An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications

Dongming WANG[1], Yu ZHANG[1], Hao WEI[1], Xiaohu YOU[1*], Xiqi GAO[1] & Jiangzhou WANG[2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China;*
[2]*School of Engineering and Digital Arts, University of Kent, Canterbury CT27NT, U.K.*

Abstract To meet the future demand for huge traffic volume of wireless data service, the research on the fifth generation (5G) mobile communication systems has been undertaken in recent years. It is expected that the spectral and energy efficiencies in 5G mobile communication systems should be ten-fold higher than the ones in the fourth generation (4G) mobile communication systems. Therefore, it is important to further exploit the potential of spatial multiplexing of multiple antennas. In the last twenty years, multiple-input multiple-output (MIMO) antenna techniques have been considered as the key techniques to increase the capacity of wireless communication systems. When a large-scale antenna array (which is also called massive MIMO) is equipped in a base-station, or a large number of distributed antennas (which is also called large-scale distributed MIMO) are deployed, the spectral and energy efficiencies can be further improved by using spatial domain multiple access. This paper provides an overview of massive MIMO and large-scale distributed MIMO systems, including spectral efficiency analysis, channel state information (CSI) acquisition, wireless transmission technology, and resource allocation.

Keywords the fifth generation mobile communication, massive MIMO, large-scale distributed antenna systems, spectral efficiency, channel state information acquisition, multi-user MIMO, resource allocation

1 Introduction

In recent years, mobile data traffic has been increasing almost exponentially. According to the most recent prediction from Ericsson, global mobile data traffic will increase nearly eleven-fold from 2015 to 2020 [1]. At the same time, as the energy consumption of information and communications technology becomes large, it is very urgent to reduce the energy consumption of mobile communication systems. This definitely leads to a great challenge for current fourth generation (4G) mobile communications, and also provides a big chance for the fifth generation (5G) mobile communication systems.

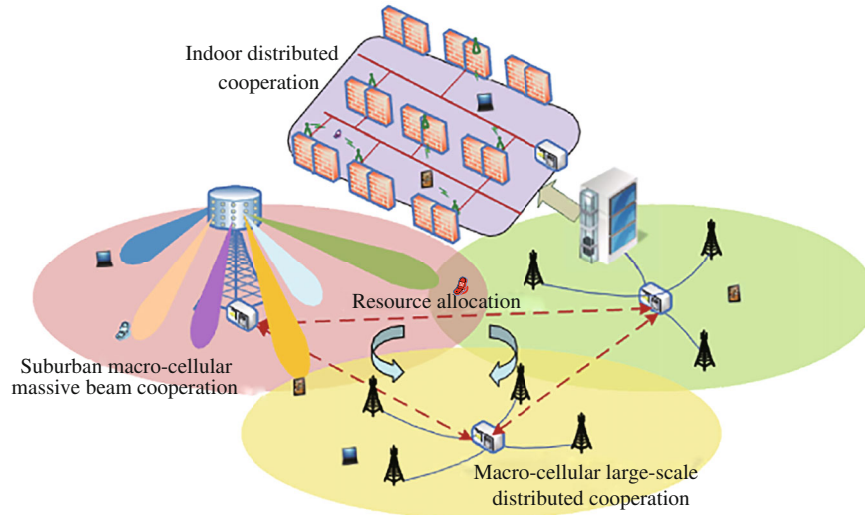* Corresponding author (email: xhyu@seu.edu.cn)

Figure 1 (Color online) Large-scale cooperative wireless communications.

In order to meet the future demand of the mobile data service, one of the fundamental objectives of 5G mobile communication systems is to further improve the spectral and energy efficiencies by one order of magnitude higher than the ones in 4G mobile communication systems. This requires new revolution of network architecture and wireless transmission technologies, to fundamentally address the problem of the spectral and energy efficiencies of mobile communication systems, as well as achieve the goals of higher data rate and green wireless communications [2,3].

Multiple-input multiple-output (MIMO) antennas, as a breakthrough communication technique in the past 20 years, is a fundamental approach to exploit spatial domain resource. MIMO offers diversity gain, multiplexing gain and power gain [4], to improve the reliability, support the spatial multiplexing of both single and multiple users, and increase the energy efficiency through beamforming techniques, respectively. So far, MIMO technology has been adopted by third generation partnership project (3GPP) long-term evolution (LTE), IEEE 802.11ac and other wireless communication standards. However, for 4G mobile communication systems, as only a small number of antennas is equipped in base stations (BSs) [5], the spatial resolution is limited and then the performance gain is not fully exploited. Furthermore, under the current system configuration, the implementation of the capacity-approaching transmission scheme is extremely difficult.

Distributed antenna system (DAS) is another approach to exploit spatial dimension resources [6–9]. In DAS, geographically dispersed remote radio units (RRUs) are equipped with multiple antennas, and are connected to a baseband unit (BBU) through high-speed backhaul links. Similar to MIMO, with the cooperation between RRUs, DAS can serve single or multiple mobile terminals in the same time-frequency resource. Then, it is also called as distributed MIMO system or cooperative MIMO system. Distributed MIMO technology could not only obtain three types of gains of MIMO, but also get the macro-diversity and the power gain due to smaller path loss [10–12].

To further improve the spectral and energy efficiencies of 4G system, both industry and academia have reached a consensus to increase the number of cooperative RRUs at hot spots [13, 14], or replace the current multiple antennas with a large-scale antenna array in each BS [5]. Both the technologies can be viewed as a large-scale collaborative wireless communication shown in Figure 1.

As shown in Figure 1, a large number of antennas could be dispersed within a cell (called large-scale distributed MIMO), or centrally deployed at a BS (referred to as massive MIMO). Theoretical research and preliminary performance assessments [5,14] demonstrated that as the number of BS antennas (or the number of distributed RRUs) becomes infinity, inter-user channels will become orthogonal. In this case, Gaussian noise and inter-cell interference from other cells will be averaged out to zero and the transmit power of any user can be arbitrarily low. The system capacity is only limited by the reuse of the pilot

resource.

In both massive MIMO and large-scale distributed MIMO systems, there exists a consensus on exploitation of wireless spatial dimension resources. That is, they have similar problems in the performance analysis and system design. Thus, in this paper, if not specifically stated, both massive MIMO and large-scale distributed MIMO systems are referred to as large-scale antenna systems (LSAS). Different from traditional MIMO system, LSAS introduces the challenges on both theoretical analysis and system design, which include spectral efficiency analysis under the constrain of pilot resource, channel state information (CSI) acquisition, wireless transmission technology, resource allocation, etc.

In the following, we will present the state of the art in LSAS. Firstly, we introduce its spectral efficiency analysis with imperfect CSI. Secondly, its bottleneck technology, CSI acquisition techniques including pilot design, channel estimation and reciprocity acquisition methods, is highlighted. Then, its transmission technology, consisting of multi-user precoding and joint multi-user detection, is reviewed. Finally, some resource allocation methods are introduced.

The symbols used in this paper are defined as follows. All boldface italic letters stand for vectors (lower case) or matrices (upper case). $\boldsymbol{I}$ denotes any suitable unit matrix. $[\boldsymbol{A}]_{ij}$ represents the ith row and jth column element of the matrix $\boldsymbol{A}$. The superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively denote the matrix transpose, conjugate and conjugate transpose. The operator $\otimes$ denotes the Kronecker product. $\mathrm{vec}(\boldsymbol{A})$ denotes vectorization of a matrix. $\mathrm{Tr}(\boldsymbol{A})$ and $\det(\boldsymbol{A})$ stand for determinant and trace, respectively. The operator $\mathrm{diag}(\boldsymbol{x})$ denotes the diagonal matrix with $\boldsymbol{x}$ along its main diagonal. A block diagonal matrix denoted as $\mathrm{diag}[\boldsymbol{A}_1 \cdots \boldsymbol{A}_n]$ has blocks $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ along its main diagnal. The operator $\mathcal{E}(\cdot)$ denotes expectation and the covariance operator is given by $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \mathcal{E}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}) - \mathcal{E}(\boldsymbol{x})\mathcal{E}(\boldsymbol{y}^{\mathrm{H}})$. $\mathcal{CN}(\mu, \sigma^2)$ stands for complex Gaussian distribution with mean μ and variance σ^2.

2 Spectral efficiency analysis of LSAS

Capacity analysis is a basis for system design. In the past 20 years, many researchers have studied the capacity of MIMO. Random matrix theory plays an important role in the theoretical analysis of MIMO capacity. The fundamental mathematical tools include the statistical properties of Wishart matrix and its eigenvalues [15], the asymptotic statistical properties of large dimensional random matrices [15], free probability theory [16] and the deterministic equivalent method [17]. With these mathematical tools, extensive research has been done on the capacity of different MIMO channels, including multi-user MIMO (MU-MIMO) and distributed MU-MIMO channels.

However, most of the studies on MIMO capacity are focused on perfect CSI at receiver side. Since a bottleneck of multi-user large-scale MIMO systems is CSI acquisition, it is necessary to study information theory under the constrained pilot resource. The spectral and energy efficiencies in both uplink and downlink massive MIMO systems were analyzed in [18], where the impact of imperfect CSI on the performance was demonstrated. In [19], the spectral efficiency of massive MIMO systems with different transceivers has been studied with the consideration of pilot contamination. The theoretical results revealed the relationship between the number of BS antennas per user and the transceiver techniques.

