growing concern [5], [6]. On the other hand, if adequate transmit power is available, then a massive MIMO system could significantly extend the range of operation compared with a single antenna system. Even though the conclusions in [4] ignore the power consumption of the radio front-end, massive MIMO is still a promising candidate for improving energy-efficiency of future networks.

The observations described above have recently sparked a flurry of research activities aimed at understanding the signal processing and information theoretic ramifications of massive MIMO system designs. In [7], massive MIMO systems are reviewed from various perspectives, including fundamental information theoretical gains, antenna and propagation aspects, and transceiver design. A follow-up tutorial [8] briefly discusses recent work. In this paper, we provide a more comprehensive and detailed overview of state-of-the-art research on this topic. In Section II, the opportunities of massive MIMO systems are viewed from an information theoretic perspective. Issues on channel estimation and signal detection are then discussed in Section III, and transmit precoding schemes are presented in Section IV. Besides the MF precoder/detector, other linear schemes such as *minimum mean-squared error* (MMSE) and *zero-forcing* (ZF) precoders/detectors are discussed based on either single-cell processing or multi-cell coordinated processing. In Section V, the so-called pilot contamination effect, caused by employing non-orthogonal pilot sequences at different users in different cells, is discussed in detail. The energy efficiency of massive MIMO systems is then analyzed in Section VI. Instead of using *orthogonal frequency division multiplexing* (OFDM) as in most MU-MIMO implementations today, the possibility of single-carrier modulation for massive MIMO systems is discussed in Section VII. Finally, the challenges and potentials related to applications of massive MIMO in future wireless communications are identified in Section VIII and conclusions are provided in Section IX.

*Notation*

Boldface lower and upper case symbols represent vectors and matrices, respectively. The transpose, conjugate, and Hermitian transpose operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, and $(\cdot)^{\mathrm{H}}$, respectively. The Moore-Penrose pseudoinverse operator is denoted by $(\cdot)^{\dagger}$. The determinant and trace operators are denoted by $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$, respectively. The norm of a vector is denoted by $\| \cdot \|$, and $a \gg b$ means $a$ is much greater than $b$.

## II. FROM REGULAR TO MASSIVE MIMO

In this section, the advantages of massive MIMO systems are reviewed from an information theoretic point of view. We start with point-to-point MIMO systems to reveal the potential opportunities that arise by equipping the terminals with a large number of antennas, and then we discuss the performance of MU-MIMO systems, where multiple single-antenna users are communicating with a BS equipped with a large number of antennas. Most results in this section are based on [7], [9], and [10].

### A. Point-to-Point MIMO

We consider a point-to-point MIMO transmission first, where the transmitter and the receiver are equipped with $N_t$ and $N_r$ antennas, respectively. We focus on the narrow-band time-invariant channel with a deterministic and constant channel matrix $\mathbf{H} \in \mathcal{C}^{N_r \times N_t}$. OFDM-based schemes are normally used to convert a frequency-selective wide-band channel into multiple parallel flat-fading narrow-band channels [11].

The received signal vector, $\mathbf{y} \in \mathcal{C}^{N_r \times 1}$, can be expressed as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n}, \qquad (1)$$

where $\mathbf{x} \in \mathcal{C}^{N_t \times 1}$ is the transmit signal vector and $\mathbf{n} \in \mathcal{C}^{N_r \times 1}$ represents noise and interference. We focus on the case that the total power of the transmit signal is normalized, i.e., $E\{\|\mathbf{x}\|^2\} = 1$, and the noise is zero-mean circularly symmetric complex Gaussian with an identity covariance matrix $\mathbf{I}$. With these assumptions, the scalar $\rho$ is the transmit power.

If we assume *independent and identically distributed* (i.i.d.) Gaussian transmit signals and that perfect CSI is available at the receiver, the instantaneous achievable rate can be expressed as

$$C = \log_2 \det \left( \mathbf{I} + \frac{\rho}{N_t}\mathbf{H}\mathbf{H}^{\mathrm{H}} \right) \; \frac{\text{bits}}{\text{s}}. \qquad (2)$$

When the propagation coefficients in the channel matrix $\mathbf{H}$ are normalized as $\mathrm{Tr}(\mathbf{H}\mathbf{H}^{\mathrm{H}}) \approx N_t N_r$, upper and lower bounds on the capacity are derived in [7] with the help of Jensen's inequality:

$$\log_2\left(1+\rho N_r\right) \le C \le \min(N_t, N_r) \log_2 \left( 1 + \frac{\rho \max(N_t, N_r)}{N_t} \right). \qquad (3)$$

The actual achievable rate depends on the distribution of the singular values of $\mathbf{H}\mathbf{H}^{\mathrm{H}}$. Among all channels with the same normalization, those whose singular values are all equal achieve the highest rate, i.e., the upper bound in (3), while those with only one non-zero singular value have the lowest rate, i.e., the lower bound in (3). The best case can be approached in the limit by a scenario where all of the propagation coefficients in the channel matrix are i.i.d., while the worst case corresponds for example to a scenario with *line-of-sight* (LOS) propagation.

Next we discuss two extreme cases, where either the number of transmit or the number of receive antennas goes to infinity.

*1) $N_t \gg N_r$ and $N_t \to \infty$:* When the number of transmit antennas goes to infinity while the number of receive antennas is constant, i.e., $N_t \gg N_r, N_t \to \infty$, the row vectors of $\mathbf{H}$ are asymptotically orthogonal, and hence we have

$$\frac{(\mathbf{H}\mathbf{H}^{\mathrm{H}})}{N_t} \approx \mathbf{I}_{N_r}. \tag{4}$$

In this case, the achievable rate in (2) can be approximated as

$$C \approx N_r \log_2(1 + \rho) \quad \frac{\text{bits}}{\text{s}}{\text{Hz}}, \tag{5}$$

which achieves the upper bound in (3).

*2) $N_r \gg N_t$ and $N_r \to \infty$:* Using similar derivations as in Case *1)*, we have

$$C \approx N_t \log_2\left(1 + \frac{\rho N_r}{N_t}\right) \quad \frac{\text{bits}}{\text{s}}{\text{Hz}}, \tag{6}$$

which achieves the upper bound in (3) as well.

The results in (5) and (6) show the advantages of equipping the arrays in a MIMO link with a large number of antennas. Note that the above discussion depends on the assumption that the row or the column vectors of $\mathbf{H}$ are asymptotically orthogonal, which is an optimistic assumption regarding the propagation coefficients. The multiplexing gain disappears in LOS propagation environments, as shown in [7].

### B. Multi-User MIMO

MU-MIMO systems can obtain the promising multiplexing gain of massive point-to-point MIMO systems while eliminating problems due to unfavorable propagation environments.

Consider a MU-MIMO system with $L$ cells, where each cell has $K$ served single-antenna users and one BS with $N$ antennas. Denote the channel coefficient from the $k$-th user in the $l$-th cell to the $n$-th antenna of the $i$-th BS as $h_{i,k,l,n}$, which is equal to a complex small-scale fading factor times an amplitude factor that accounts for geometric attenuation and large-scale fading:

$$h_{i,k,l,n} = g_{i,k,l,n} \sqrt{d_{i,k,l}}, \tag{7}$$

where $g_{i,k,l,n}$ and $d_{i,k,l}$ represent complex small-scale fading and large-scale fading coefficients, respectively. The small-scale fading coefficients are assumed to be different for different users or for different antennas at each BS while the large-scale fading coefficients are the same for different antennas at the same BS, but are user-dependent. Then, the channel matrix from all $K$ users in the $l$-th cell to the $i$-th BS can be expressed as

$$\mathbf{H}_{i,l} = \begin{pmatrix} h_{i,1,l,1} & \cdots & h_{i,K,l,1} \\ \vdots & \ddots & \vdots \\ h_{i,1,l,N} & \cdots & h_{i,K,l,N} \end{pmatrix} = \mathbf{G}_{i,l}\mathbf{D}_{i,l}^{1/2}, \tag{8}$$

where

$$\mathbf{G}_{i,l} = \begin{pmatrix} g_{i,1,l,1} & \cdots & g_{i,K,l,1} \\ \vdots & \ddots & \vdots \\ g_{i,1,l,N} & \cdots & g_{i,K,l,N} \end{pmatrix}, \tag{9}$$

$$\mathbf{D}_{i,l} = \begin{pmatrix} d_{i,1,l} & & \\ & \ddots & \\ & & d_{i,K,l} \end{pmatrix}. \tag{10}$$

Consider a single-cell ($L = 1$) MU-MIMO system with $K$ single-antenna users and a BS with $N$ antennas. For simplicity, the cell and the BS indices are dropped when single-cell systems are considered.

*1) Uplink:* For uplink signal transmission, the received signal vector at a single BS, which we denote by $\mathbf{y}_u \in \mathcal{C}^{N \times 1}$, has the same expression as in (1):

$$\mathbf{y}_u = \sqrt{\rho_u}\mathbf{H}\mathbf{x}_u + \mathbf{n}_u, \tag{11}$$

where $\mathbf{x}_u \in \mathcal{C}^{K \times 1}$ is the signal vector from all users, $\mathbf{H} \in \mathcal{C}^{N \times K}$ is the uplink channel matrix defined in (8) by dropping the cell and the BS indices, $\mathbf{n}_u \in \mathcal{C}^{N \times 1}$ is a zero-mean noise vector with complex Gaussian distribution and identity covariance matrix, and $\rho_u$ is the uplink transmit power. The transmitted symbol from the $k$-th user, $x_k^u$, is the $k$-th element of $\mathbf{x}_u = [x_1^u, \ldots, x_K^u]^{\mathrm{T}}$ with $E[|x_k^u|^2] = 1$.

Based on the assumption that the small-scale fading coefficients for different users are independent, the column channel vectors from different users are asymptotically orthogonal as the number of antennas at the BS, $N$, grows to infinity [2]. Then, we have

$$\mathbf{H}^{\mathrm{H}}\mathbf{H} = \mathbf{D}^{1/2}\mathbf{G}^{\mathrm{H}}\mathbf{G}\mathbf{D}^{1/2} \approx N\mathbf{D}^{1/2}\mathbf{I}_K\mathbf{D}^{1/2} = N\mathbf{D}. \tag{12}$$

A detailed discussion of this result can be found in [9]. Based on the result in (12), the overall achievable rate of all users becomes

$$\begin{aligned} C &= \log_2 \det(\mathbf{I} + \rho_u\mathbf{H}^{\mathrm{H}}\mathbf{H}) \\ &\approx \log_2 \det(\mathbf{I} + N\rho_u\mathbf{D}) \\ &= \sum_{k=1}^{K} \log_2(1 + N\rho_u d_k) \quad \frac{\text{bits}}{\text{s}}{\text{Hz}}. \end{aligned} \tag{13}$$

In the following, we show that simple MF processing at the BS can achieve the capacity in (13). When MF processing is used, the BS processes the signal vector by multiplying the conjugate-transpose of the channel, as

$$\begin{aligned} \mathbf{H}^{\mathrm{H}}\mathbf{y}_u &= \mathbf{H}^{\mathrm{H}}(\sqrt{\rho_u}\mathbf{H}\mathbf{x}_u + \mathbf{n}_u) \\ &\approx N\sqrt{\rho_u}\mathbf{D}\mathbf{x}_u + \mathbf{H}^{\mathrm{H}}\mathbf{n}_u, \end{aligned} \tag{14}$$

where (12) is used. Note that the channel vectors are asymptotically orthogonal when the number of antennas at the BS grows to infinity. Therefore, $\mathbf{H}^{\mathrm{H}}$ does not color the noise. Since $\mathbf{D}$ is a diagonal matrix, the MF processing separates the signals from different users into different streams and there is asymptotically no inter-user interference. The signal transmissions from each user can thus be treated as if originating from a SISO channel. From (14), the *signal-to-noise ratio* (SNR) for the $k$-th user is $N\rho_u d_k$. Consequently, the rate achievable by using MF is the same as the limit in (13), which implies that simple MF processing at the BS is optimal when the number of antennas at the BS, $N$, grows to infinity.