In the following, we consider multi-cell multi-user DAS as an example, of which massive MIMO is its special case. We assume that the pilot resource is limited and pilot sequences are reused by the users in the systems. We present a closed-form expression of asymptotic capacity. Furthermore, we also give some research directions.

2.1 Theoretical analysis of spectral efficiency

Consider multi-cell multi-user DAS with L hexagonal cells, as shown in Figure 2. Each cell consists of N RRUs and K single-antenna users. Each RRU is equipped with M antennas. All the RRUs in each cell are connected to a BBU through optical fibers. In this paper, cell 1 is the reference cell. The uplink channel vector between the kth user of cell l to all of the RRUs of the cell 1 is denoted as

$$\boldsymbol{g}_{l,k} \triangleq \boldsymbol{R}_{l,k}^{\frac{1}{2}} \boldsymbol{\Lambda}_{l,k}^{\frac{1}{2}} \boldsymbol{h}_{l,k}, \tag{1}$$
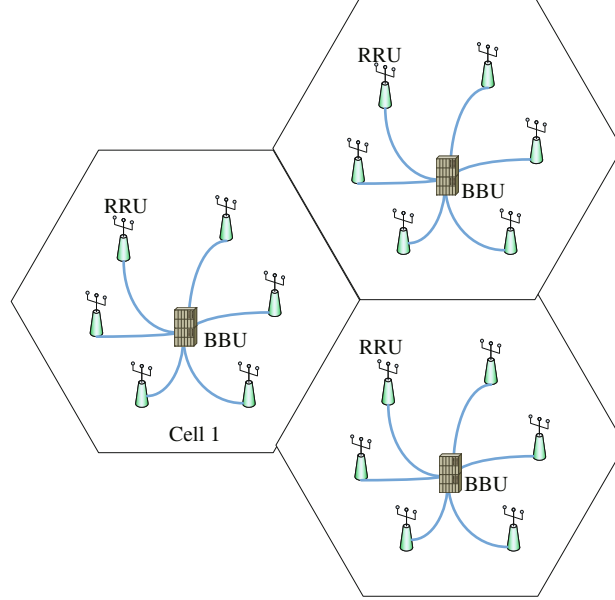
Figure 2 (Color online) Multi-cell multi-user large-scale distributed antenna system.

where

$$\boldsymbol{\Lambda}'_{l,k} \triangleq \operatorname{diag}\left[\ \lambda_{l,1,k}\ \cdots\ \lambda_{l,N,k,}\ \right], \tag{2}$$

$$\boldsymbol{\Lambda}_{l,k} \triangleq \boldsymbol{\Lambda}'_{l,k} \otimes \boldsymbol{I}_M, \tag{3}$$

$$\boldsymbol{R}_{l,k} \triangleq \operatorname{diag}\left[\ \boldsymbol{R}_{l,1,k}\ \cdots\ \boldsymbol{R}_{l,N,k}\ \right], \tag{4}$$

$$\boldsymbol{h}_{l,k} = \left[\ \boldsymbol{h}_{l,1,k}^{\mathrm{T}}\ \cdots\ \boldsymbol{h}_{l,N,k}^{\mathrm{T}}\ \right]^{\mathrm{T}}. \tag{5}$$

$\lambda_{l,n,k}$ represents the large-scale and shadow fading between the kth user of cell l and the nth RRU of cell 1 and $\boldsymbol{\Lambda}'_{l,k}$ is a diagonal matrix representing the large-scale and shadow fading matrix between the kth user of cell l and all the RRUs of cell 1. We assume the links between the user and all the antennas of a RRU have the same large-scale and shadow fading and $\boldsymbol{\Lambda}_{l,k}$ represents the large-scale and shadow fading matrix between the kth user of cell l and all the antennas of N RRUs of cell 1. $\boldsymbol{R}_{l,n,k}$ is an $M \times M$ receive correlation matrix of the kth user of cell l to the nth RRU of cell 1 and $\boldsymbol{R}_{l,k}$ is an $MN \times MN$ block diagonal receive correlation matrix of the kth user of cell l to all the RRUs of cell 1. $\boldsymbol{h}_{l,n,k}$ represents the small-scale fading between the kth user of cell l and the nth RRU of cell 1, and it is a vector of size M which contains independent identically distributed (i.i.d.) zero mean circularly symmetric complex Gaussian (ZMCSCG) random variables with unit variance. $\boldsymbol{h}_{l,k}$, an $MN \times 1$ vector, represents the small-scale fading between the kth user of cell l and all the antennas of N RRUs of cell 1. The channel matrix between K users of cell l and all antennas of N RRUs of cell 1 is defined as $\boldsymbol{G}_l = [\boldsymbol{g}_{l,1} \cdots \boldsymbol{g}_{l,K}]$.

We assume the BBU obtains uplink CSI by uplink pilot sequences from the users. $K \times K$ identity matrix is used as the pilot sequences set for K users of any cell. That is, for inner-cell users, the pilot sequences are orthogonal. For example, a user sends pilot symbol 1 on a time-frequency resource when other users in the same cell send pilot symbol 0. Since the pilot sequences are reused by the other cell users, there is a severe pilot contamination among the adjacent cells.

Because $\mathcal{E}[\boldsymbol{g}_{l,k}\boldsymbol{g}_{l',k'}^{\mathrm{H}}] = 0$, for $k \neq k'$, the estimation of $\boldsymbol{g}_{l,k}$ can be processed individually. To estimate $\boldsymbol{g}_{l,k}$, we have the following observations:

$$\boldsymbol{y}_{\mathrm{P},k} = \boldsymbol{g}_{l,k} + \sum_{i \neq l} \boldsymbol{g}_{i,k} + \boldsymbol{z}_{\mathrm{P},k}, \tag{6}$$

where $\boldsymbol{z}_{\mathrm{P},k}$ is the equivalent Gaussian noise with variance γ_{P}. From [20], we can conclude that

$$\hat{\boldsymbol{g}}_{l,k} = \boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\boldsymbol{Q}_k^{-1}\boldsymbol{y}_{\mathrm{P},k}, \quad l = 1, \ldots, L, \tag{7}$$

$$\boldsymbol{Q}_k = \sum_{i=1}^{L} \boldsymbol{R}_{i,k}\boldsymbol{\Lambda}_{i,k} + \gamma_{\mathrm{P}}\boldsymbol{I}_{MN}. \tag{8}$$

Channel estimation error is defined as $\tilde{\boldsymbol{g}}_{l,k} = \boldsymbol{g}_{l,k} - \hat{\boldsymbol{g}}_{l,k}$ and its covariance matrix can be expressed as

$$\mathrm{cov}\left(\tilde{\boldsymbol{g}}_{l,k}, \tilde{\boldsymbol{g}}_{l,k}\right) = \boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k} - \boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\boldsymbol{Q}_k^{-1}\boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}. \tag{9}$$

Define $\hat{\boldsymbol{h}}_k \triangleq \boldsymbol{Q}_k^{-\frac{1}{2}}\boldsymbol{y}_{\mathrm{P},k}$. It can be seen that $\hat{\boldsymbol{h}}_k$ obeys $\hat{\boldsymbol{h}}_k \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_{MN}\right)$ and is not related to l. Then, the channel vector can be modelled as

$$\hat{\boldsymbol{g}}_{l,k} = \boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\boldsymbol{Q}_k^{-\frac{1}{2}}\hat{\boldsymbol{h}}_k. \tag{10}$$

After channel estimation, the channel vector between the kth user of cell l and all of the RRUs of cell 1 can be expressed as a correlated fading channel with the Rayleigh fading part $\hat{\boldsymbol{h}}_k$. The main difference between the realistic channel model and the equivalent channel model is that the Rayleigh fading part of $\hat{\boldsymbol{g}}_{l,k}$ is not related to l while that of $\boldsymbol{g}_{l,k}$ is statistically independent for any l.

By defining the following channel matrix $\hat{\boldsymbol{G}}_l = [\hat{\boldsymbol{g}}_{l,1} \cdots \hat{\boldsymbol{g}}_{l,K}]$, the relationship between the estimated channel matrix $\hat{\boldsymbol{G}}_l$ and the transmit and receive signal can be expressed as

$$\boldsymbol{y}_1 = \hat{\boldsymbol{G}}_1\boldsymbol{x}_1 + \sum_{l=2}^{L}\hat{\boldsymbol{G}}_l\boldsymbol{x}_l + \sum_{l=1}^{L}\tilde{\boldsymbol{G}}_l\boldsymbol{x}_l + \boldsymbol{z}_1. \tag{11}$$