*2) Downlink:* Denote $\mathbf{y}_d \in \mathcal{C}^{K \times 1}$ as the received signal vector at all $K$ users. For most prior work in massive MIMO, *time-division duplexing* (TDD) mode is assumed as discussed in Section III-A, where the downlink channel is the transpose of the uplink channel matrix. Then, the received signal vector can be expressed as

$$\mathbf{y}_d = \sqrt{\rho_d}\mathbf{H}^{\mathrm{T}}\mathbf{x}_d + \mathbf{n}_d, \tag{15}$$

where $\mathbf{x}_d \in \mathcal{C}^{N \times 1}$ is the signal vector transmitted by the BS, $\mathbf{n}_d \in \mathcal{C}^{K \times 1}$ is additive noise defined as before, and $\rho_d$ is the transmit power of the downlink. To normalize the transmit power, we assume $E\{\|\mathbf{x}_d\|^2\} = 1$.

As we see in Section III, the BS usually has CSI corresponding to all users based on uplink pilot transmission. Therefore, it is possible for the BS to perform power allocation to maximize the sum transmission rate. With power allocation, the sum capacity for the system is [10]

$$C = \max_{\mathbf{P}} \log_2 \det(\mathbf{I}_N + \rho_d \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{H}})$$
$$\approx \max_{\mathbf{P}} \log_2 \det(\mathbf{I}_K + \rho_d N \mathbf{P}\mathbf{D}) \quad \frac{\text{bits}}{\frac{\text{s}}{\text{Hz}}}, \tag{16}$$

where (12) is used and $\mathbf{P}$ is a positive diagonal matrix with the power allocations $(p_1, \ldots, p_K)$ as its diagonal elements and $\sum_{k=1}^{K} p_k = 1$.

If the MF precoder is used, the transmitted signal vector is

$$\mathbf{x}_d = \mathbf{H}^*\mathbf{D}^{-1/2}\mathbf{P}^{1/2}\mathbf{s}_d, \tag{17}$$

where $\mathbf{s}_d \in \mathcal{C}^{K \times 1}$ is the source information vector. Then, the received signal vector at all $K$ users is

$$\mathbf{y}_d = \sqrt{\rho_d}\mathbf{H}^{\mathrm{T}}\mathbf{H}^*\mathbf{D}^{-1/2}\mathbf{P}^{1/2}\mathbf{s}_d + \mathbf{n}_d$$
$$\approx \sqrt{\rho_d}N\mathbf{D}^{1/2}\mathbf{P}^{1/2}\mathbf{s}_d + \mathbf{n}_d, \tag{18}$$

where the second line of (18) is for the case when the number of antennas at the BS, $N$, grows to infinity, and (12) is used. Since $\mathbf{P}$ and $\mathbf{D}$ are both diagonal matrices, the signal transmission from the BS to each user can be treated as if originating from a SISO transmission, and again, inter-user interference has been suppressed. The overall achievable data rate in (18) can be maximized by appropriate choice of the power allocation as in (16), which demonstrates that the capacity can be achieved using the simple MF precoder.

Above, we have shown that based on the favorable propagation assumption of (12), the simple MF precoder/detector can achieve the capacity of a massive MU-MIMO system when the number of antennas at the BS, $N$, is much larger than the number of users, $K$, and grows to infinity, i.e., $N \gg K$ and $N \to \infty$. The results for another scenario that both the number of antennas at the BS and the number of users grow large while their ratio is bounded, i.e., $N/K = c$ as $N, K \to \infty$, where $c$ is a constant, are different. The detailed results for both scenarios are discussed in the following sections.

## III. CHANNEL ESTIMATION AND SIGNAL DETECTION

In this section, channel estimation and signal detection at the BS are discussed. We first discuss channel estimation methods for massive MIMO and explain why TDD mode is usually



Fig. 2. Multi-User MIMO TDD protocol.

assumed. Then, both linear and non-linear signal detection methods are presented.

### A. Channel Estimation

For regular MIMO systems, multi-user precoding in the downlink and detection in the uplink require CSI at the BS. The resource, time or frequency, required for channel estimation in a MIMO system is proportional to the number of the transmit antennas and is independent of the number of the receive antennas.

If FDD is used, that is, uplink and downlink use different frequency bands, the CSI corresponding to the uplink and downlink is different. Channel estimation for the uplink is done at the BS by letting all users send different pilot sequences. The time required for uplink pilot transmission is independent of the number of antennas at the BS. However, to get CSI for the downlink channel in FDD systems, a two-stage procedure is required. The BS first transmits pilot symbols to all users, and then all users feed back estimated CSI (partial or complete) for the downlink channels to the BS. The time required to transmit the downlink pilot symbols is proportional to the number of antennas at the BS. As the number of BS antennas grows large, the traditional downlink channel estimation strategy for FDD systems becomes infeasible. For example, consider a 1 ms × 100 kHz channel coherence interval, which can support transmission of 100 complex symbols. When there are 100 antennas at the BS, the whole coherence interval will be used for downlink training if orthogonal pilot waveforms are used for channels to each antenna, while there is no symbol left for data transmission.

Fortunately, the channel estimation strategy in TDD systems can be utilized to solve the problem. Based on the assumption of channel reciprocity, only CSI for the uplink needs to be estimated. In [12], a TDD protocol, shown in Fig. 2, was proposed. According to this protocol, all the users in all the cells first synchronously send uplink data signals. Next, the users send pilot sequences. BSs use these pilot sequences to estimate CSI to the users located in their cells. Then, BSs use the estimated CSI to detect the uplink data and to generate beamforming vectors for downlink data transmission. However, due to the limited channel coherence time, the pilot sequences employed by users in neighboring cells may no longer be orthogonal to those within the cell, leading to a pilot contamination problem [2], which will be discussed in detail in Section V.

Linear MMSE based channel estimation is commonly used, which can provide near-optimal performance with low complexity. Besides MMSE estimation, a compressive sensing-based channel estimation approach is proposed in [13], which exploits the fact that the degrees of freedom of the physical channel matrix are much smaller than the number of free parameters. To improve the spectral efficiency of the system, a time-frequency training sequence design is developed in [14].

The proposed structure achieves the benefits of both time- and frequency-domain estimation while avoiding their individual drawbacks.

### B. Signal Detection

Linear signal detectors with low complexity, such as the MF, ZF, and MMSE detectors, are practical candidates for massive MIMO systems. They can asymptotically achieve capacity as the number of antennas at the BS is large enough compared to the number of users and the channel vectors from different users are independent [2], [7]. The performance of massive MIMO systems based on various linear receivers has been studied from various perspectives [15]–[18]. A performance comparison between the MMSE and the MF receivers in realistic system settings is provided in [15]. It shows that the MMSE receiver can achieve the same performance as the MF receiver with fewer antennas, especially when there exists inter-cell interference. The scenario with a bounded ratio of the number of antennas to the number of users has been investigated in [16] and [17] for the MMSE and ZF receivers, respectively. In [16], an expression for the asymptotic *signal-to-interference-plus-noise-ratio* (SINR) of the MMSE receiver for a single-cell system with a bounded ratio of the number of antennas to the number of users is obtained. Two types of MMSE receivers are considered: the optimal MMSE receiver taking different transmit power levels from different users into account and a suboptimal MMSE receiver assuming equal transmit power. In [17], the exact data rate, symbol-error rate, and outage performance of the ZF receivers are derived. Besides centralized MIMO systems, the sum rate of the ZF receivers in distributed MIMO systems is also analyzed and lower and upper bounds on the sum rate are derived in [18]. A rough calculation of the complexity order for the ZF and MMSE receivers is $\mathcal{O}(NK + NK^2 + K^3)$ [7].

In addition to linear detection methods, non-linear detection can also be used to achieve better performance at the cost of higher computational complexity. Complexity reduction for non-linear detectors in massive MIMO systems is the key issue, and some work has been done on this topic. In [19], a *block-iterative generalized decision feedback equalizer* (BI-GDFE) is proposed, and its asymptotic SINR performance is evaluated. The complexity order of the proposed BI-GDFE is $\mathcal{O}((NK + N^2K + N^3)N_{iter})$, where $N_{iter}$ is the number of iterations. For random MIMO channels, the proposed BI-GDFE can approach the single-user MF bound within only a few iterations even if the number of antennas is large. Complexity reduction schemes for the existing local neighborhood search, including *likelihood ascent search* (LAS) [20], [21] and *tabu search* (TS) [22], are presented. LAS-based detection in [21] can achieve a better bit-error rate with the same order of complexity as traditional LAS. The layered TS method in [22] performs detection in a layered manner and works well in large MIMO systems with low complexity, which has the complexity order as $\mathcal{O}(((N+N_{tabu})N_{neighbor}+NK)N_{iter}+NK^2+K^3)$, where $N_{tabu}$ and $N_{neighbor}$ are the maximum number of entries and the number of neighboring vectors used by the algorithm. Low-complexity graph-based schemes are proposed in [23] and [24]. In [23], a low-complexity receiver based on cooperative

particle swarm optimization and factor-graph data detection is investigated. To obtain good features from both the local neighborhood search algorithm and the factor-graph based *belief propagation* (BP) algorithm, a hybrid reactive TS-BP approach is developed in [24]. It can achieve near-optimal performance with low complexity for signals with high-order modulation. Moreover, the element-based *lattice-reduction* (LR) algorithms in [25] can provide better performance than other LR approaches with lower complexity. Some other detection related work can be found in [26]–[28].

## IV. PRECODING

In this section, precoding at the BS is discussed. We first discuss basic precoding methods and then extend the discussion to multi-cell precoding. Finally, some practical issues related to precoding are discussed.

For regular MIMO systems, both non-linear and linear precoding techniques can be used. Compared with linear precoding methods, non-linear methods, such as *dirty-paper-coding* (DPC) [29], *vector perturbation* (VP) [30] and lattice-aided methods [31], have better performance albeit with higher implementation complexity. However, with an increase in the number of antennas at the BS, linear precoders, such as MF and ZF, are shown to be near-optimal [2], [7]. Thus, it is more practical to use low-complexity linear precoding techniques in massive MIMO systems. Therefore, we mainly focus on linear precoding techniques in this section.

### A. Basic Precoding

Basic precoding methods include MF and ZF. When MF is used, the transmit signal from the BS can be expressed as

$$\mathbf{x}_d^{MF} = \frac{1}{\sqrt{\alpha}}(\mathbf{H}^{\mathrm{T}})^{\mathrm{H}}\mathbf{s}_d = \frac{1}{\sqrt{\alpha}}\mathbf{H}^*\mathbf{s}_d, \qquad (19)$$

where $\alpha$ is a power normalization factor. The impact of normalization techniques is discussed in [32], which shows that vector normalization is better for ZF while matrix normalization is better for MF.

When ZF is used, the transmit signal from the BS can be expressed as

$$\mathbf{x}_d^{ZF} = \frac{1}{\sqrt{\alpha}}(\mathbf{H}^{\mathrm{T}})^{\dagger}\mathbf{s}_d = \frac{1}{\sqrt{\alpha}}\mathbf{H}^*(\mathbf{H}^{\mathrm{T}}\mathbf{H}^*)^{-1}\mathbf{s}_d. \qquad (20)$$

For *regularized ZF* (RZF), a diagonal loading factor is added prior to the inversion of the matrix $\mathbf{H}^{\mathrm{T}}\mathbf{H}^*$, and the transmit signal at the BS is expressed as

$$\mathbf{x}_d^{RZF} = \frac{1}{\sqrt{\alpha}}\mathbf{H}^*(\mathbf{H}^{\mathrm{T}}\mathbf{H}^* + \delta\mathbf{I})^{-1}\mathbf{s}_d, \qquad (21)$$

where $\delta > 0$ is the regularization factor, and can be optimized based on the design requirements. The RZF precoder becomes the ZF precoder as $\delta \to 0$, and becomes the MF precoder as $\delta \to \infty$.

The performance of (R)ZF precoding for a single-cell massive MIMO system is analyzed in [33] when the number of antennas at the BS, $N$, is much larger than the number of users, $K$, and grows to infinity, i.e., $N \gg K$ and $N \to \infty$. The analysis is based on estimated CSI, where the estimated channel matrix,

$\widehat{\mathbf{H}}$, is used instead of the accurate channel matrix $\mathbf{H}$, in (20) and (21). A lower bound on the sum rate for the ZF precoder is derived. The ZF precoder outperforms MF in the high spectral efficiency region while MF is better in the low spectral efficiency region. The computational complexity of these precoders is discussed in [33] as well. To reach maximum spectral efficiency, the ZF precoder requires less computation than the MF precoder, a surprising result based on the fact that fewer users are served for the ZF precoder at peak spectral efficiency. Note that the user selection problem may become more complex as the number of users grows.

Besides the scenario where the number of antennas at the BS is much larger than the number of users, another scenario is investigated in [34]–[36] where both the number of antennas at the BS and the number of users grow large while their ratio is bounded, i.e., $N/K = c$ as $N, K \rightarrow \infty$, where $c$ is a constant. In this setup, ZF precoding achieves an SNR that tends to the optimal SNR for an interference-free system with $N - K$ transmit antennas when $N, K \rightarrow \infty$ [7]. With perfect CSI, the asymptotic SINR performance of RZF precoding is derived in [34], which depends on the user-to-antenna ratio, $K/N$, the regularization parameter, $\delta$, and the SNR. In [35], the analysis is extended to take the transmit correlation into account. With an estimated channel matrix, $\widehat{\mathbf{H}}$, the equivalent deterministic SINRs for ZF and RZF precoding are derived in [36], where a tight approximation is derived even for systems with a finite number of BS antennas.

### B. Multi-Cell Precoding

Linear precoders can also be used in multi-cell massive MIMO systems, where cooperating BSs are designed to jointly serve users in different cells [37]. Depending on the overhead of the information exchange among the BSs, there are three scenarios: single-cell processing, coordinated beamforming, and network MIMO multi-cell processing. Single-cell processing is based on the assumption that BSs only have channel information for users in their own cells and no information about users in other cells. Coordinated beamforming exploits channel information from a BS to users of all cells. The network MIMO multi-cell processing concept is based on full cooperation among BSs, where not only channel information but also data are globally shared. Among the above three scenarios, single-cell processing can avoid the information exchange overhead, but it cannot mitigate inter-cell interference. Based on the single-cell processing assumption, the RZF precoder can achieve performance similar to MF with fewer antennas [15]. Processing based on network MIMO provides the best performance but has the highest information exchange overhead [38]–[40]. The network-MIMO-based scheme proposed in [40] can achieve the performance limit with one order of magnitude fewer BS antennas than single-cell processing [2]. As discussed below, coordinated beamforming can obtain a tradeoff between performance and the overhead of information exchange [41].

Compared with regular MIMO systems, the use of coordinated beamforming with massive MIMO is considerably more difficult; with a large number of BS antennas, it becomes more and more impractical to share instantaneous CSI among the BSs.

Therefore, different schemes based on sharing statistical CSI are investigated. Scenarios where the number of antennas at the BS is much larger than the number of users, $N \gg K$, where their ratio is bounded, $N/K = c$ as $N, K \rightarrow \infty$, are both studied. For the first scenario, a beamformer is designed to minimize total transmit power across all BS's in [42]. Random matrix theory is utilized to obtain an asymptotically optimal distributed beamformer. A two-layer precoding approach is proposed in [43] to relax the requirement that one BS should get full CSI for its own cell, where the BS only needs to estimate the channels within the subspace determined by the outer precoder. Thus, the proposed scheme also reduces the number of pilot symbols required for channel estimation in addition to the overhead due to information exchange. For the second scenario, a min-max-fair coordination beamformer is proposed in [44], [45], which exploits a nested ZF structure to recursively solve a series of min-max fairness problems of decreasing dimension. To further reduce the information exchange overhead for power allocation updating, the scheme proposed in [46] maximizes the minimum weighted SINR of the users.

Distributed MIMO based on limited CSI exchange is investigated in [47], where the proposed linear Hermitian precoding in each cell uses only CSI for its own cell and the large-scale fading coefficients of users in other cells. The performance loss due to the use of limited CSI is negligible compared to the case with full CSI exchange. Thus, the design can be used to reduce the information exchange overhead, especially for massive MIMO systems.

Note that for the coordinated schemes in [42]–[47], how to obtain the required CSI is not taken into account. The analysis is based on the assumption that the required CSI, either instantaneous or statistical, is already known at the BSs. In [48], [49], and [50], the problem of multi-cell precoding has been considered in more general and realistic settings. In particular, channel estimation error and pilot contamination are explicitly taken into account. These results will be discussed in Section V.

### C. Practical Considerations

The precoding designs in [51]–[53] take practical antenna array structures into account. To build a large array in practice, the use of non-linear but power-efficient *radio-frequency* (RF) front-end amplifiers is preferred. To avoid signal distortion, the transmit signals are required to have a low *peak-to-average-power-ratio* (PAPR). A precoding method based on per-antenna constant envelope constraints is developed to reduce the PAPR of the transmit signal in [51], [52]. Compared to the average-only transmit power constraint, extra transmit power is required to achieve a desired rate with per-antenna power constraints [54], [55]. However, precoding with such constraints facilitates the use of power-efficient amplifiers, leading to improvement in the power efficiency of the entire system. Instead of considering precoding only, a framework based on joint precoding, modulation, and PAPR reduction is proposed in [56]. The approach described in [56] has low PAPR and enables the use of low-cost RF amplifiers. Moreover, a comprehensive performance analysis regarding the use of large and dense antenna arrays is provided in [53].

Since the computational complexity of the precoder increases with the number of antennas at the BS, low-complexity precoding methods are necessary. A low-complexity norm descent search precoder is proposed in [57] for systems with a large number of antennas, whose bit-error rate curves drop exponentially with SNR. A low-complexity method based on matrix inversion is proposed in [58] to decrease the computational complexity of the ZF and the MMSE precoders. In general, the design of low-complexity precoders is critical for massive MIMO systems.

## V. PILOT CONTAMINATION

As we have discussed in Section III-A, only TDD is likely to be employed with massive MIMO systems due to the high overhead and complexity associated with channel estimation and channel sharing under FDD. As shown in Fig. 2, for TDD-based massive MIMO transmission systems, pilot sequences are transmitted from users in the uplink to estimate channels. Let $\boldsymbol{\psi}_{k,l} = (\psi_{k,l}^{[1]}, \ldots, \psi_{k,l}^{[\tau]})^{\mathrm{T}}$ be the pilot sequence of user $k$ in cell $l$, where $\tau$ denotes the length of the pilot sequence. Though it is not necessary, it is convenient to assume that $|\psi_{k,l}^{[j]}| = 1$ and we use this assumption in what follows. Ideally, the pilot sequences employed by users within the same cell and in the neighboring cells should be orthogonal, that is

$$\boldsymbol{\psi}_{k,l}^{\mathrm{H}} \boldsymbol{\psi}_{j,l'} = \delta[k-j]\delta[l-l'], \tag{22}$$

where $\delta[.]$ is defined as

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

In this case, a BS can obtain uncontaminated estimation of the channel vectors in the sense that they are not correlated to the channel vectors of other users.

However, the number of orthogonal pilot sequences with a given period and bandwidth is limited, which in turn limits the number of users that can be served [2]. In order to handle more users, nonorthogonal pilot sequences are used in neighboring cells. Thus for some different $k, j, l,$ and $l'$, we may have

$$\boldsymbol{\psi}_{k,l}^{\mathrm{H}} \boldsymbol{\psi}_{j,l'} \neq 0. \tag{24}$$

As a result, the estimate of the channel vector to a user becomes correlated with the channel vectors of the users with non-orthogonal pilot sequences. Below, we consider the pilot contamination issue in more details and discuss possible methods for its mitigation.

### A. Pilot Contamination Effect

There are many schemes for assigning pilot sequences to users in different cells. One simple scheme is to reuse the same set of orthogonal pilot sequences, say $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K$, in all cells. This means that the $k$-th user in any cell will be assigned the pilot sequence $\boldsymbol{\psi}_k$. Identical pilot sequences assigned to users in neighboring cells will interfere with each other causing pilot contamination. This situation is shown in Fig. 3. Only users with
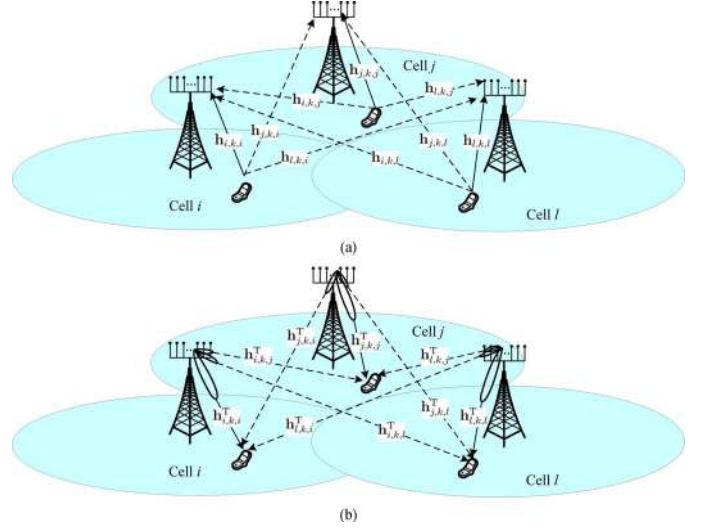


Fig. 3. Illustration of pilot contamination concept. (a) Uplink Transmission. (b) Downlink Transmission.

the same pilot sequence $\boldsymbol{\psi}_k$ are shown. During downlink transmission, this results in the BS beamforming signals not only to its own users, but also to users in the neighboring cells, and therefore creates a strong source of directional interference. This interference, unlike the intra-cell interference, will not disappear as the number of BS antennas increases. A similar effect occurs during the uplink transmission.

Consider a system with $L$ cells. Each cell is assumed to have $K$ single-antenna users and a BS with $N$ antennas, where $N \gg K$. For purposes of illustration, we assume that all $L$ cells use the same set of $K$ pilot sequences, represented by the $\tau \times K$ orthogonal matrix $\boldsymbol{\Phi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K)$ satisfying $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} = \tau\mathbf{I}$. Assume further that the pilot transmission from different cells is synchronized, as in Fig. 3(a). The received signal matrix at the $i$-th BS, $\mathbf{Y}_i^p \in \mathcal{C}^{N \times \tau}$, can be expressed as

$$\mathbf{Y}_i^p = \sqrt{\rho_p} \sum_{l=1}^{L} \mathbf{H}_{i,l} \boldsymbol{\Phi}^{\mathrm{T}} + \mathbf{N}_i^p, \tag{25}$$

where $\mathbf{H}_{i,l} \in \mathcal{C}^{N \times K}$, defined in (8), is the channel matrix from all $K$ users in the $l$-th cell to the $i$-th BS. The $k$-th column of $\mathbf{H}_{i,l}$, denoted by $\mathbf{h}_{i,k,l}$, is the channel vector from the $k$-th user in the $l$-th cell to the $i$-th BS. $\mathbf{N}_i^p \in \mathcal{C}^{N \times \tau}$ is the noise matrix at the $i$-th BS during the pilot transmission phase, whose entries are i.i.d. circular complex Gaussian random variables with zero-mean and unit variance, and $\rho_p$ is the pilot transmit power.