Using the minimum mean square error (MMSE) receiver, the sum-rate can be written as [21]

$$C = \log_2 \det\left(\sum_{l=1}^{L}\hat{\boldsymbol{G}}_l\hat{\boldsymbol{G}}_l^{\mathrm{H}} + \boldsymbol{\Sigma}\right) - \log_2 \det\left(\sum_{l=2}^{L}\hat{\boldsymbol{G}}_l\hat{\boldsymbol{G}}_l^{\mathrm{H}} + \boldsymbol{\Sigma}\right), \tag{12}$$

where

$$\boldsymbol{\Sigma} = \sum_{l=1}^{L}\sum_{k=1}^{K}\left(\boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k} - \boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\boldsymbol{Q}_k^{-1}\boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\right) + \gamma_{\mathrm{UL}}\boldsymbol{I}_{MN}, \tag{13}$$

and γ_{UL} is the Gaussian noise variance in the uplink channel. As mentioned above, the equivalent model of multi-cell multi-user LSAS was given with imperfect CSI. Obviously, the above formulas are similar to that of MIMO channel capacity with multi-cell interference. However, it is more challenging to thoeretically analyze the capacity in LSAS, owing to the correlation of $\hat{\boldsymbol{G}}_l$. It has been proved in [21] that as $MN \to \infty$, the sum-rate satisfies

$$C - C_{\mathrm{inf}} \to 0, \tag{14}$$

where

$$C_{\mathrm{inf}} = \sum_{k=1}^{K}\log_2\left[1 + \xi_{1,1,k} - \boldsymbol{\xi}_{1,2:L,k}^{\mathrm{T}}(\boldsymbol{\Xi}'_k + \boldsymbol{I}_{L-1})^{-1}\boldsymbol{\xi}_{2:L,1,k}\right], \tag{15}$$

$$\xi_{i,l,k} \triangleq \mathrm{Tr}\left(\boldsymbol{Q}_k^{-1}\boldsymbol{\Lambda}_{i,k}\boldsymbol{R}_{i,k}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{l,k}\boldsymbol{\Lambda}_{l,k}\right), \tag{16}$$

$$\boldsymbol{\Xi}' = \begin{bmatrix} \xi_{2,2,k} & \cdots & \xi_{2,L,k} \\ \vdots & \ddots & \vdots \\ \xi_{L,2,k} & \cdots & \xi_{L,L,k} \end{bmatrix}, \tag{17}$$

$$\boldsymbol{\xi}_{1,2:L,k} = \begin{bmatrix} \xi_{1,2,k} & \cdots & \xi_{1,L,k} \end{bmatrix}^{\mathrm{T}}, \tag{18}$$

$$\boldsymbol{\xi}_{2:L,1,k} = \left[\begin{array}{ccc} \xi_{2,1,k} & \cdots & \xi_{L,1,k} \end{array} \right]^{\mathrm{T}}. \tag{19}$$

The above equations are general expressions for the capacity of multi-cell multi-user DAS, which covers those of massive MIMO. It is easy to verify that the expression of the sum-rate in [20] is a special case of (15). In particular, when $\boldsymbol{R}_{l,k} = \boldsymbol{I}_M$ for all l and k, $N = 1$, and $M \to \infty$, C_{inf} satisfies

$$C_{\mathrm{inf}} = \sum_{k=1}^{K} \log_2 \left[1 + \frac{\lambda_{1,1,k}^2}{\sum_{l=2}^{L} \lambda_{l,1,k}^2 + \frac{\varepsilon + \gamma_{\mathrm{UL}}}{M} \left(\sum_{l=1}^{L} \lambda_{l,1,k} + \gamma_{\mathrm{P}} \right)} \right], \tag{20}$$

where

$$\varepsilon = \sum_{k=1}^{K} \sum_{l=1}^{L} \lambda_{l,1,k} - \sum_{k=1}^{K} \left(\sum_{l=1}^{L} \lambda_{l,1,k}^2 \right) \left(\sum_{l=1}^{L} \lambda_{l,1,k} + \gamma_{\mathrm{P}} \right)^{-1}. \tag{21}$$

It can be seen that when $M \to \infty$,

$$C_{\mathrm{inf}} = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\lambda_{1,1,k}^2}{\sum_{l=2}^{L} \lambda_{l,1,k}^2} \right), \tag{22}$$

which is the same as the sum-rate with the maximal ratio combining (MRC) receiver [5].

The above theoretical results show that the capacity increases with the number of users, when the number of antennas at the BS is larger than the number of users. However, if orthogonal pilots are used in the uplink, pilot overhead increases linearly with the number of users, and then there exists a largest number of users served by the system. It can also be seen that LSAS is interference-limited and its performance is limited by the users using the same pilot sequence in other cells. The downlink capacity of large-scale DAS was studied with maximun ratio transmission (MRT) precoder according to the above model with pilot contamination in [22]. The results in [21, 22] demonstrated that under the same condition, the uplink capacity of large-scale distributed MIMO systems has 100% improvement than that of massive MIMO systems with MMSE receiver, while the improvement is 50% with MRT in the downlink.

2.2 System-level spectral efficiency

System-level spectral efficiency is a key parameter indicator of cellular mobile communication systems in industry, and is usually obtained by complex and time-consuming system simulation. In recent years, in order to demonstrate the impact of the system parameters, some researchers have studied the theoretical analysis of the system-level spectral efficiency.

There are two kinds of methods to study system-level spectral efficiency. The first one is using the stochastic geometry. As a representative paper, Ref. [23] derived the system-level spectral efficiency of multi-cell cellular systems. The main idea was that the spectral efficiency is obtained by Shannon formula according to signal to interference plus noise ratio (SINR), and then assuming that the BS locations are modeled as Poisson point process, the system-level spectral efficiency can be given by an elegant closed-form expression. The stochastic geometry model attracts more and more attention. For example, it has been applied to spectral efficiency analysis of cooperative BS systems [24], capacity analysis of heterogeneous network [25], capacity analysis of DAS [26], and capacity analysis of massive MIMO systems [27]. However, in cooperative BSs or DAS, the closed expressions obtained by using Poisson point process is too complex to give an explicit relationship between system-level spectral efficiency and system parameters. Furthermore, when imperfect CSI (especially pilot contamination) is considered, due to the complicated expression of SINR, the analysis of system-level spectral efficiency has not been seen.

Another method is based on the model which has been widely used in the system-level simulation. The theoretical analysis has two steps: first, the locations of the users in the system are fixed, and the ergodic spectral efficiency is obtained by averaging the spectral efficiency over the small-scale fading. Obviously, the ergodic spectral efficiency is a function of the users' location. After that, assuming that the BS

locations are fixed and all the users are uniformly distributed in every cell, we obtain the average of the ergodic spectral efficiency. The closed expressions for both distributed and centralized antenna systems were firstly derived in [28]. In [29], the average spectral efficiency of DAS was further obtained with all RRUs on a circle. According to asymptotic upper and lower bounds of system-level spectral efficiency for non-cooperative multi-cell multi-user cellular systems, the relationship between system-level spectral efficiency and system parameters was derived (including bandwidth, the number of users and RRUs) in [30]. In [31], an approximate closed-form expression of average spectral efficiency for multi-cell multi-user massive MIMO systems was given when pilot reuse was considered.

2.3 Topics for future research

When pilot contamination is considered, spectral efficiency analysis should be further studied for large-scale DAS with zero forcing (ZF) or regularized ZF (RZF) precoding. Recently, the impact of non-ideal hardware on system capacity attracts attention [32, 33]. Moreover, it was found that non-reciprocity between uplink and downlink has a substantial impact on spectral efficiency for time division duplex (TDD) systems. Finally, considering imperfect CSI, asymptotic analysis of system-level spectral efficiency is also a challenging research direction.

3 CSI acquisition in LSAS

In LSAS, with an increasing number of BS antennas and spatial-division users, CSI acquisition becomes a bottleneck in system implementation. When orthogonal pilots are employed, pilot overhead increases linearly with the total number of antennas of the users or the total number of antennas at BS for uplink or downlink, respectively. To achieve the capacity of broadcast channel, CSI of downlink should be known at BS. However, since the number of BS antennas may be much larger than the total number of users' antennas, downlink CSI acquisition becomes an obstacle for LSAS. For TDD systems, during the coherence time, due to the reciprocity of the uplink and downlink channels, the estimation of uplink channels at BS can be utilized to design downlink precoding. That means the overhead of both downlink pilots and the CSI feedback can be avoided for TDD systems. While for frequency division duplex (FDD) systems, the lack of the reciprocity becomes a main challenge.

CSI acquisition in LSAS is facing the following problems. First, the pilot overhead still linearly increases with the total number of users' antennas, even for TDD systems. How to reduce the pilot overhead or utilize pilot resource efficiently should be studied. Second, under the limited pilot resource, how to improve the accuracy of channel estimation is also important. Finally, although for TDD system, the uplink/downlink wireless channels are reciprocal, the overall channels are nonreciprocal due to the mismatch of transceiver radio frequency (RF) circuits. In the following, we will give the detailed investigation on the pilot design, channel estimation methods, and reciprocity calibration for LSAS.

3.1 Pilot design

Since CSI plays an important role in both transmitter and receiver, reference signal design has always been the key technology for mobile communication systems. According to the functions of reference signals, they can be divided into CSI reference signal (CSI-RS) and demodulation reference signal (DM-RS). CSI-RS is usually sent omnidirectionally, which can be used to obtain the channel quality indicator, or the statistical channel information, etc. CSI-RS is usually sparse in time and frequency domain. DM-RS are mainly served as data demodulation. To reduce the overhead, precoded pilots are usually adopted for DM-RS.

Pilot design is usually divided into two kinds: orthogonal design and non-orthogonal design. Orthogonal pilot sequences have been widely used in wireless communication systems. The idea of orthogonal design includes time division multiplexing, frequency division multiplexing, code division multiplexing and their combinations. These techniques have been applied in 3GPP LTE standard. The advantage of

orthogonal design is that it can achieve optimal performance with least square channel estimation. However, the disadvantage is that the overhead is large for LSAS, especially for multi-cell multi-user LSAS. Hence, there is an urgent request to reduce the pilot overhead. Nonorthogonal pilot sequences are often adopted to reduce the pilot overhead. For LSAS, two solutions have been proposed. One uses the idea of superposition of pilot symbols and data symbols, and another solution is pilot reuse. The former will introduce extra interference between pilot and data, and the later will suffer from pilot contamination.