To estimate the channel, the $i$-th BS projects its received signal on $\boldsymbol{\Phi}^*$ to get sufficient statistics for estimating $\mathbf{H}_{i,i}$ [59]. From (25), the resulting estimated channel matrix is

$$\begin{aligned} \widehat{\mathbf{H}}_{i,i} &= \frac{1}{\sqrt{\rho_p}\tau} \mathbf{Y}_i^p \boldsymbol{\Phi}^* \\ &= \mathbf{H}_{i,i} + \sum_{l \neq i} \mathbf{H}_{i,l} + \frac{1}{\sqrt{\rho_p}\tau} \mathbf{N}_i^p \boldsymbol{\Phi}^*, \end{aligned} \tag{26}$$

where we have used $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} = \tau\mathbf{I}$. The $k$-th column, $\widehat{\mathbf{h}}_{i,k,i}$, of $\widehat{\mathbf{H}}_{i,i}$ is the estimate of the channel vector, $\mathbf{h}_{i,k,i}$. From (26), we have

that the estimate, $\widehat{\mathbf{h}}_{i,k,i}$, is a linear combination of the channel vectors, $\mathbf{h}_{i,k,l}, l = 1, \ldots, L$, of the users in different cells with the same pilot sequence. In [2], this phenomenon is referred to as *pilot contamination*.

In a MIMO system with no pilot contamination, the channel estimation and precoding can be completely decoupled from each other. In the case of pilot contamination this is not the right approach. Because of the pilot contamination, the estimated channel vector in any cell is a linear combination of channel vectors of users in other cells that use the same pilot sequence. In [60], an MMSE-based precoding is proposed that takes this particular form of the estimated channel vector into account and attempts to minimize i) the sum of squares of errors for users located in cell $l$; and ii) the sum of squares of interferences for users located in other cells. MMSE-based precoding provides significant improvement compared with traditional precoding methods like ZF, single-cell MMSE precoding from [61], and others.

Consider now the uplink data transmission stage after channel estimation. The received signal at the $i$-th BS is

$$\mathbf{y}_i^u = \sqrt{\rho_u} \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbf{h}_{i,k,l} x_{k,l}^u + \mathbf{n}_i^u, \tag{27}$$

where $x_{k,l}^u$ is the symbol from the $k$-th user in the $l$-th cell and $\mathbf{n}_i^u$ is the additive noise vector during uplink transmission. When the MF detector is used, the BS processes the signal vector by multiplying with the conjugate-transpose of the channel estimate. Thus the detected symbol from the $k$-th user of the $i$-th cell, $\widehat{x}_{k,i}^u$, is

$$\begin{aligned}
\widehat{x}_{k,i}^u &= \left( \widehat{\mathbf{h}}_{i,k,i} \right)^{\mathrm{H}} \mathbf{y}_i^u \\
&= \left( \sum_{l=1}^{L} \mathbf{h}_{i,k,l} + \widehat{\mathbf{v}}_i \right)^{\mathrm{H}} \left( \sqrt{\rho_u} \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbf{h}_{i,k,l} x_{k,l}^u + \mathbf{n}_i^u \right),
\end{aligned} \tag{28}$$

where $\widehat{\mathbf{v}}_i$ is the $i$-th column of $(1/\sqrt{\rho_p}\tau)\mathbf{N}_i^p\boldsymbol{\Phi}^*$ in (26).

From (28) and (12), as the number of BS antennas grows large, i.e., $N \to \infty$, the SINR of the $k$-th user in the $i$-th cell tends to the following limit [2]

$$\text{SINR}_{k,i}^u = \frac{d_{i,k,i}^2}{\sum_{l \neq i} d_{i,k,l}^2}, \tag{29}$$

where $d_{i,k,l}$ is the large-scale channel fading coefficient in (7). From (29), the SINR depends only on the large-scale fading factors of the channels while the small-scale fading factors and noise are averaged out. Moreover, from (26), when non-orthogonal pilot sequences are used in different cells, the BS cannot distinguish among the channel vectors from its own cell to channels from other cells. Furthermore, the SINR limit due to pilot contamination will not disappear as the number of antennas increases. Similar results hold for the ZF or the MMSE detectors.

The pilot contamination affects the downlink transmission as well, as shown in Fig. 3(b). For the downlink, the power varies from one coherent interval to another if (26) is used directly as beamforming vectors. Thus, a normalized version for

the beamforming vectors is commonly proposed in [62]. Let the MF beamforming vector from the $i$-th BS to the $k$-th user in the $i$-th cell be

$$\mathbf{w}_{k,i}^d = \frac{\widehat{\mathbf{h}}_{i,k,i}}{\|\widehat{\mathbf{h}}_{i,k,i}\|} = \frac{\widehat{\mathbf{h}}_{i,k,i}}{\alpha_{k,i}\sqrt{N}}, \tag{30}$$

where the scalar $\alpha_{k,i} = \|\widehat{\mathbf{h}}_{i,k,i}\|/\sqrt{N}$ is a normalization factor. Then the $i$-th BS transmits a $N$-dimensional vector

$$\mathbf{x}_i^d = \sum_{k=1}^{K} \mathbf{w}_{k,i}^d s_{k,i}^d, \tag{31}$$

where $s_{k,i}^d$ is the source symbol for the $k$-th user in the $i$-th cell. The received signal at the $k$-th user of the $i$-th cell is

$$y_{k,i}^d = \sqrt{\rho_d} \sum_{l=1}^{L} \mathbf{h}_{l,k,i}^{\mathrm{T}} \sum_{k'=1}^{K} (\mathbf{w}_{k',l}^d)^* s_{k',l}^d + n_{k,i}^d, \tag{32}$$

where $n_{k,i}^d$ is the additive noise. Based on a derivation similar to that for the uplink, the downlink SINR of the $k$-th user in the $i$-th cell, as $N \to \infty$, tends to [62]

$$\text{SINR}_{k,i}^d = \frac{\frac{d_{i,k,i}^2}{\alpha_{k,i}^2}}{\sum_{l \neq i} \frac{d_{l,k,i}^2}{\alpha_{k,l}^2}}, \tag{33}$$

where

$$\alpha_{k,l}^2 = \sum_{j=1}^{L} d_{l,k,j} + \frac{1}{\rho_p\tau}. \tag{34}$$

The expression in (33) is slightly different from that in [2], [7] due to the normalization for the MF precoder.

Besides the normalization factor, the statistical properties of the interference terms in (29) and (33) are different. In the uplink, interference comes from the users with the same pilot sequence as the $k$-th user in the $j$-th cell. The denominator in (29) depends on the distances from these interfering users to the $j$-th BS. In the downlink, interference comes from the neighboring BSs that transmit using contaminated channel estimates. The denominator in (33) depends on the distances from the neighboring BSs to the $k$-th user in the $j$-th cell. Moreover, the coefficients in the denominator of (29) can be treated as independent parameters while the coefficients in the denominator of (33) are all correlated. To see this, it is enough to note that distances from the $k$-th user in the $i$-th cell to the neighboring cells, say the $l$-th, the $j$-th and the $m$-th cells, almost precisely determine the coordinates of the user, and therefore also determine the distances to other cells. Even though the statistical properties are different for the uplink and downlink, the dissimilarities have little impact on performance [2], [62].

Since it is an important issue that limits the performance of multi-cell massive MIMO systems, the pilot contamination effect under different scenarios is investigated in several papers [15], [63]–[65]. To demonstrate how the SINR changes with the number of antennas at the BS, the SINR convergence rate is derived in [63]. From (29) and (33), a massive MIMO system with strong pilot contamination obviously has a lower

performance limit than one with weak pilot contamination. However, it reaches the (lower) performance limit with fewer antennas at the BS. In [64], the sum rate lower bound is derived when pilot contamination is present. The impact of more practical channel models on pilot contamination is analyzed in [65]. The model takes the correlation of channels corresponding to different users into account and is also suitable for environments without rich scattering. Using general channel models and practical precoders/detectors and taking the pilot contamination effect into account, it is demonstrated in [15] that the RZF/MMSE precoder/detector can perform as well as the MF precoder/detector with almost one order of magnitude fewer antennas when the number of antennas at the BS is not extremely large compared to the number of users in each cell.

The impact of pilot contamination when the number of antennas at the BS and the number of users grow to infinity while maintaining a fixed ratio is studied in [66], [67]. In particular, performance limits based on MF and MMSE filters are derived in [66], [67], and the asymptotic SINR in both cases depends on both pilot contamination and interference averaging. The difference is that the MMSE filter can obtain an interference suppression gain, leading to higher asymptotic SINR performance. When the ratio of the number of users and the number of antennas at the BS tends to zero, similar SINR results as in (29) and (33) can be obtained where only the pilot contamination effect appears in the asymptotic SINR expression.

## B. Mitigating Pilot Contamination

Several methods have been developed to mitigate the pilot contamination effect. We introduce some of them here.

*1) Protocol-Based Methods:* One way to reduce the effect is through frequency reuse or reducing the number of served users that use non-orthogonal pilot sequences [2], [68]. For some specific cases, the performance can be improved [68], but in general, frequency reuse may make little difference since fewer users are served simultaneously even though the SINRs for specific users increase [2].

In the transmission protocol proposed in [2], all users transmit pilot sequences synchronously, as shown in Fig. 2. In order to mitigate the pilot contamination, a scheme based on a time-shifted (asynchronous) protocol is proposed in [69], [70] and [62]. The basic idea is to partition cells into several groups $A_1, \ldots, A_\Gamma$ and to use a time-shifted protocol in each group. An example of this approach with $\Gamma = 3$ is shown in Fig. 4. While the users from group $A_1$ transmit pilots, the BSs from $A_2$ transmit downlink data signals. This avoids pilot contamination among users from $A_1$ and $A_2$. At the same time the BSs from $A_1$ have to estimate their channel vectors in the presence of downlink signals transmitted by the BSs from $A_2$ and $A_3$. Since downlink transmit power, $\rho_d$, is usually significantly higher than that of pilot sequences, $\rho_p$, it is not a priori clear that this approach provides any gain. From [70] and [62], as $N \to \infty$, the uplink and downlink SINRs tend to the following limits,

$$\text{SINR}_{k,i}^u = \frac{d_{i,k,i}^2}{\sum\limits_{j \in A_\gamma, j \neq i} d_{i,k,j}^2}, \tag{35}$$
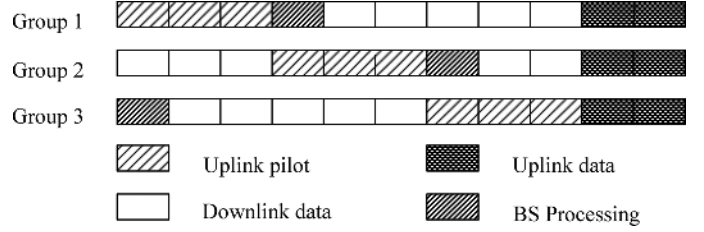


Fig. 4.   Time-shifted pilot scheme with $\Gamma = 3$.

and,

$$\text{SINR}_{k,i}^d = \frac{\frac{d_{i,k,i}^2}{\alpha_{k,i}^2}}{\sum\limits_{j \in A_\gamma, j \neq i} \frac{d_{l,k,i}^2}{\alpha_{k,l}^2}}. \tag{36}$$

From these expressions, only users from the same group create interference to each other. This leads to better performance compared to (29) and (33). This scheme yields the same SINRs as the scheme with frequency reuse $\Gamma$ while still allowing use of the entire band in all cells. Instead of using the same uplink and downlink transmit powers, $\rho_u$ and $\rho_d$, optimized individual powers, $\rho_{k,l}^u$ and $\rho_{k,l}^d$, can be used for each user. It is shown in [62] that power allocation can substantially improve the SINR compared with (29) and (33).

*2) Precoding Methods:* A distributed single-cell precoding method is proposed in [60]. According to this method, the precoding matrix at one BS is designed to minimize the sum of the squared error of its own users and interference to the users in all other cells. The distributed single-cell precoding method is shown to provide better performance than traditional single-cell ZF precoding.

The precoding methods based on multi-cell cooperation in [38] and [39] can mitigate the pilot contamination effect. However, the information exchange overhead required among the BSs increases with the number of antennas. Therefore, these methods are only feasible for MIMO systems with a limited number of antennas.

To obtain the benefit of cooperation while limiting the information exchange overhead, a *pilot contamination precoding* (PCP) method is proposed in [48]. PCP is based on two assumptions that the source signals for all users in all cells are accessible at each BS and that large-scale fading coefficients, $d_{i,k,l}$, are accessible to all BS's or a network hub. Due to pilot contamination, the channel estimate, $\widehat{\mathbf{h}}_{i,k,i}$, is a linear combination of the channel vectors $\mathbf{h}_{i,k,l}, l = 1, \ldots, L$ from (26). Instead of mitigating interference caused by this contaminated estimate, each BS uses the pilot contamination for transmitting information to all users in the network as described below.

In order to transmit information to the $k$-th users of *all cells*, the $i$-th BS transmits signal

$$\widehat{s}_{k,i}^d = \sum_{l=1}^L a_{i,k,l} s_{k,l}^d, \tag{37}$$

instead of the source symbol $s_{k,i}^d$. The coefficients $a_{i,k,l}$ are called *PCP coefficients*. They are computed as a function of the large-scale fading coefficients $d_{j,k,l}$. The choice $a_{i,k,l} = \delta[i-l]$

corresponds to the case of no PCP. Note that $N$-dimensional MF beamforming vectors, $\mathbf{w}_{k,i}^d$, are computed exactly as in (30), that is completely locally without any cooperation between BSs. Uplink PCP is formulated similarly.

If the PCP coefficients $a_{i,k,l}$ are computed using the ZF criterion proposed in [48], it is referred as *ZF-PCP*. ZF-PCP allows one to simultaneously remove the additive noise and inter-cell interference. Thus, ZF-PCP theoretically yields infinite SINRs, $\mathrm{SINR}_{k,i}^d = \infty$, as $N \to \infty$.

It is interesting to note, however, that ZF PCP does not give good results for finite values of $N$. From [48] that for finite $N$, we have

$$\mathrm{SINR}_{k,l}^d = \frac{N\rho_d\rho_u\tau \left| \sum_{j=1}^{L} d_{j,k,l}a_{j,k,l} \right|^2}{T_1 + N \cdot T_2}, \quad (38)$$

where

$$T_1 = \sum_{j=1}^{L} \sum_{n=1}^{K} \left( \frac{1}{L} + \rho_d d_{j,k,l} \right) \left( 1 + \rho_d \tau \sum_{s=1}^{L} d_{j,n,s} \right) \sum_{u=1}^{L} |a_{j,n,u}|^2 \quad (39)$$

and

$$T_2 = \rho_d \rho_u \tau \sum_{u \neq l} \left| \sum_{j=1}^{L} d_{j,k,l}a_{j,k,u} \right|^2. \quad (40)$$

When the coefficients $a_{i,k,l}$ are numerically optimized to directly maximize $\mathrm{SINR}_{k,l}^d$ in (38), the approach is referred as optimal PCP.

The expression in (38) allows one to conduct Monte-Carlo simulation of massive MIMO systems in a very simple fashion, by simply generating random large-scale fading coefficients $d_{i,k,l}$. Fig. 5 shows the *cumulative distribution function* (CDFs) of user rates for different PCPs. The user downlink rates are computed as $\log_2(1 + \mathrm{SINR}_{k,l}^d)$. The network consists of $L = 7$ wrapped around cells. The number of BS antennas is $N = 64$ and the number of users in each cell is $K = 10$. One can see that ZF-PCP has significantly worse performance than no PCP scenario. Computations show that only at $N$ exceeding $10^6$, ZF-PCP starts giving some gain over no-PCP. However, PCP with numerically optimized coefficients provides a very large improvement over no-PCP scenario. In particular, at the outage probability 5%, optimal PCP provides 7500-fold gain in terms of downlink user rates.

*3) AOA-Based Methods:* As shown in [71]–[73], under realistic channel models, some users with identical or non-orthogonal pilot sequences may have no interference with each other. According to the multipath channel model for linear antenna arrays, the channel vectors from the $k$-th user in the $l$-th cell to the $i$-th BS have the form [74]

$$\mathbf{h}_{i,k,l} = \frac{1}{\Theta} \sum_{\theta=1}^{\Theta} \mathbf{a}(\phi)\eta_{i,k,l}, \quad (41)$$

where $\Theta$ is the number of paths, $\eta_{i,k,l} \sim \mathcal{CN}(0, \sigma_{i,k,l}^2)$ is independent of the path index $\theta$, $\sigma_{i,k,l}^2$ is the user's average path
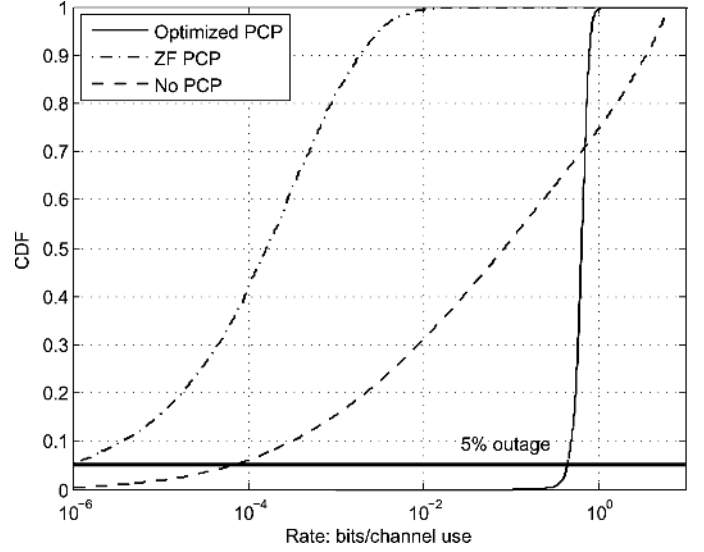


Fig. 5. CDF of downlink user rates for different PCP algorithms.

loss, and $\mathbf{a}(\phi)$ is the steering vector. For a uniform linear array, the steering vector can be expressed as

$$\mathbf{a}(\phi) = \begin{bmatrix} 1 \\ e^{-j2\pi(D/\lambda)cos(\phi)} \\ \vdots \\ e^{-j2\pi((N-1)D/\lambda)cos(\phi)} \end{bmatrix}, \quad (42)$$

where $D$ is the antenna spacing, $\lambda$ is the wavelength of the carrier, and $\phi$ is a random angle-of-arrival (AOA) with *probability density function* (PDF), $f(\phi)$. In [71]–[73], it is shown that users with mutually non-overlapping AOA PDFs hardly contaminate each other even if they use the same pilot sequence. A coordinated scheme for assigning identical pilot sequences only to users of this type is developed in [73]. This scheme achieves a significant reduction in inter-cell interference and a corresponding increase in uplink and downlink SINRs.

*4) Blind Methods:* The blind methods based on subspace partitioning in [75]–[77] can also mitigate pilot contamination. In [75], an eigenvalue-decomposition-based (EVD) channel estimation and iterative least-square with projection (ILSP) estimation of channel vectors is proposed. The EVD-based estimation is based on the assumption that the channel vectors from different users are orthogonal. This assumption allows one to estimate channel vectors using the statistics of the received data. The multiplicative scalar ambiguity inherent to this type of estimation can be resolved by using mutually orthogonal cell pilot sequences assigned to the network cells.

This approach can be summarized as follows. During the uplink transmission, the users first transmit data signals and then cell pilot sequences. The estimation consists of the following steps:

Step 1: The covariance matrix of the received signal is estimated by

$$\mathbf{R}_{\mathbf{y}_l^u} = \mathbb{E}[\mathbf{y}_l^u(\mathbf{y}_l^u)^{\mathrm{H}}] \approx \widehat{\mathbf{R}}_{\mathbf{y}_l^u} = \frac{1}{T} \sum_{t=1}^{\mathrm{T}} \mathbf{y}_l^u[t](\mathbf{y}_l^u[t])^{\mathrm{H}}, \quad (43)$$

where $\mathbf{y}_l^u[t]$ is the received signal by the $l$-th BS at time $t$ defined in (27).

Step 2: A $N \times K$ matrix $\mathbf{U}_l$ is found, whose $k$-th column is the eigenvector of $\widehat{\mathbf{R}}_{\mathbf{y}_l^u}$ corresponding to the eigenvalue that is closest to the quantity, $N\rho_u d_{l,l,k}$. Furthermore, with the help of the pilot sequences, an estimate $\widehat{\Xi}_l$, of the multiplicative matrix, $\Xi_l$ is obtained [75].

Step 3: The estimate $\widehat{\mathbf{G}}_{l,l} = \mathbf{U}_l\widehat{\Xi}_l$ is calculated.

The ILSP estimation is based on the idea of joint iterative estimation of the channel vectors and transmitted data. Let $\mathbf{X}_l$ be the $K \times T$ matrix composed of the $T$ consecutively transmitted signals of the $K$ users from the $l$-th cell, $\mathcal{X}$ be the set of all possible $\mathbf{X}_l$, and $\mathbf{Y}_l$ be corresponding matrix of received symbols. Let $\widehat{H}_{l,l}^{(0)}$ be an initial estimate of $\mathbf{H}_{l,l}$, obtained for example with help from the EVD-based estimation. Then ILSP consists of a multiple repetitions of the following steps:

Step 1: $\widehat{\mathbf{X}}_l^{(i)} = \arg\max_{\mathbf{X}_l \in \mathcal{X}} \|(1/\rho_u)\widehat{\mathbf{H}}_{l,l}^{(i-1)\mathrm{H}} \mathbf{Y}_l - \mathbf{X}_l\|$,

Step 2: $\widehat{\mathbf{H}}_{l,l}^{(i)} = (1/\rho_u)\mathbf{Y}_l\widehat{\mathbf{X}}_l^{(i)\mathrm{H}}, i = i + 1$.

In [76], [77], a blind pilot decontamination scheme is proposed for systems with a power-controlled handoff strategy, and this scheme can separate the interference subspace from the desired signal subspace, resulting in no pilot contamination. Other precoding and detection schemes that require less CSI, such as the Hermitian precoding method in [47], can help reduce pilot contamination as well.

## VI. ENERGY EFFICIENCY

Besides spectral efficiency, massive MIMO technology can improve power efficiency as well [78]. In this section, we first present the power scaling law for massive MIMO systems and then discuss the spectral and energy efficiency tradeoff.

The power scaling law for massive MU-MIMO systems has been derived in [4]. As before, consider uplink transmission in a single-cell system with $K$ single-antenna users, a BS with $N$ antennas where $N \gg K$. The channel vectors from different users are uncorrelated. In this scenario, linear detectors [2], [7], such as MF, ZF and MMSE, are good enough. Based on [4] and (14), for the MF detector with perfect CSI at the BS, the ergodic achievable uplink data rate for the $k$-th user when the number of BS antennas, $N$, goes to infinity is

$$R_k^u \approx \log_2(1 + Nd_k\rho_u) \quad \frac{\frac{\mathrm{bits}}{\mathrm{s}}}{\mathrm{Hz}}. \tag{44}$$

For comparison, the data rate for a user with transmit power $\rho_u$ through a SISO link with large-scale fading coefficient $d_k$ only is

$$R_k^{u,SISO} = \log_2(1 + d_k\rho_u) \quad \frac{\frac{\mathrm{bits}}{\mathrm{s}}}{\mathrm{Hz}}. \tag{45}$$

From (44) and (45), when $N$ is large, the performance of a user with transmit power $\rho_u/N$ in the MU-MIMO system with $N$ antennas at the BS is the same as a SISO system with transmit power $\rho_u$ without small-scale fading. Consequently, the power can be scaled down by $N$ times for one user when perfect CSI is available at the BS. Moreover, the spectral efficiency increases by a factor of $K$ in serving $K$ users simultaneously.