In [34], a multi-cell time-shifted pilot scheme was presented, and the interference between pilots and data can be suppressed by interference cancellation. Ref. [34] showed that with this method, pilot contamination can be mitigated when the number of BS antennas approaches infinity. However, when the number of spatial-division users is large, for time-shifted pilot systems, to achieve the same performance as the synchronized pilot systems, there is need for more antennas in the BS. In [35], the semi-orthogonal pilot design was proposed to allow simultaneous data and pilot transmission by using the successive interference cancellation. Simulation and theoretical analysis showed that the method can greatly reduce the pilot overhead and then improve the spectral efficiency. The idea behind [34, 35] is the superposition of pilot and data.

Pilot reuse [5] has been studied since massive MIMO was firstly proposed. In [36], with long-term channel information, such as large-scale fading, three low-complexity pilot scheduling approaches were further proposed to maximize the achievable sum rate, including the greedy algorithm, the tabu search (TS) algorithm and the greedy TS algorithm. The research on massive MIMO channel model showed that there are sparse characteristics in the spatial domain and time domain when both the number of antennas and the bandwidth greatly increase. Pilot assignment using the sparsity of the massive MIMO channels can mitigate pilot contamination effectively. In [37], it has been proved that the mean square error of channel estimation can be minimized, provided that the angle of arrival intervals of users are non-overlapping. Therefore, pilot reuse in spatial correlated massive MIMO channels is feasible when channels are sparse in the angle domain. Moreover, the second-order statistics of channels can be exploited to improve the performance of pilot assignment [36–38], and therefore, pilot contamination can be mitigated. In the time domain, the sparsity of the wideband MIMO channels can also be used to suppress pilot contamination [39, 40].

For large-scale DAS, the distributed MIMO channels are sparse in the power domain [14]. We can also use this property of distributed MIMO to do pilot assignment. With the users' locations, an inter-user interference (IUI) matrix can be constructed to quantify the inter-user pilot interference. From this point of view, pilot reuse and frequency allocation are similar. Some methods used in frequency allocation, such as fractional frequency reuse or the advanced graph coloring algorithm, can be employed to reduce the pilot contamination in distributed MIMO systems [41, 42].

For CSI acquisition in massive MIMO systems, there are more challenges in FDD than in TDD [43, 44]. Assuming that the BS and the users share a common set of training signals, open-loop and closed-loop training frameworks were proposed in [45]. In open-loop training, the BS transmits training signals in a round-robin manner, so that the receiver can estimate the current channel using spatial or temporal correlations as well as previous channel estimations. In closed-loop mode, the user feeds back the sequence number of the best training signal which is selected from the previously received training signals. Then, with the index feedback from the users, the BS determines the training signal to be sent out.

3.2 Channel estimation methods

Exploiting the sparsity of massive MIMO channel is an efficient way to improve the performance of channel estimation. Based on a physically motivated channel model, a critical relationship for massive MIMO orthogonal frequency division multiplexing (OFDM) has been revealed between the space-frequency domain channel covariance matrix (SFCCM) and the channel power angle-delay spectrum [46]. When the number of antennas at BS is sufficiently large, the SFCCM can be well approximated by [46]

$$\boldsymbol{R}_k \approx \left(\boldsymbol{F}_{N_c \times N_g} \otimes \boldsymbol{V}_M \right) \operatorname{diag} \left[\operatorname{vec} \left(\boldsymbol{\Omega}_k \right) \right] \left(\boldsymbol{F}_{N_c \times N_g} \otimes \boldsymbol{V}_M \right)^{\mathrm{H}}, \tag{23}$$

where N_c is the number of subcarriers, N_g is the length of the guard interval, $\boldsymbol{F}_{N_c \times N_g}$ denotes the matrix composed of the first N_g columns of N_c dimensional unitary discrete Fourier transform (DFT) matrix, $\boldsymbol{V}_M$ is a matrix containing the spatial information, and $\boldsymbol{\Omega}_k$ is referred to the angle-delay domain channel power matrix, which describes the sparsity of wireless channels in the angle-delay domain. Therefore, the eigenvectors of the SFCCMs for different users tend to be the same in the asymptotically large array regime, which shows that massive MIMO-OFDM channels can be asymptotically decorrelated by the fixed space-frequency domain statistical eigendirections, while the eigenvalues depend on the corresponding channel power angle-delay spectrum. Furthermore, with a presented equivalent channel model, the sum of mean square error of channel estimation can be minimized if the channel power distributions of different users in the angle-delay domain can be made non-overlapping by proper pilot phase shift scheduling.

Since channels in wideband massive MIMO have sparsity in both the angular [47] and delay domain [39], a parametric channel model can be utilized to represent this type of channels. Then, subspace channel estimation methods [48, 49] can be applied to parameterized channel estimation. In [50] , more accurate channel estimates could be obtained through direction of arrival and direction of departure estimation of resolvable paths.

Compress sensing, as another effective way to estimate sparse channels, can achieve more accurate channel estimation with lower pilot overhead. In [51], with the inherent sparsity of wireless channels, the channel estimation was modeled as a joint sparse recovery problem, which can be solved by an improved algorithm named block optimized orthogonal matching pursuit. In [52], a distributed Bayesian method, which is based on the support agnostic Bayesian matching pursuit algorithm [53], has been developed to improve channel estimates in massive MIMO. This approach essentially exploits the channel sparsity and common support properties to estimate sparse channels with a small number of pilots.

The problem of pilot contamination is very similar to the channel estimation problem in code division multiple access (CDMA) systems. Blind channel estimation which has been extensively studied in CDMA system by many researchers was reconsidered to solve the problem of the pilot overhead for LSAS. In TDD systems, with spatial asymptotic orthogonality of channels in massive MIMO systems, Ngo et al. proposed a blind channel estimation algorithm based on eigenvalue decomposition [54]. Using only a few uplink pilots, it can remove the ambiguity of the blind channel estimation. Based on subspace projection, a blind channel estimation method without pilot symbols was presented to further overcome pilot contamination [55]. However, the computational complexity becomes a major obstacle for the practical implementation.

In addition, data-aided channel estimation is also a traditional method to improve the performance of channel estimation. The performance can be further enhanced by using iterative receiver with joint data detection and channel estimation [56]. Nevertheless, due to the non-orthogonality of the data symbols, the performance improvement can be obtained only when the data length is long enough. Correspondingly, in return, its computational complexity increases significantly with the number of users and the data length.

By exploiting the sparsity in the time domain, a joint design of channel estimation and pilot assignment was proposed to mitigate pilot contamination as well as improve the performance in [39]. The main idea behind this method is that with the orthogonal property of different users in the time-delay domain, pilot contamination is randomized by pilot assignment in the different time slots. Then, the power delay profile (PDP) is estimated via multipath time-delay estimation and multipath component extraction. In the ideal case, the estimated PDP approximates to that without pilot contamination. Finally, pilot contamination can be eliminated by the estimated PDP.

In an FDD system, feedback is a typical way to improve the accuracy of channel estimation. In [57], a distributed compressive CSI estimation scheme was proposed, where the compressed measurements were observed at the users locally, while the CSI recovery was performed at the BS jointly. Considering the spatial sparsity and slow-varying characteristics, a distributed sparsity adaptive matching pursuit algorithm was proposed in [58]. By exploiting the spatially common sparsity of massive MIMO channels during multiple time blocks, a closed-loop channel tracking scheme was also proposed to track the channels [58].

In large-scale distributed MIMO, both the small-scale fading coefficients (SSFCs) and the large-scale fading coefficients (LSFCs) play important roles. Usually, the estimation of LSFCs has been largely neglected, assuming somehow perfectly known prior to SSFCs estimation. In [59], a channel estimation algorithm was proposed to obtained the LSFCs by taking advantage of the spatial samples and the channel hardening effect.

3.3 Reciprocity calibration for TDD systems

In the practical wireless communication systems, the baseband channel matrix is composed of the wireless propagation channel and the coefficients of transceiver RF circuits. For TDD systems, although the wireless propagation channels are reciprocal, the RF circuits usually include antennas, mixers, filters, analog to digital converters, power amplifiers, etc. which are usually not symmetric for transmitter and receiver. This non-symmetric of the transceiver RF (also called as RF mismatches) makes the whole channel matrix non-reciprocal.

Considering the RF mismatches, the uplink and downlink channels are modeled by

$$\boldsymbol{G}_{\mathrm{UL}} = \boldsymbol{C}_{\mathrm{BS,r}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{C}_{\mathrm{UE,t}}, \tag{24}$$

$$\boldsymbol{G}_{\mathrm{DL}} = \boldsymbol{C}_{\mathrm{UE,r}} \boldsymbol{H} \boldsymbol{C}_{\mathrm{BS,t}}, \tag{25}$$

where $\boldsymbol{H}$ represents the wireless channel matrix, $\boldsymbol{C}_{\mathrm{BS,t}}$ and $\boldsymbol{C}_{\mathrm{BS,r}}$ denote the transmit and receive RF matrices of the BS, respectively, and $\boldsymbol{C}_{\mathrm{UE,t}}$ and $\boldsymbol{C}_{\mathrm{UE,r}}$ denote the transmit and receive RF matrices of the user equipment (UE), respectively. All of these RF matrices are modeled as a diagonal matrix. Each diagonal can be modeled as a random variable with log-normal distributed amplitude and uniform distributed phase.

Using the transpose of the uplink channel as the downlink channel, the ZF precoding matrix is written as

$$\boldsymbol{W} = \boldsymbol{G}_{\mathrm{UL}}^* \big(\boldsymbol{G}_{\mathrm{UL}}^{\mathrm{T}} \boldsymbol{G}_{\mathrm{UL}}^* \big)^{-1}. \tag{26}$$

It can be seen that $\boldsymbol{G}_{\mathrm{DL}} \boldsymbol{W}$ is not a diagonal matrix and there exists IUI. If we use the following precoding matrix

$$\boldsymbol{W} = \boldsymbol{C}_{\mathrm{BS,t}}^{-1} \boldsymbol{C}_{\mathrm{BS,r}} \boldsymbol{G}_{\mathrm{UL}}^* \big(\boldsymbol{G}_{\mathrm{UL}}^{\mathrm{T}} \boldsymbol{G}_{\mathrm{UL}}^* \big)^{-1}, \tag{27}$$

$\boldsymbol{G}_{\mathrm{DL}} \boldsymbol{W}$ becomes a diagonal matrix which means the IUI is avoided. Therefore, $\boldsymbol{C}_{\mathrm{BS,t}}^{-1} \boldsymbol{C}_{\mathrm{BS,r}}$ is called as the calibration matrix.

The impact of the RF mismatches on the massive MIMO systems was analyzed in [60–62]. Considering the RF mismatches, the lower bound of the ergodic sum-rate for multi-user massive MIMO can be given by [62]

$$\mathcal{E}\left(R^{\mathrm{mis}}\right) \geqslant R_{\mathrm{LB}}^{\mathrm{perfect}} - \Delta R_{\mathrm{BS}}^{\mathrm{mis}} - \Delta R_{\mathrm{UE}}^{\mathrm{mis}},$$

where

$$R_{\mathrm{LB}}^{\mathrm{perfect}} = K \cdot \left[\log\left(\rho\right) + \log\left(\frac{M-K}{K}\right) \right],$$

$$\Delta R_{\mathrm{BS}}^{\mathrm{mis}} = K \cdot \left\{ \log\left(\rho\right) + \log\left(\frac{\lambda_1}{\lambda_2}\right) + \log\left[\frac{MK}{(M-K)(K-1)}\right] \right\},$$

$$\lambda_1 = \mathrm{sinc}^2\left(\theta_{\mathrm{BS,r}}\right) \cdot \mathrm{sinc}^2\left(\theta_{\mathrm{BS,t}}\right),$$

$$\lambda_2 = \mathrm{e}^{2\delta_{\mathrm{BS,t}}^2} + \mathrm{e}^{2\delta_{\mathrm{BS,r}}^2} - 2\mathrm{e}^{\delta_{\mathrm{BS,t}}^2/2 + \delta_{\mathrm{BS,r}}^2/2} \cdot \mathrm{sinc}\left(\theta_{\mathrm{BS,r}}\right) \cdot \mathrm{sinc}\left(\theta_{\mathrm{BS,t}}\right),$$

$$\Delta R_{\mathrm{UE}}^{\mathrm{mis}} = K \left(\log \mathrm{e} \cdot 2\delta_{\mathrm{UE,t}}^2 \right).$$

ρ denotes the signal-to-noise ratio (SNR), δ^2 denotes the variance of the amplitude mismatch, and θ denotes the range of the phase mismatch for corresponding transceiver (the subscripts 't' and 'r' denote transmitter and receiver, respectively).

From Figure 3 it can be seen that the RF phase mismatch of BS will introduce more than 40% performance degradation for $\theta_{\mathrm{BS,t}} = \theta_{\mathrm{BS,r}} = \pi/3$. However, the phase mismatch of UE does not degrade
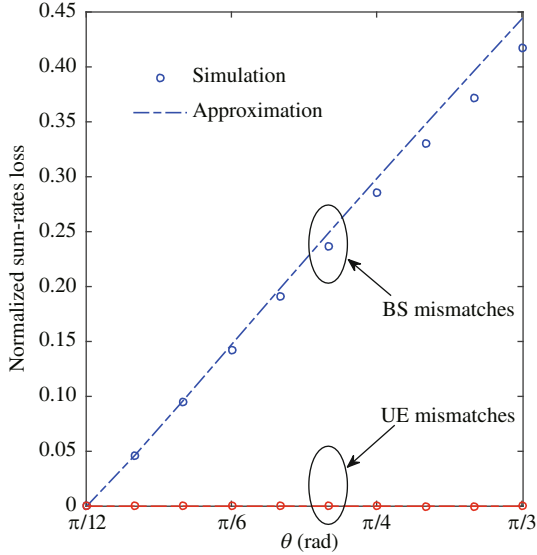
Figure 3 (Color online) Normalized sum-rates loss versus the phase range of RF mismatches. The transmit SNR is set to be 10 dB.
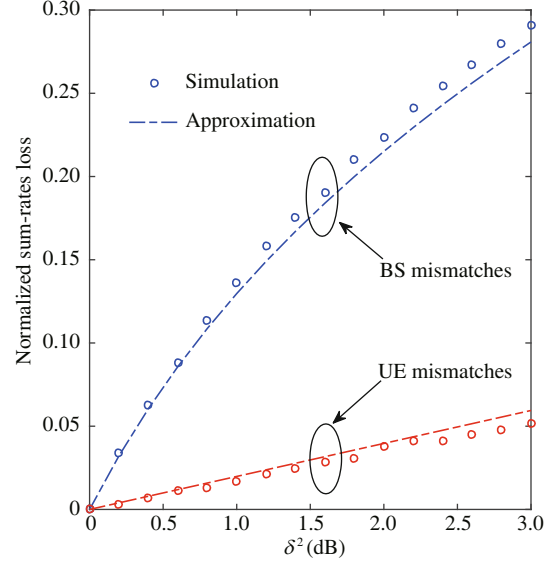

Figure 4 (Color online) Normalized sum-rates loss versus the amplitude variance of RF mismatches. The transmit SNR is set to be 10 dB.

the system throughput. The simulation results are consistent with the theoretical analysis. From Figure 4 it can be seen that for the BS, the amplitude mismatch also introduces severe performance degradation. For example, the performance loss is about 28% for $\delta_{\mathrm{BS,t}}^2 = \delta_{\mathrm{BS,r}}^2 = 3$ dB. While for the UE, the performance decreases slightly if the variance of the amplitude mismatch is small.

Therefore, it is very important to perform the reciprocity calibration for the implementation of massive MIMO systems and large-scale distributed MIMO systems, especially for the BS. There are two types of calibration methods, one is called hardware calibration method, and another is named signal-space calibration method.

The idea of the hardware calibration method is to use the couplers and multiple switches to connect the transmit circuits of each antenna with the receive circuits of the rest of the antennas [63]. Then, the self-transmitted calibration signals are used to perform the RF chains measurement. The hardware calibration method is preferred for massive MIMO systems due to the advantages of fastness and simplicity. However, this method requires a huge hardware cost, which is a heavy overhead for massive MIMO systems. Besides, this method may not be used for large-scale distributed MIMO systems, since the antennas of different RRUs are geographically separated. To avoid extra hardware circuits, the signal-space calibration method was proposed, where the calibration coefficients are calculated by the calibration signals. Then, this method is also named over-the-air (OTA) calibration method.

OTA calibration can be classified as partial calibration and full calibration. The partial calibration was proposed to just calibrate the RF mismatches at the BS, for example the calibration matrix in (27). While, the full calibration compensates for the RF mismatches at both the BS and the UEs. For full calibration method, we need to exchange the calibration signals between BS and UEs. Then, the BS computes the calibration coefficients according to the received uplink calibration signals and the downlink calibration signals feedback from the UEs. In [64], total least squares (TLS) based calibration method was proposed to achieve the optimal performance. But, in practice, the UEs are desired to be excluded from the calibration procedure, and the feedback will become very heavy with the increment of the antennas in massive MIMO systems.

Actually, we only need to perform calibration at the BS, since the RF mismatches of the UEs have a negligible impact on the system performance [60]. In [65], Argos method was proposed to perform calibration for an experimental massive MIMO system. It relies on the sequential transmission and reception of pilots between the reference antenna and the rest of the antennas. For distributed MIMO

systems, the calibration coefficients are obtained by exchanging calibration signals between the reference RRU and the other RRUs [66]. Nevertheless, the system performance will degrade significantly, when the SNR between the reference antenna and the other antennas is very small [65, 66]. Thus, to improve the calibration performance, LS calibration was presented in [67]. Not only relying on the reference antenna, the LS method utilizes the received calibration signals of all antennas to obtain the better performance. Theoretically, in [68] the LS method has been proved to be equivalent to the TLS method, since the LS method can be regarded as the extension of the TLS method [64] to the case of self-calibration. To avoid the eigenvalue decomposition required by the LS method, an iterative coordinate decent method was proposed to achieve the performance of the LS method yet with low computation complexity [68].

3.4 Topics for future research

At present, CSI acquisition is regarded as the key technology to implement LSAS whether in academia or industry. Exploiting the channel characteristic is often the preferable way to improve the performance of channel estimation. The parametric modeling and the sparse nature of LSAS are beneficial for channel parameters estimation and the pilot overhead reduction. For large-scale distributed MIMO systems, frequency and time synchronization among RRUs are basic prerequisites for achieving the joint processing gain [69]. In addition, calibration methods should be further studied. Although some calibration methods, such as TLS and LS, could achieve good performance in TDD systems, the computational complexity is still very high. For FDD systems, it is always a hot topic to reduce the feedback overhead for MU-MIMO systems. The reciprocity of statistical CSI (e.g., the correlation matrix of antennas) in FDD systems has been looked as an efficient way to reduce the feedback overhead. However, the necessity of calibration for the statistical CSI should be validated by experimental systems.