The results based on imperfect CSI at the BS are somewhat different. Based on [4], when an MMSE channel estimate is obtained during the pilot transmission phase and is used for up-

link data detection, the ergodic achievable rate for the $k$-th user based on the MF detector is

$$R_{k,I}^u \approx \log_2(1 + \tau Nd_k^2\rho_u^2) \quad \frac{\frac{\mathrm{bits}}{\mathrm{s}}}{\mathrm{Hz}} \tag{46}$$

as $N \to \infty$, where $\tau$ is the length of the pilot sequence and the pilot power satisfies $\rho_p = \tau\rho_u$. From (45) and (46), the rate of a user with transmit power $\rho_u/\sqrt{N}$ in the MU-MIMO system with $N$ BS antennas is asymptotically the same as the performance under SISO transmission with transmit power $\tau d_k\rho_u^2$ without small-scale fading. Thus, the transmit power of one user can be scaled down by $1/\sqrt{N}$ to get the same performance. Similar results for ZF and MMSE receivers for both perfect and imperfect CSI are also obtained in [4].

In addition to the single-cell scenario, the power scaling law is still valid for multi-cell systems; that is, one user can scale down the transmit power proportional to $1/N$ or $1/\sqrt{N}$ based on perfect and imperfect CSI, respectively, to get the same performance as in the SISO case [4]. Note that this result still holds even with pilot contamination [4].

Energy efficiency is defined as the ratio of the spectral efficiency and the transmit power. The tradeoff for the uplink is studied in [4]. With perfect CSI, the energy efficiency decreases as the spectral efficiency increases. However, a different result is obtained if imperfect CSI is available. In the low transmit power region, energy efficiency increases with the spectral efficiency, while with high transmit power, energy efficiency decreases as the spectral efficiency increases. The tradeoff for the downlink is discussed in [33]. Compared to the MF precoder, ZF can provide better performance in scenarios with high spectral efficiency and low energy efficiency, while the converse holds for scenarios with high energy efficiency and low spectral efficiency.

Circuit power consumption is not considered in [4], [33]. When the circuit power is considered, the question of how to perform antenna selection to improve energy efficiency is investigated in [79] and [80]. In [79], RF chain selection for configurations both with and without CSI is studied to maximize spectral efficiency for a given total power consumption constraint. It is shown that for a MISO case without CSI, the optimal number of RF chains is about half the maximum number of RF chains that can be supported by the power budget. In [80], mutual information based on antenna selection is derived, and the variance of the mutual information is found to decrease as the number of antennas increases. This work also shows that in order to maximize the energy efficiency, all antennas should be used if the circuit power can be ignored compared to the transmit power while only a subset of the antennas should be chosen if the circuit power is comparable to the transmit power.

## VII. SINGLE-CARRIER TRANSMISSION

Existing results for massive MIMO systems mainly focus on flat fading channels, which would imply that OFDM is assumed for wide-band frequency-selective channels. OFDM decomposes the wideband channel into a set of narrow-band flat fading subchannels so that techniques for flat fading channels can be used. However, a significant drawback of OFDM signals is their high PAPR, which can result in low power amplifier

efficiency. Moreover, to deal with *intersymbol interference* (ISI), a *cyclic-prefix* (CP) is inserted in the time-domain, which decreases the spectral and power efficiency. The computational burden for calculating the *discrete Fourier transform* (DFT) for OFDM modulation and demodulation increases with the number of antennas as well. As an alternative, single-carrier transmission has been studied for massive MIMO systems [81], which can reduce complexity incurred by the DFT. Furthermore, single-carrier transmission has a lower PAPR than OFDM. Note that, regardless of whether single-carrier or OFDM modulation is used, the methods in [51], [52] for frequency flat channels and in [82] for frequency selective channels can be used to shape the samples transmitted by each antenna to have constant envelope.

For a frequency-selective single-cell MU-MISO downlink channel, the channel between the $n$-th transmit antenna to the $k$-th user can be modeled as a finite impulse response filter with $\Theta$ taps, where the channel coefficient for tap $\theta$, $h_{k,n}[\theta]$, has the same form as in (7), where the small-scale and large-scale fading components are denoted as $g_{k,n}[\theta]$ and $d_k[\theta]$, respectively. It is assumed that the small-scale fading coefficients remain the same for each block of $M$ symbols and the large-scale fading coefficients are fixed during the entire transmission. Denote $\mathbf{x}_d[t] \in \mathcal{C}^{N \times 1}$ as the transmit signal vector from the BS at time $t$. Then, the received signal vector at time $t$ combined together for all $K$ users can be expressed as

$$\mathbf{y}_d[t] = \sqrt{\rho_d} \sum_{\theta=0}^{\Theta-1} \mathbf{H}^{\mathrm{T}}[\theta]\mathbf{x}_d[t-\theta] + \mathbf{n}_d[t]$$

$$= \sqrt{\rho_d} \sum_{\theta=0}^{\Theta-1} \mathbf{D}^{1/2}[\theta]\mathbf{G}^{\mathrm{T}}[\theta]\mathbf{x}_d[t-\theta] + \mathbf{n}_d[t], \quad (47)$$

where $\mathbf{n}_d[t] = \left[n_1^d[t], \ldots, n_K^d[t]\right]^{\mathrm{T}} \in \mathcal{C}^{K \times 1}$ is additive white Gaussian noise at all users and $\mathbf{H}[\theta] = (h_{k,n}[\theta])_{k,n=1}^{K,N}$ is an $N \times K$ channel matrix for channel tap $\theta$.

To suppress multi-user interference and ISI, the following precoder has been proposed in [81]:

$$\mathbf{x}_d[t] = \sqrt{\frac{1}{NK}} \sum_{\theta=0}^{\Theta-1} \mathbf{H}^*[\theta]\mathbf{s}_d[t+\theta], \quad (48)$$

where $\mathbf{s}_d[t]$ is the information symbol vector at time $t$, each of whose elements represents an information symbol for a given user, and the scalar $\sqrt{1/NK}$ is for power normalization. The proposed precoder in (48) takes the time-inverse and complex-conjugated image of the channel impulse response and has a form similar to MF. It is shown in [81] that multi-user interference and ISI can be efficiently suppressed with low total transmit-power-to-receiver-noise ratio, leading to near-optimal sum rate performance independent of the channel power delay profile. The proposed precoder can also achieve an array power gain proportional to the number of BS antennas.

For the uplink transmission, the operation can be treated as a dual process of the downlink transmission. MF and MMSE detectors can be used for signal detection. The *symbol-error-rate* (SER) performance with *quadrature-phase-shift-keying* (QPSK) modulation for the exponential channel model and TU
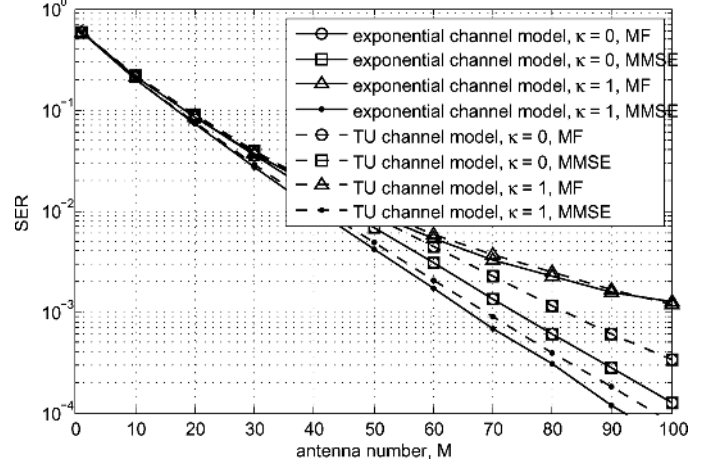


Fig. 6. SER performance of the uplink single carrier transmission.

channel model with 20 taps is shown in Fig. 6, where SNR is $-5$ dB. A LOS path is considered for the first arrival path and the strength of the LOS component is determined by the Rician factor, $\kappa$. In general, the SER performance for exponential channel models is a little better than that for the TU channel model and the MMSE receiver performs better than the MF receiver.

Besides the work in [81], there are some discussions about single-carrier transmission in massive MIMO systems in [56] and [83]. To validate the potentials of single-carrier transmission in massive MIMO systems, further investigation is needed in more general and complicated scenarios.

## VIII. CHALLENGES AND POTENTIALS

To make massive MIMO systems a reality, there are still many issues that need to be studied and addressed. Some of these challenges are discussed below.

### A. Basic Issues

*1) Propagation Models:* Most existing work on massive MIMO is based on the premise that as the number of antennas grows, the individual user channels are still spatially uncorrelated and their channel vectors asymptotically become pairwise orthogonal under favorable propagation conditions. Theoretical studies of massive MIMO typically assume i.i.d. complex Gaussian (Rayleigh fading) conditions [2], [4], which is hard to justify in some situations. A number of experimental studies with multiple antenna arrays have been performed at 2.6 GHz [84]–[86] by employing antennas in small cylindrical or large linear arrays with 128 or 112 elements, respectively, and have led to several interesting observations. From [86], the real antenna correlation coefficients are significantly larger than would be expected under i.i.d. channel assumptions. Furthermore, very highly correlated channel vectors cannot be rendered orthogonal by increasing the number of antennas. This suggests that user scheduling should be a critical component of massive MIMO systems and is much more important than in regular MIMO implementation where more complicated signal processing can be used to separate spatially correlated users. From [85], there is significant performance degradation in the case of closely spaced users with a LOS path to the BS

for different system settings; instead of achieving 80–90% of the optimum dirty paper coding bound, performance drops to about 55%. Nevertheless, both sets of studies have concluded that despite significant differences between the ideal i.i.d. assumption and the measured channels, a large fraction of the theoretical performance gains of large antenna arrays can in practice still be achieved.

*2) TDD and FDD Modes:* As discussed earlier in Section III, research on massive MIMO systems is normally based on the TDD transmission mode due to channel estimation and feedback issues. There are several possible ways to enable FDD mode in massive MIMO systems. One way is to design efficient precoding methods based on partial CSI [87] or even no CSI. Another way is to use the idea of compressed sensing to reduce the feedback overhead [13], [88]. At the BS, the antennas are usually correlated when placing so many antennas in a finite area. The correlation among channel responses from different antennas indicates that we do not need CSI for each individual antenna. Instead, the CSI can be compressed first and then only the necessary information is fed back. The BS can reconstruct the CSI according to the received information. In this way, the overhead for CSI feedback can be greatly reduced. Moreover, even though the uplink and downlink are allocated with different frequencies, they are not always independent of each other. Several strategies use channel reciprocity in FDD systems [89]. Frequency correction algorithms are needed to achieve channel reciprocity in FDD systems. Several frequency correction algorithms, such as frequency correction based on direction of arrival, covariance matrix, and spatio-temporal correlation, have been discussed and compared in [89]. To utilize it in massive MIMO systems, more investigation is needed.

*3) Modulation:* To construct a BS with a large number of antennas, low-cost power-efficient RF amplifiers are necessary, and problems with high PAPR can impede good performance for OFDM [56]. As discussed in Section VII, single-carrier transmission is able to achieve near-optimal sum-rate performance at low-transmit-power-to-receiver-noise-power ratios, without requiring equalization at the receiver and multi-user resource allocation. Whether this is possible for more general and complicated scenarios needs further investigation.

*4) Pilot Contamination:* In a typical multi-cell massive MIMO system, users from neighboring cells may use non-orthogonal pilots. The reason for this is very simple — the number of orthogonal pilots is smaller than the number of users. The use of non-orthogonal pilots results in the pilot contamination problem. Pilot contamination causes directed inter-cell interference, which, unlike other sources of interferences, grows together with the number of BS antennas and significantly damages the system performance. Various channel estimation, precoding, and cooperation methods have been proposed to resolve this issue. However, more efficient methods with good performance, low complexity, and limited or zero cooperation between BSs are worth more intensive study.

*5) Hardware Impairments:* Only limited work has been done on the impact of hardware impairments on massive MIMO systems. Initial studies in [90], [91] have shown that hardware effects can lead to channel estimation error and a capacity ceiling even though a high array gain can still be achieved with an increase in the number of antennas at the BS. The user side impairment is more severe compared to the BS side. Fortunately, the pilot contamination issue in the non-ideal case is actually easier to mitigate compared to the ideal case [91]. Some other initial work has studied the impact of phase noise [92], [93], per-antenna power constraints [51], [52], mutual coupling [94] and hybrid analog/digital beamforming architectures [53], [95], but they are only limited to signal processing models rather than actual transceiver implementations. For massive MIMO to become a reality, more studies on the topic are desired.

*6) Antenna Arrays:* There are several practical issues regarding antenna arrays that are relevant to massive MIMO systems. One concerns the configuration and deployment of the arrays. In [96], a massive MIMO system is proposed where antennas are placed in a 2D grid. Moreover, 3D and distributed array structures are candidates as well. The second issue is the mutual coupling effect among antenna elements. The mutual coupling effect can be ignored only when the antennas are well separated from each other. For massive MIMO systems, antennas may be compactly arranged, and, the coupling issue cannot be ignored [97]. Finally, the increased hardware and computational costs due to the use of very large antenna arrays should be considered. To reduce cost while maintaining performance, an *electromagnetic lens antenna* (ELA) is proposed in [98], [99], integrated to a large antenna array. The proposed ELA can provide spatial multipath separation and energy focusing functions, which can be exploited to improve the performance of massive MIMO and/or reduce its implementation cost and complexity. In general, issues related to array design and implementation are critical for massive MIMO systems.

### B. Application Issues

The use of massive MIMO in wireless networks has the potential to achieve dramatic improvements in capacity and energy efficiency. It also fits very well with applications involving heterogeneous networks and millimeter wave communications, among many others.

*1) Heterogeneous Networks:* In heterogeneous networks (HetNets), low-cost small cells referred to as pico- or femto-cells are flexibly deployed in order to provide dense coverage and ubiquitous high throughput. The use of massive MIMO in coordination with HetNets in order to provide improved interference management and energy efficiency is an important future research direction.

Interference management among coexisting massive MIMO systems and small cells is a critical issue. It is important that a macro-cell BS with a large number of antennas be able to communicate with its own macro-cell users without interfering with users in small cells. A precoding method based on reversed TDD is proposed in [100] to address this issue. In this approach, the macro-cell BS estimates the null space to the small cells during their downlink transmission and then projects its downlink data transmission into the null space of the small cells, generating interference-free transmission. A similar idea has also been proposed before in [101] for cognitive radio networks. However, by employing a large number of antennas at the macro-cell BS, higher degrees of freedom can be used to improve system performance.

Energy efficiency issues in massive MIMO systems with Het-Nets are addressed in [102], [103]. The energy efficiency of massive MIMO systems and HetNets are compared in [102] under the assumption that a given area is covered either by a massive MIMO system or a certain number of small cells. From [102], when the number of cells is large, HetNets are more energy efficient than a corresponding massive MIMO system, but otherwise a massive MIMO system is better. In other settings, users may be simultaneously served by both a massive MIMO BS and small cell BS's. A beamforming method to minimize consumed power under given rate requirements in such a scenario is proposed in [103] when perfect CSI is available and interference coordination is used. Based on the results in [103], it is usually optimal to assign one BS per user even though users could be served by several BS's.

To deploy small cells, a high-capacity and easily accessible backhaul network is required. The use of massive MIMO BSs for wireless backhaul is proposed in [104]. Additional work is necessary to make such an approach more practical.

*2) Millimeter Waves:* Another technology for achieving dramatic improvement in capacity and spectral efficiency is communications over *millimeter wave* (MMW) bands around or above 30 GHz, where the spectrum is less crowded and available bandwidths are broader. MMW technology fits well with massive MIMO and HetNet technology. An encouraging factor is the apparent symbiosis among these three concepts: smaller cell sizes are attractive for operation at MMW frequencies where RF path loss is significantly higher, the shorter wavelength associated with higher frequencies is appealing for massive MIMO designs since the size of the antenna array and associated electronics is reduced, and the large beamforming gain achievable with a very large number of antennas can extend coverage to help overcome the high MMW path loss.

A very thorough MMW measurement campaign has been completed in New York City at 28 GHz and in Austin, Texas, at 38 GHz [105]. Using directional horn antennas for the data collection, the experiments established operational ranges in both outdoor and indoor environments, and provided measurements of path loss exponent, penetration loss through different types of windows and walls, delay spread, number of resolvable multipaths, and reflection coefficients. The results in [105] are promising for MMW MIMO systems, but many important questions remain. For example, propagation at MMW frequencies does not obey a Rayleigh-fading model, it tends to be LOS or near-LOS. This obviously has important implications for the achievable spatial multiplexing gain. Furthermore, the use of MMW frequencies means higher Doppler shifts for a given velocity, and hence potentially shorter coherence times. However, there are two mitigating factors that compensate for this. First, as mentioned above, millimeter-wave systems will primarily be used for relatively short-range applications (e.g., femto- and pico-cells) due to the increased path loss relative to current systems that operate near 2–3 GHz. Consequently, the mobility of users in such systems will be relatively low, and we can expect that the order-of-magnitude increase in carrier frequency will likely be accompanied by a corresponding order-of-magnitude decrease in user velocity. Second, the ability of massive MIMO implementations to achieve very narrow spatial selectivity (e.g.,

"pencil" beams) coupled with (near-)LOS propagation will result in a significant reduction in delay spread, and hence a corresponding increase in coherence bandwidth. Together, these factors should lead to systems that do not require a significant increase in channel update rates, but additional work is needed to confirm this conjecture.

## IX. CONCLUSIONS

In this article, we have comprehensively described massive MIMO systems from several different perspectives. By equipping a BS with a large number of antennas, spectral and energy efficiency can be dramatically improved. However, to make the benefits of massive MIMO a reality, significant additional research is needed on a number of issues, including channel correlation, hardware implementations and impairments, interference management, and modulation.

## REFERENCES

[1] T. L. Marzetta, "Multi-cellular wireless with base stations employing unlimited numbers of antennas," in *Proc. UCSD Inf. Theory Applicat. Workshop*, Feb. 2010.

[2] 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.

[3] E. G. Larsson, "Very large MIMO systems: Opportunities and challenges," 2012 [Online]. Available: http://www.kth.se/polopoly_fs/1. 303070!/Menu/general/column-content/attachment/Large_MIMO.pdf

[4] 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, Apr. 2013.

[5] G. Y. Li, Z.-K. Xu, C. Xiong, C.-Y. Yang, S.-Q. Zhang, Y. Chen, and S.-G. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun. Mag.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[6] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[7] 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. 40–46, Jan. 2013.

[8] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[9] M. Matthaiou, M. R. MacKay, P. J. Smith, and J. A. Nossek, "On the condition number distribution of complex Wishart matrices," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1705–1717, Jun. 2010.

[10] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[11] G. L. Stuber, J. Barry, S. McLaughlin, Y. G. Li, M. A. Ingram, and T. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.

[12] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. 40th Asilomar Conf. Signals, Syst., Comput. (ACSSC)*, Pacific Grove, CA, USA, Oct. 2006, pp. 359–363.

[13] S. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[14] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[15] 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.

[16] Y.-C. Liang, G. M. Pan, and Z. D. 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.

[17] H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multiuser MU-SIMO systems with ZF receivers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4471–4483, Nov. 2013.

[18] M. Matthaiou, C. Zhong, M. R. McKay, and T. Ratnarajah, "Sum rate analysis of ZF receivers in distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 180–191, Feb. 2013.

[19] Y.-C. Liang, S. Sun, and C. Ho, "Block-iterative generalized decision feedback equalizers BI-GDFE for large MIMO systems: Algorithm design and asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2035–2048, Jun. 2006.

[20] K. V. Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 473–485, Apr. 2008.

[21] P. Li and R. D. Murch, "Multiple output selection-LAS algorithm in large MIMO systems," *IEEE Commun. Lett.*, vol. 14, pp. 399–401, May 2010.

[22] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[23] C. Knievel, M. Noemm, and P. A. Hoeher, "Low-complexity receiver for large-MIMO space-time coded systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, San Francisco, CA, USA, Sep. 2011.

[24] R. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "A hybrid RTS-BP algorithm for improved detection of large-MIMO M-QAM signals," in *Nat. Conf. Commun. (NCC)*, Bangalore, India, Jan. 2011.

[25] Q. Zhou and X. Ma, "Element-based lattice reduction algorithms for large MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 274–286, Feb. 2013.

[26] M. Cirkic and E. G. Larsson, "SUMIS: Near-optimal soft-in soft-out MIMO detection with low and fixed complexity," *IEEE Trans. Signal Process.*, 2014, to be published.

[27] P. Suthisopapan, K. Kasai, V. Imtawil, and A. Meesomboon, "Approaching capacity of large MIMO systems by non-binary LDPC codes and MMSE detection," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 1712–1716.

[28] J. Choi, B. Lee, B. Shim, and I. Kang, "Low complexity detection and precoding for massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2857–2861.

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

[30] B. M. Hochwald, C. B. Peel, and A. L. Swindlehust, "A vector-perturbation technique for near-capacity multiantenna communication-part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 5, pp. 537–544, May 2005.

[31] 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.

[32] 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. IEEE Globecom Workshops (GC Workshops)*, Anaheim, CA, Dec. 2012, pp. 233–237.

[33] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[34] V. K. Nguyen and J. S. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," in *Proc. IEEE Global Commun. Conf. (Globecom)*, New Orleans, LA, USA, Dec. 2008.

[35] R. Muharar and J. Evans, "Downlink beamforming with transmit-side channel correlation: A large system analysis," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011.

[36] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4506–4537, Jul. 2012.

[37] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.

[38] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[39] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[40] 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.

[41] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[42] S. Lakshminarayana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic analysis of distributed multi-cell beamforming," in *Proc. IEEE Int. Symp. Personal, Indoorl, Mobile Radio Commun.*, Sep. 2010, pp. 2105–2110.

[43] A. Liu and V. Lau, "Hierarchical interference mitigation for large MIMO cellular networks," *IEEE Trans. Signal Process.*, 2013, submitted for publication.

[44] R. Zakhour and S. V. Hanly, "Min-max fair coordination beamforming via large system analysis," in *Proc. IEEE Proc. Int. Symp. Inf. Theory*, Aug. 2011, pp. 1990–1994.

[45] 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.

[46] 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. 1536–1276, Jun. 2013.

[47] J. Zhang, X. Yuan, and L. Ping, "Hermitian precoding for distributed MIMO systems with individual channel state information," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 241–250, Feb. 2013.

[48] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1137–1141.

[49] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for large scale antenna system: A max-min formulation," in *Proc. the 51th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2013.