4 Transmission methods of multi-user LSAS

As known, for uplink multi-user MIMO systems, the capacity-approaching receiver is maximum a posteriori detection. For downlink transmission, the capacity can be achieved by dirty-paper coding with the downlink channel information [70]. However, with the increasing number of antennas and users, it is difficult to implement optimal transmission. The theoretical results of [5] and [21] demonstrated that for both massive MIMO and large-scale distributed MIMO systems, when the number of antennas tends to infinity, the channel capacity can be approached by low-complexity MRT in the downlink and MRC in the uplink. Furthermore, the results also showed that the channel capacity can be achieved by RZF precoding or linnear MMSE detection with not-so-large number of antennas [19]. That is, for LSAS, the capacity-approaching transceiver is more practical.

Nevertheless, due to the capabilities of the current technologies, the scale of antennas cannot be very large. When RZF precoding or linnear MMSE detection is employed, the computational complexity of the matrix inversion is still very large for a large number of users. Currently, there are two approaches to overcome the obstacle. One is to find a low-complexity matrix inversion method. Another way is to reduce the complexity of the transceiver by making use of the sparsity or the statistical property of the channel. In fact, due to the duality of the uplink and downlink, the mechanisms of both uplink and downlink transmission methods are very similar. For instance, in [71, 72], a polynomial expansion based precoding and a polynomial expansion based iterative receiver were proposed for massive MIMO to simplify matrix inversion, respectively. In order to exploit the sparsity of LSAS, belief propagation (BP) was extended to downlink precoding [73] and iterative detection [14].

In the rest of this section, we will firstly introduce the space division multiple access (SDMA) transmission methods, and then present a joint transmission method for large-scale DAS. Finally, we focus on capacity-approaching receiver for LSAS.

4.1 SDMA based on statistical CSI

For the multi-user downlink transmission, in order to achieve the capacity of multi-user broadcast channel, it is usually assumed that downlink CSI is known at the transmitter. Therefore, under this condition, efficient information transmission relies on the design of multi-user signals at the transmitter. However, different from tranditional MU-MIMO, there are two main difficulties to be solved when we design the downlink transmitter of multi-user massive MIMO. As we have discussed, first one is the complexity of CSI acquisition, and second on is the complexity of the spatial division for large-scale users. In order to solve both of the problems, multi-user transmission methods with statistical CSI have been proposed, including beam division multiple access (BDMA) [74] and joint spatial division and multiplexing (JSDM) [75].

The main idea of JSDM is a two-stage precoding scheme by using statistical CSI. First of all, the users in service are divided into groups, and each group of users have similar transmit correlation matrix. Then, the first-stage precoding is applied to each group of users semi-statically. Subsequently, the second-stage precoding is applied to the equivalent channel with the reduced dimension. It can be seen that the performance of the systems is largely dependent on user grouping.

For wideband massive MIMO systems, a beam domain channel model was proposed for theoretical analysis [74]. Based on the channel model, an upper bound of achievable ergodic sum-rate was derived. Ref. [74] showed that SDMA in the beam domain is optimal to maximize the upper bound of the ergodic sum-rate. That means by selecting users within non-overlapping beams, the MU-MIMO channels can be equivalently decomposed into multiple single-user MIMO channels. Then, a BDMA transmission scheme was proposed where multiple users are served simultaneously via different beams. With BDMA, both the complexity of transceiver and the overhead of channel estimation can be reduced significantly.

4.2 Downlink transmission in large-scale DAS

In large-scale DAS, the sparsity in the power domain can be utilized to reduce the complexity of RZF precoding. In [73], RZF beamforming (RZFBF) was realized in a distributed manner by using BP, so that it can avoid large dimensional matrix inversion. Using Bayesian inference and Gaussian approximation, BP-RZFBF was further developed. By exploiting the channel hardening property in large-scale MIMO systems, the AMP-RZFBF based on spatial channel covariance information (CCoI), called CCoI-aided AMP-RZFBF, was also proposed, which results in much simpler implementations in terms of computation.

In TDD large-scale distributed MIMO-OFDM systems, a low-complexity transmission method was presented by a joint design of uplink and downlink. The idea is shown in Figure 5. Firstly, we obtain the wideband statistical CSI matrix by uplink channel estimation. After TDD calibration and sparsification, we compute the precoding matrix by using the large-scale sparse matrix inversion. Then, we obtain the IUI suppression matrix. Finally, we perform uplink precoding and downlink precoding (called short-term single-user precoding) for every served user, according to the equivalent channel with the reduced dimension, i.e., the composite channel consisting of each user's practical channel and the interference suppression matrix. During the uplink transmission, each user sends its precoded signals, and at the BS, the received signals are passed to the interference suppression matrix, and then to the single-user detection. For the downlink, single-user precoding is performed for each user before IUI suppression precoding. Notice that the method could reduce the system complexity greatly, since we only compute the inverse of large dimensional statistical CSI matrix by using its sparsity.

4.3 Low-complexity receivers

The capacity-approaching receiver technology has always attracted much attention in multi-antenna systems. For MIMO receiver, the trade-off between computational complexity and performance is also a hot-topic. In particular, since LSAS is interference-limited, it is a quite challenging task to design a receiver which achieves its capacity under the interference channels.

For LSASs, the following difficulties are encountered in the design of a receiver, which covers imperfect CSI, inter-cell interference and high-complexity of the multi-user detection. In [76] a joint channel
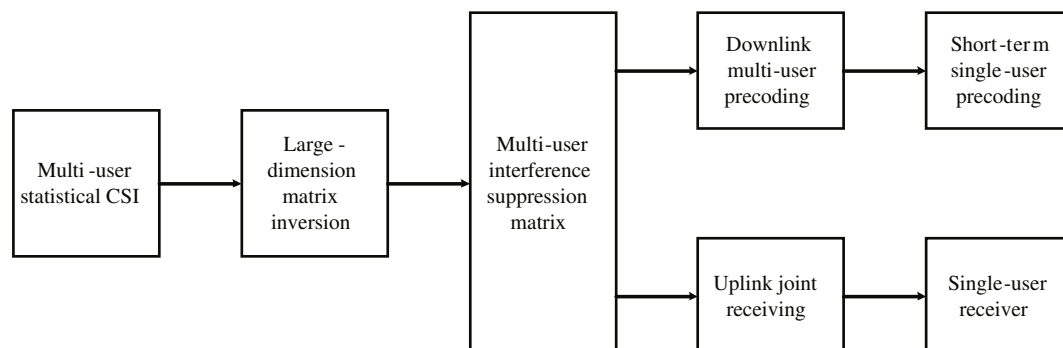
Figure 5 Transmitter and receiver design for large-scale distributed MIMO.

estimation and data detection algorithm with message passing was proposed by exploiting the channel hardening effect of massive MIMO. To overcome the complexity of the detection, recently, some researchers have studied large dimensional MIMO detection including simplified matrix inversion [72, 77] and sparsity-based detection [14, 78].

In [72], a turbo receiver has been proposed for massive MIMO. The iterative receiver consists of channel estimation, interference estimation, low-complexity soft-input soft-output (SISO) detection and decoding. Firstly, the channel parameters and preliminary statistical interference characteristics are estimated, and then linear prefiltering is peformed. Next, a low complexity SISO detector based on singular value decomposition is applied to the reduced-dimension channel matrix. To further reduce the detection complexity, a SISO detector based on polynomial expansion is presented to avoid matrix inversion. The results showed that due to the effect of large-scale antennas, the performance of the proposed turbo receiver is very close to the theoretical channel capacity performance.

4.4 Topics for future research

There are still many challenges on the wireless transmission of LSAS. Firstly, for high mobile speed users, it remains unknown that whether statistical CSI can be used to reduce the complexity of the implementation. The results in [79] indicated that large-scale MIMO has the ability to enhance the transmission rate of the high-speed users. However, the effectiveness of statistical beamforming and channel estimation of LSAS for high-mobility users should be further studied. Secondly, the wireless transmission methods of LSAS should be further studied under imperfect CSI, including channel estimation error and hardware impairment, such as I/Q imbalance [80], reciprocity mismatch [81], and a low-resolution analog-to-digital converter [82, 83]. Finally, the receiver of the BS has been paid much attention. However, for downlink transmission, with BDMA or other SDMA using statistical CSI, due to the residual interference, the user equipment is usually confronted with the interference channel. How to improve the performance of the receiver of the UE is also a challenge.

5 Resource allocation in LSAS

In LSAS, in order to further improve the efficiency of radio resource in the network, it is necessary to jointly optimize the allocation of space-time resources, power resources and user groups, etc., which brings new challenges to radio resource management. Hence, a low-complexity and high-performance resource allocation strategy is essential for the implementation of LSAS. Many researchers have devoted to the time-frequency resource allocation for MIMO-OFDM systems [84–86]. Nevertheless, this paper mainly focuses on the spatial resource allocation strategies and the RRUs grouping strategies in LSAS. In the following, we will discuss the spatial resource allocation strategies for massive MIMO and large-scale distributed MIMO, respectively.