[50] A. Adhikary, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for large scale antenna system: Uplink," 2013, accepted for publication.

[51] 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, Nov. 2012.

[52] 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, Mar. 2013.

[53] O. N. Alrabadi, E. Tsakalaki, H. Huang, and G. F. Pedersen, "Beamforming via large and dense antenna arrays above a clutter," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 314–325, Feb. 2013.

[54] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[55] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On the Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[56] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 303–313, Feb. 2013.

[57] S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity precoder for large multiuser MISO systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2008, pp. 797–801.

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

[59] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[60] 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.

[61] K. S. Gomadam, H. C. Papadopoulos, and C.-E. W. Sundberg, *Techniques for Multi-User MIMO with Two-Way Training*, pp. 3360–3366, May 2008.

[62] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[63] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *Proc. Signal Process. Adv. Wireless Commun. (SPAWC)*, San Francisco, CA, USA, Jun. 2011, pp. 381–385.

[64] D. Wang, C. Ji, X. Gao, S. Sun, and X. You, "Uplink sum-rate analysis of multi-cell multi-user massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 5404–5408.

[65] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[66] 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.*, Jun. 2012, pp. 1220–1224.

[67] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with Pilot contamination: Large system analysis," 2013 [Online]. Available: http://arxiv.org/pdf/1307.4388v1.pdf, submitted for publication

[68] Y. Li, Y.-H. Nam, B. L. Ng, and J. C. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Anaheim, CA, USA, Dec. 2012, pp. 4500–4504.

[69] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Cape Town, South Africa, May 2010.

[70] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Interference reduction on cellular networks with large antenna arrays," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012.

[71] M. Filippou, D. Gesbert, and H. Yin, "Decontaminating pilots in cognitive massive MIMO networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Aug. 2013, pp. 816–820.

[72] H. Yin, D. M. F. Gesbert, and Y. Liu, "Decontaminating pilots in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 3170–3175.

[73] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[74] J. A. Tsai, R. Bueher, and B. Woener, "The impact of AoA energy distribution on the spatial fading correlation of linear antenna array," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2002, pp. 933–937.

[75] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimations for multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speed, Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[76] R. R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*, Stuttgart, Mar. 2013.

[77] L. Cottatellucci, R. R. Müller, and M. Vehkaperä, "Analysis of pilot decontamination based on power control," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Dresden, Germany, Jun. 2013.

[78] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[79] Y. Pei, T. Pham, and Y. Liang, "How many RF chains are optimal for large-scale MIMO systems when circuit power is considered?," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Anaheim, Dec. 2012, pp. 3868–3873.

[80] H. Li, L. Song, D. Zhu, and M. Lei, "Energy efficiency of large scale MIMO systems with transmit antenna selection," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013.

[81] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug. 2012.

[82] 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. 10, pp. 547–550, Oct. 2013.

[83] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A novel Monte Carlo sampling based receiver for large-scale uplink multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.

[84] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very large MIMO channels," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, San Francisco, CA, USA, Sep. 2011.

[85] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 295–299.

[86] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Aug. 2012, pp. 811–815.

[87] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," in *Proc. 46th Annu. Conf. Inf. Sci., Syst. (CISS)*, Princeton, NJ, USA, Mar. 2012.

[88] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[89] Y. Han, J. Ni, and G. Du, "The potential approaches to achieve channel reciprocity in FDD system with frequency correction algorithms," in *Proc. Int. ICST Conf. Commun. Netw. in China (Chinacom)*, Beijing, China, Aug. 2010.

[90] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits," in *Proc. Int. Conf. Digital Signal Process. (DSP)*, Fira, Greece, Jul. 2013.

[91] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, 2013, submitted for publication.

[92] A. Pitarokoilis, S. Mohammed, and E. Larsson, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. 50th Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2012, pp. 1190–1197.

[93] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, Jun. 2013, submitted for publication.

[94] X. Artiga, B. Devillers, and J. Perruisseau-Carrier, "Mutual coupling effects in multi-user massive MIMO base stations," in *Proc. IEEE Antennas Propag. Int. Symp. (APSURISI)*, Chicago, IL, USA, Jul. 2012.

[95] O. Ayach, R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, USA, Jun. 2012, pp. 3724–3729.

[96] Y.-H. Nam, B. L. Ng, K. Sayana, Y. Li, J. Zhang, Y. Kim, and J. Lee, "Full-dimension MIMO FD-MIMO for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–179, Jun. 2013.

[97] P. S. Taluja and B. L. Hughes, "Diversity limits of compact broadband multi-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 326–337, Feb. 2013.

[98] Y. Zeng, R. Zhang, and Z. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO," in *Proc. IEEE/CIC Int. Conf. Commun. in China (ICCC)*, Xi'an, China, Aug. 2013, pp. 454–459.

[99] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing enabled massive MIMO: Performance improvement and cost reduction," 2013 [Online]. Available: http://arxiv.org/abs/1312.0718, submitted for publication

[100] K. Hosseini, J. Hoydis, S. Brink, and M. Debbah, "Massive MIMO and small cells: How to densify heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 5442–5447.

[101] R. Zhang, F. Gao, and Y. C. Liang, "Cognitive beamforming made practical: Effective interference channel and learning-throughput tradeoff," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 706–718, Feb. 2010.

[102] W. Liu, S. Han, C. Yang, and C. Sun, "Massive MIMO or small cell network: Who is more energy efficient?," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Shanghai, China, Apr. 2013, pp. 24–29.

[103] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. Int. Conf. Telecommun. (ICT)*, Casablanca, Morocco, May 2013.

[104] T. L. Marzetta and H. Yang, "Dedicated LSAS for macro-cell wireless backhaul – Part I: downlink", Tech. Rep. Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ, USA, 2012.

[105] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access (Available Online at IEEExplore)*, vol. 1, pp. 335–349, 2013.
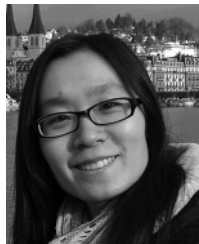
**Lu Lu** (S'09) received her B.S.E degree and M.S.E degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2007 and 2010, respectively. She then got her licentiate degree from Royal Institute of Technology (KTH) in 2011. She is currently working toward the Ph.D. degree with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Her research interests include MIMO, cooperative communications, and cognitive radio networks. She is a co-recipient of the Best Paper Award at ICCCN 2011, Maui, Hawaii.

**Geoffrey Ye Li** (S'93–M'95–SM'97–F'06) received his B.S.E. and M.S.E. degrees in 1983 and 1986, respectively, from the Department of Wireless Engineering, Nanjing Institute of Technology, Nanjing, China, and his Ph.D. degree in 1994 from the Department of Electrical Engineering, Auburn University, Alabama.

He was a Teaching Assistant and then a Lecturer with Southeast University, Nanjing, China, from 1986 to 1991, a Research and Teaching Assistant with Auburn University, Alabama, from 1991 to 1994, and a Post-Doctoral Research Associate with the University of Maryland at College Park, Maryland, from 1994 to 1996. He was with AT&T Labs – Research at Red Bank, New Jersey, as a Senior and then a Principal Technical Staff Member from 1996 to 2000. Since 2000, he has been with the School of Electrical and Computer Engineering at the Georgia Institute of Technology as an Associate Professor and then a Full Professor. He is also holding the Cheung Kong Scholar title at the University of Electronic Science and Technology of China since March 2006.

His general research interests include statistical signal processing and communications, with emphasis on cross-layer optimization for spectral- and energy-efficient networks, cognitive radios and opportunistic spectrum access, and practical issues in LTE systems. In these areas, he has published over 300 refereed journal and conference papers in addition to 25 granted patents. His publications have been cited over 16,000 times from Google Citations and he is listed as a highly cited researcher by Thomson Reuters. He has been involved in editorial activities in more than 10 technical journals for the IEEE Communications and Signal Processing Societies. He organized and chaired many international conferences, including technical program vice-chair of IEEE ICC'03, technical program co-chair of IEEE SPAWC'11, and general chair of IEEE GlobalSIP'14. He has been awarded IEEE Fellow for his contributions to signal processing for wireless communications since 2006. He won the Stephen O. Rice Prize Paper Award in 2010 and the WTC Wireless Recognition Award in 2013 from the IEEE Communications Society. He also received the James Evans Avant Garde Award in 2013 from the IEEE Vehicular Technology Society.

**A. Lee Swindlehurst** (S'83–M'84–SM'99–F'04) received the B.S., summa cum laude, and M.S. degrees in Electrical Engineering from Brigham Young University, Provo, Utah, in 1985 and 1986, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University in 1991. From 1986–1990, he was employed at ESL, Inc., of Sunnyvale, CA, where he was involved in the design of algorithms and architectures for several radar and sonar signal processing systems. He was on the faculty of the Department of Electrical and Computer Engineering at Brigham Young University from 1990–2007, where he was a Full Professor and served as Department Chair from 2003–2006. During 1996–1997, he held a joint appointment as a visiting scholar at both Uppsala University, Uppsala, Sweden, and at the Royal Institute of Technology, Stockholm, Sweden. From 2006–07, he was on leave working as Vice President of Research for ArrayComm LLC in San Jose, California. He is currently the Associate Dean for Research and Graduate Studies in the Henry Samueli School of Engineering and a Professor of the Electrical Engineering and Computer Science Department at the University of California Irvine. His research interests include sensor array signal processing for radar and wireless communications, detection and estimation theory, and system identification, and he has over 240 publications in these areas.

Dr. Swindlehurst is a Fellow of the IEEE, a past Secretary of the IEEE Signal Processing Society, past Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and past member of the Editorial Boards for the EURASIP Journal on Wireless Communications and Networking, *IEEE Signal Processing Magazine*, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is a recipient of several paper awards: the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 and 2010 IEEE Signal Processing Society's Best Paper Awards, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, and is co-author of a paper that received the IEEE Signal Processing Society Young Author Best Paper Award in 2001.

**Alexei Ashikhmin** (M'00–SM'08) is a research scientist in the Communications and Statistical Sciences Department of Bell Labs, Murray Hill, New Jersey. Alexei Ashikhmin's research interests include communications theory, large scale antenna arrays, classical and quantum information theory, and the theory of error correcting codes. From 2003 to 2006, and from 2011 to 2014 Dr. Ashikhmin served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY.

In 2004 Dr. Ashikhmin received S. O. Rice Award for the 2004 best paper of IEEE TRANSACTIONS ON COMMUNICATIONS. In 2002, 2010, and 2011 Dr. Ashikhmin received Bell Laboratories President Awards for breakthrough research in wired and wireless communication projects.

In 2005–2007 Alexei Ashikhmin was an adjunct professor at Columbia University. He taught courses "Error Correcting Code" and "Communications Theory."

**Rui Zhang** (S'00–M'07) received the B.Eng. (First-Class Hons.) and M.Eng. degrees from the National University of Singapore in 2000 and 2001, respectively, and the Ph.D. degree from the Stanford University, Stanford, CA USA, in 2007, all in electrical engineering. Since 2007, he has worked with the Institute for Infocomm Research, A-STAR, Singapore, where he is now a Senior Research Scientist. Since 2010, he has joined the Department of Electrical and Computer Engineering of the National University of Singapore as an Assistant Professor. His current research interests include multiuser MIMO, cognitive radio, cooperative communication, energy efficient and energy harvesting wireless communication, wireless information and power transfer, smart grid, and optimization theory.

Dr. Zhang has authored/coauthored over 180 internationally refereed journal and conference papers. He was the co-recipient of the Best Paper Award from the IEEE PIMRC in 2005. He was the recipient of the 6th IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2010, and the Young Investigator Award of the National University of Singapore in 2011. He is now an elected member of IEEE Signal Processing Society SPCOM and SAM Technical Committees, and an editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.