The SDMA with statistical CSI can reduce the implementation complexity in massive MIMO systems. However, its performance is largely dependent on user grouping. A greedy algorithm based on statistical

CSI has been applied to do user grouping by maximizing the sum capacity [74]. As mentioned above, those users who have similar eigenvectors of their channel covariance matrices should first be assigned into a same cluster for JSDM. Thus, the performance of the grouping method depends on IUI. A K-Means clustering method using the chordal distance as its grouping metric was proposed in [87]. In [88], an improved grouping method using the weighted likelihood similarity measure and subspace projection based similarity measure has been proposed to provide a better system throughput.

In large-scale distributed MIMO systems, the clustering of RRUs and users is regarded as the key technology to reduce the system complexity. The clustering algorithms can roughly be classified into three kinds: the first one is the user-centric clustering, the second one is the dynamic clustering and the last one is the quasi-static interlaced clustering.

The user-centric wireless communication was first proposed in the development of 4G [89], and it is still a hot topic for 5G mobile communication at present. A large-scale DAS where each user chooses a few surrounding RRUs as its virtual cell was analyzed and the optimal size of the virtual cell was given in [90]. Under the capacity constraint of the inter-RRU backhaul, a method combining the user-centric dynamic clustering, the user scheduling and beamforming was proposed by maximizing the weighted sum rate with the generalized weighted mean square error [91].

Statistical CSI can also be used in the dynamic clustering [92]. In [93], using the sparsity of the channels, the channel matrix was transformed into a bilateral block diagonal matrix, with which the dynamic clustering and the parallel baseband processing were realized.

A quasi-static overlapping for BS clustering was presented in [94]. All the RRUs in the system are divided into disjoint clusters, and different clustering methods lead to different cluster patterns between which there exists an overlap. In each cluster pattern, all the users and RRUs employ the same frequency resources, while orthogonal time-frequency resources are allocated in different patterns. Combined with a power allocation method, the clustering method can significantly improve the throughput of the whole cell as well as its boundary.

There has always been much interest in resource allocation in wireless communication. Moreover, it is of great value to design resource allocation and multi-user scheduling methods with lower complexity in LSAS.

6 Conclusion

In this paper, the state-of-the-art research progress on the wireless transmission theory and technology in LSAS has been summarized, which covers the theoretical analysis of the spectral efficiency, CSI acquisition, uplink and downlink wireless communication, and resource allocation. Besides, this paper also presented potential research topics. Although there have already been some experimental large-scale MIMO systems, and the 3GPP standard organization is also promoting the evolution of relevant technology, a large number of theoretical and engineering problems should be addressed.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Ericsson. Ericsson mobility report. http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf. 2015

2 You X H, Pan Z W, Gao X Q, et al. The 5G mobile communication: the development trends and its emerging key techniques (in Chinese). Sci Sin Inform, 2014, 44: 551–563

3 Ma Z, Zhang Z Q, Ding Z G, et al. Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives. Sci China Inf Sci, 2015, 58: 041301

4 Paulraj A J, Gore D A, Nabar R U, et al. An overview of MIMO communications-a key to gigabit wireless. Proc IEEE, 2004, 2: 198–218

5 Marzetta T L. Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun, 2010, 11: 3590–3600

6 You X H, Wang D M, Sheng B, et al. Cooperative distributed antenna systems for mobile communications. IEEE Wirel Commun, 2010, 17: 35–43

7 Zhu H. Performance comparison between distributed antenna and microcellular systems. IEEE J Sel Area Commun, 2011, 29: 1151–1163

8 Wang J, Zhu H, Gomes N. Distributed antenna systems for mobile communications in high speed trains. IEEE J Sel Area Commun, 2012, 30: 675–683

9 Osman H, Zhu H, Toumpakaris D, et al. Achievable rate evaluation of in-building distributed antenna systems. IEEE Trans Wirel Commun, 2013, 12: 3510–3521

10 Dai L. A comparative study on uplink sum capacity with co-located and distributed antennas. IEEE J Sel Area Commun, 2011, 29: 1200–1213

11 Wang J, Dai L. Asymptotic rate analysis of downlink multi-user systems with co-located and distributed antennas. IEEE Trans Wirel Commun, 2015, 14: 3046–3058

12 Wang D, Wang J, You X, et al. Spectral efficiency of distributed MIMO systems. IEEE J Sel Area Commun, 2013, 10: 2112–2127

13 Huh H, Caire G, Papadopoulos H C, et al. Achieving massive MIMO spectral efficiency with a not-so-large number of antennas. IEEE Trans Wirel Commun, 2012, 9: 3226–3239

14 Wang D M, Zhao Z L, Huang Y Q, et al. Large-scale multi-user distributed antenna system for 5G wireless communications. In: Proceedings of IEEE 81st Vehicular Technology Conference Spring, Glasgow, 2015. 1–5

15 Tulino A M, Verdu S. Random matrix theory and wireless communications. In: Foundations and Trends in Communications and Information Theory. Norwell: Now Publishers Inc, 2004

16 Lu A, Gao X Q, Xiao C S. A free deterministic equivalent for the capacity of MIMO MAC with distributed antenna sets. In: Proceedings of IEEE International Conference on Communications, London, 2015. 1751–1756

17 Zhang J, Wen C K, Jin S, et al. On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas. IEEE J Sel Area Commun, 2013, 2: 133–148

18 Ngo H Q, Larsson E G, Marzetta T L. Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans Commun, 2013, 4: 1436–1449

19 Hoydis J, Brinkz S, Debbah M. Massive MIMO in the UL/DL of cellular networks: how many antennas do we need. IEEE J Sel Area Commun, 2013, 2: 160–171

20 Wang D M, Ji C, Gao X Q, et al. Uplink sum-rate analysis of multi-cell multi-user massive MIMO system. In: Proceedings of IEEE International Conference on Communications, Budapest, 2013. 5404–5408

21 Wang D M, Ji C, Sun S H, et al. Spectral efficiency of multicell multi-user DAS with pilot contamination. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, 2013. 3208–3212

22 Li J M, Wang D M, Zhu P C, et al. Downlink spectral efficiency of multi-cell multi-user large-scale DAS with pilot contamination. In: Proceedings of IEEE International Conference on Communications, London, 2015. 2011–2016

23 Andrews J, Baccelli F, Ganti R. A tractable approach to coverage and rate in cellular networks. IEEE Trans Commun, 2011, 11: 3122–3134

24 Baccelli F, Giovanidis A. A stochastic geometry framework for analyzing pairwise-cooperative cellular networks. IEEE Trans Wirel Commun, 2015, 2: 794–808

25 Fei Z S, Ding H C, Xing C W, et al. Performance analysis for range expansion in heterogeneous networks. Sci China Inf Sci, 2014, 57: 082305

26 Lin Y, Yu W. Ergodic capacity analysis of downlink distributed antenna systems using stochastic geometry. In: Proceedings of IEEE International Conference on Communications, Budapest, 2013. 3338–3343

27 Bai T Y, Heath R W. Analyzing uplink SIR and rate in massive MIMO systems using stochastic geometry. arXiv:1510.02538. 2015

28 Wang D, You X, Wang J, et al. Spectral efficiency of distributed MIMO cellular systems in a composite fading channel. In: Proceedings of IEEE International Conference on Communications, Beijing, 2008. 1259–1264

29 Yang A, Jing Y, Xing C, et al. Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas. IEEE Trans Wirel Commun, 2015, 10: 5659–5671

30 Aggarwal R, Koksal C E, Schniter P. On the design of large scale wireless systems. IEEE J Sel Area Commun, 2013, 2: 215–225

31 Xin Y, Wang D, Li J. Area spectral efficiency and area energy efficiency of massive MIMO cellular systems. IEEE Trans Veh Tech, in press. doi: 10.1109/TVT.2015.2436896

32 Bjornson E, Hoydis J, Kountouris M, et al. Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits. IEEE Trans Inf Theory, 2015, 11: 7112–7139

33 Gustavsson U, Sanchez-Perez C, Eriksson T, et al. On the impact of hardware impairments on massive MIMO. In: Globecom Workshops (GC Wkshps), Austin, 2014. 294–300

34 Fernandes F, Ashikhmin A, Marzetta T L. Inter-cell interference in non-cooperative TDD large scale antenna systems. IEEE J Sel Area Commun, 2013, 2: 192–201

35 Zhang H, Zheng X, Xu W, et al. On massive MIMO performance with semi-orthogonal pilot-assisted channel estimation. EURASIP J Wirel Commun Netw, in press. doi: 10.1186/1687-1499-2014-220

36 Jin S, Li M M, Huang Y M, et al. Pilot scheduling schemes for multi-cell massive multiple-input-multiple-output transmission. IET Commun, 2015, 9: 689–700

37 You L, Gao X, Xia X G, et al. Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels. IEEE Trans Wirel Commun, 2015, 6: 3352–3366

38 Yin H, Gesbert D, Filippou M, et al. A coordinated approach to channel estimation in large-scale multiple-antenna systems. IEEE J Sel Area Commun, 2013, 2: 264–273

39 Chen Z, Yang C. Pilot decontamination in massive MIMO systems: exploiting channel sparsity with pilot assignment. In: Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, 2014. 637–641

40 Gao Z, Dai L, Wang Z. Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems. Electron Lett, 2014, 12: 896–898

41 Yang Y, Bai B, Chen W. How much spectrum can be reused in 5G cellular networks a matrix graph approach. arXiv: 1401.4750. 2014

42 Atzeni I, Arnau J, Debbah M. Fractional pilot reuse in massive MIMO systems. arXiv:1503.07321. 2015

43 Choi J, Chance Z, Love D J, et al. Noncoherent trellis coded quantization: a practical limited feedback technique for massive MIMO systems. IEEE Trans Commun, 2013, 12: 5016–5029

44 Noh S, Zoltowski M D, Sung Y, et al. Pilot beam pattern design for channel estimation. IEEE J Sel Topics Signal Process, 2014, 5: 787–801

45 Choi J, Love D, Bidigare P. Downlink training techniques for FDD massive MIMO systems: open-loop and closed-loop training with memory. IEEE J Sel Topics Signal Process, 2014, 8: 802–814

46 You L, Gao X, Swindlehurst A L, et al. Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots. IEEE Trans Signal Process, 2015, 6: 1461–1476

47 Adeogun R O. Channel prediction for mobile MIMO wireless communication systems. Dissertation for Ph.D. Degree. Wellington: Victoria University of Wellington, 2015. 1–313

48 Huang M, Chen X, Zhou S, et al. Low-complexity subspace tracking based channel estimation method for OFDM systems in time-varying channels. In: Proceedings of IEEE International Conference on Communications (ICC), Istanbul, 2006. 4618–4623

49 Simeone O, Bar-Ness Y, Spagnolini U. Pilot-based channel estimation for OFDM systems by tracking the delay-subspace. IEEE Trans Wirel Commun, 2004, 1: 315–325

50 Zhu Y, Liu L, Wang A, et al. DoA estimation and capacity analysis for 2D active massive MIMO systems. In: Proceedings of IEEE International Conference on Communications, Budapest, 2013. 4630–4634

51 Qi C, Huang Y, Jin S, et al. Sparse channel estimation based on compressed sensing for massive MIMO systems. In: Proceedings of IEEE International Conference on Communications (ICC), London, 2015. 4558–4563

52 Masood M, Afify L H, Al-Naffouri T Y. Efficient coordinated recovery of sparse channels in massive MIMO. IEEE Trans Signal Process, 2015, 1: 104–118

53 Masood M, Al-Naffouri T Y. Sparse reconstruction using distribution agnostic Bayesian matching pursuit. IEEE Trans Signal Process, 2013, 21: 5298–5309

54 Ngo B Q, Larsson E G. EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays. In: Proceedings of IEEE International Conference on Acoust, Speech, Signal Processing, Kyoto, 2012. 3249–3252

55 Muller R F, Cottatellucci L, Vehkaper M. Blind pilot decontamination. IEEE J Sel Topics Signal Process, 2014, 5: 773–786

56 Ma J, Li P. Data-aided channel estimation in large antenna systems. IEEE Trans Signal Process, 2014, 12: 3111–3124

57 Rao X, Lau V K N. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. IEEE Trans Signal Process, 2014, 12: 3261–3271

58 Gao Z, Dai L, Wang Z, et al. Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO. IEEE Trans Signal Process, 2015, 23: 6169–6183

59 Chen K F, Liu Y C, Su Y T. On composite channel estimation in wireless massive MIMO systems. In: Proceedings of IEEE Globecom Workshops, Atlanta, 2013. 135–139

60 Wei H, Wang D M, Zhu H L, et al. Mutual coupling calibration for multiuser massive MIMO systems. IEEE Trans Wirel Commun, 2016, 15: 606–619

61 Zhang W, Ren H, Pan C, et al. Large-scale antenna systems with UL/DL hardware mismatch: achievable rates analysis and calibration. IEEE Trans Commun, 2015, 4: 1216–1229

62 Wei H, Wang D M, Wang J Z, et al. Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding. Sci China Inf Sci, 2016, 59: 022302

63 Nishimori K, Hiraguri T, Ogawa T, et al. Effectiveness of implicit beamforming using calibration technique in massive MIMO system. In: Proceedings of IEEE International Workshop on Electromagnetics (iWEM), Sapporo, 2014. 117–118

64 Kaltenberger F, Jiang H, Guillaud M. Relative channel reciprocity calibration in MIMO/TDD systems. In: Proceedings of IEEE Future Network and Mobile Summit, Florence, 2010. 1–10

65 Shepard C, Yu H, Anand N. Argos: practical many-antenna base stations. In: Proceedings of the 18th annual

International Conference on Mobile Computing and Networking, Istanbul, 2012. 53–64

66 Rogalin R, Bursalioglu O Y, Papadopoulos H C. Hardware-impairment compensation for enabling distributed large-scale MIMO. In: Proceedings of IEEE Information Theory and Applications Workshop (ITA), San Diego, 2013. 1–10

67 Rogalin R, Bursalioglu O Y, Papadopoulos H, et al. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. IEEE Trans Wirel Commun, 2014, 13: 1815–1831

68 Wei H, Wang D M, Wang J Z, et al. TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent. Sci China Inf Sci, in press. doi: 10.1007/s11432-015-5441-4

69 Rahul H S, Kumar S, Katabi D. JMB: scaling wireless capacity with user demands. In: Proceedings of ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. New York: ACM, 2012. 235–246

70 Yu W. Competition and cooperation in multiuser communication environments. Dissertation for Ph.D. Degree. Stanford: Stanford University, 2002

71 Kammoun A, Muller A, Bjornson E, et al. Linear precoding based on polynomial expansion: large-scale multi-cell MIMO systems. IEEE J Sel Topics Signal Process, 2014, 8: 861–875

72 Huang Y, Tang W, Li J, et al. On the performance of iterative receivers in massive MIMO systems with pilot contamination. In: proceedings of IEEE 9th Conference on Industrial Electronics and Applications (ICIEA), Hangzhou, 2014. 52–57

73 Wen C K, Chen J C, Wong K K, et al. Message passing algorithm for distributed downlink regularized zero-forcing beamforming with cooperative base stations. IEEE Trans Wirel Commun, 2014, 13: 2920–2930

74 Sun C, Gao X, Jin S, et al. Beam division multiple access transmission for massive MIMO communications. IEEE Trans Commun, 2015, 6: 2170–2184

75 Nam J, Ahn J Y, Caire G. Joint spatial division and multiplexing – the large-scale array regime. IEEE Trans Inf Theory, 2013, 10: 6441–6463

76 Narasimhan T L, Chockalingam A. Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems. IEEE J Sel Topics Signal Process, 2014, 8: 847–860

77 Dai L L, Gao X Y, Su X, et al. Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multi-user large-scale MIMO systems. IEEE Trans Veh Tech, 2014, 64: 4839–4845

78 Fadlallah Y, Aissa A, Amis K, et al. New iterative detector of MIMO transmission using sparse decomposition. IEEE Trans Veh Tech, 2014, 64: 3458–3464

79 Cao J, Wang D, Li J, et al. Uplink sum-rate analysis of massive MIMO system with pilot contamination and CSI delay. Wirel Personal Commun, 2014, 1: 297–312

80 Zhang W, Lamare R C, Pan C, et al. Widely linear block diagonalization type precoding in massive MIMO systems with IQ imbalance. In: Proceedings of IEEE International Conference on Communications, London, 2015. 1789–1794

81 Han S, Yang C, Wang G, et al. Coordinated multipoint transmission strategies for TDD systems with non-ideal channel reciprocity. IEEE Trans Commun, 2013, 10: 4256–4270

82 Fan L, Jin S, Wen C K, et al. Uplink achievable rate for massive MIMO systems with low-resolution ADC. IEEE Commun Lett, 2015, 19: 2186–2189

83 Zhang T C, Wen C K, Jin S, et al. Mixed-ADC massive MIMO detectors: performance analysis and design optimization. arXiv:1509.07950. 2015

84 Zhu H, Wang J. Chunk-based resource allocation in OFDMA systems-part I: chunk allocation. IEEE Trans Commun, 2009, 9: 2734–2744

85 Zhu H, Wang J. Chunk-based resource allocation in OFDMA Systems-Part II: joint chunk, power and bit allocation. IEEE Trans Commun, 2012, 2: 499–509

86 Zhu H, Karachontzitis S, Toumpakaris D. Low-complexity resource allocation and its application to distributed antenna systems. IEEE Wirel Commun, 2010, 3: 44–50

87 Nam J Y, Adhikary A, Ahn J Y, et al. Joint spatial division and multiplexing: opportunistic beamforming, user grouping and simplified downlink scheduling. IEEE J Sel Topics Signal Process, 2014, 8: 876–890

88 Xu Y, Yue G, Mao S. User grouping for massive MIMO in FDD systems: new design methods and analysis. IEEE Access, 2014, 2: 947–959

89 Xu X D, Wu C L, Tao X F, et al. Maximum utility principle access control for beyond 3G mobile system. Wirel Commun Mobile Comput, 2007, 7: 951–959

90 Dai L. An uplink capacity analysis of the distributed antenna system (DAS): from cellular das to das with virtual cells. IEEE Trans Wirel Commun, 2014, 13: 2717–2731

91 Dai B B, Yu W. Sparse beamforming and user-centric clustering for downlink cloud radio access network. IEEE Access, 2014, 2: 1326–1339

92 Liu J, Wang D. An improved dynamic clustering algorithm for multi-user distributed antenna system. In: Proceedings of IEEE International Conference on Wireless Communications Signal Processing (WCSP), Nanjing, 2009. 1–5

93 Fan C, Zhang Y J, Yuan X. Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs. IEEE Trans Wirel Commun, 2016, 15: 1881–1894

94 Ratnam V V, Caire G, Molisch A F. Capacity analysis of interlaced clustering in a distributed antenna system. In: Proceedings of IEEE International Conference on Communications (ICC), London, 2015. 578–582