Optimal Antenna Cluster Size in Cell-Free Large-Scale Distributed-Antenna Systems with Imperfect CSI and Inter-Cluster Interference

Zhiyuan Jiang, *Student Member, IEEE*, Sheng Zhou, *Member, IEEE*, and Zhisheng Niu, *Fellow, IEEE*

***Abstract*—In a cell-free Large-Scale Distributed-Antenna Systems (L-DAS), the antennas are distributed over the intended coverage area. Introducing cooperation among the antennas can significantly improve the system throughput. Whereas only finite-cluster-size (limited) cooperation with L-DAS antenna-clusters is realistic due to practical limitations. In this paper, the impact of antenna cluster size on the downlink sum rate of frequency-division duplex (FDD) cell-free L-DAS is analyzed considering *imperfect channel state information* (CSI) and *inter-cluster interference* (ICLI). We investigate the optimal cluster size, in terms of maximizing the downlink sum rate with respect to the overhead of imperfect channel training and feedback. When the number of users is sufficiently large, closed-form lower bounds of the ergodic sum capacity, which are leveraged to analyze the system performance, are derived by constructing different rate-achieving user scheduling schemes. Both analog and digital feedback schemes are considered to evaluate the imperfect channel feedback. Based upon these, closed-form expressions of the optimal cluster size for one-dimensional antenna-topology systems and the corresponding achievable rates are derived. The scaling law of the optimal cluster size for two-dimensional antenna-topology systems is also given. Numerical results demonstrate the impact of signal-to-noise ratio (SNR) and the block length on the achievable rates and the optimal cluster size, which agree with our analytical results.**

***Index Terms*—Distributed large-scale antenna systems, Frequency-division-duplex, Broadcast channel, Imperfect CSI.**

I. Introduction

Multiuser multiple-input-multiple-output (MU-MIMO) technology enables simultaneous (on the same time-frequency recourse) data transmissions to a multiplicity of autonomous terminals via distinguishable spatial data streams. To reach its true potential, the number of base station (BS) antennas in MU-MIMO systems is scaled up to hundreds or even thousands, which is often dubbed the massive MIMO system, further improving the system capacity and radiated energy efficiency with simple linear precoding and decoding schemes [1]. The multiple BS antennas can be co-located at the BS side, which is referred to as co-located antenna systems (CAS). Alternatively, they can be distributed throughout the intended coverage area without the traditional concepts of "cells" [2], or referred to as distributed antenna systems (DAS). Recent literature has shown that in addition to capacity improvement, the distributed antennas can enhance the system coverage, energy efficiency and battery life of user-terminals, due to the reduced pathloss from BS antennas to user-terminals [3][4].

Ideally, the numerous distributed antennas should all cooperate with each other, performing joint precoding and decoding, to fully exploit the cooperation gain. Whereas in practice, due to the large amount of overheads for channel state information (CSI) acquisition and user data sharing, only limited cooperation, i.e., a finite number of antennas forming a cluster, is realistic (see e.g. [5][6][7])). Antennas in each cluster are connected via a high-capacity backhaul, where the number of cooperating antennas is referred to as the cluster size. Signals from transmit antennas of other clusters are handled as inter-cluster interference (ICLI).

In principle, there exists a tradeoff between the *cooperation gain* and the *overhead*. Specifically, the cooperation gain comes from the fact that the cooperating antennas perform joint signal processing within each cluster, and the ICLI is better suppressed when the cluster size is larger, in which case most users are farther away from the interfering antennas. On the other hand, the CSI at transmitter (CSIT) acquisition overhead scales with the cluster size[1]. In practice, the CSIT is usually obtained via uplink channel training, utilizing the channel reciprocity, which is only feasible for time-division-duplex (TDD) mode. While in frequency-division-duplex (FDD) and uncalibrated TDD systems, the antennas have to first send downlink training sequences and then listen to the user CSI feedback. The TDD mode offers the more practical way to acquire timely and accurate CSI in massive MIMO systems, due to the fact that the CSI acquisition overhead with TDD mode scales with the number of users, irrespective of the number of antenna cluster size. Nonetheless, the FDD system still represents the far more majority of the currently deployed cellular systems. Unlike the large quantity of downlink cooperating antennas enabled by the TDD mode, the cluster size of the FDD system is limited by the CSIT acquisition overhead, as well as the imperfect CSIT brought by practical training

The authors are with Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China. Email: jiang-zy10@mails.tsinghua.edu.cn, {sheng.zhou,niuzhs}@tsinghua.edu.cn.

Part of the work has been presented at IEEE Globecom 2012. This work is sponsored in part by the National Basic Research Program of China (973 Program: 2012CB316001), the National Science Foundation of China (NSFC) under grant No. 61201191 and No. 61321061, and Hitachi R&D Headquarter.

[1]Generally, the cluster size is also affected by the backhaul capacity, computation overhead due to exchange of information among antennas. However these factors can be alleviated by the more and more efficient optical transmissions and processing chips. However the channel estimation overhead is a more inherited limitation.

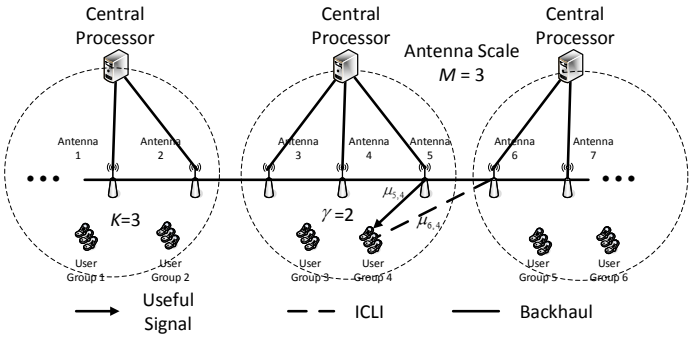


Fig. 1. Specifications of antenna-clusters and user-groups. The total number of users in each antenna-cluster is $N = KG = 6$ in this case. The dash circles represent the antenna clusters. Note that only two channel gains are depicted for brevity.

and feedback process.

In this paper, we consider the *optimal antenna cluster size* in a cell-free FDD large-scale distributed-antenna system (L-DAS). See Fig. 1 for example,[2] antennas and users are distributed in the coverage area. Our analysis aims to determine the optimal number of cooperating antennas to optimize the system throughput. Given the system setting, leveraging the Wyner model [8], which is a tractable mathematical abstraction of the multi-site cellular systems, we analyze the optimal cluster size in this scenario to maximize the sum ergodic rate of the system, considering the CSIT acquisition overhead, imperfect CSIT, and the impact of ICLI. The main contributions of this paper include:

- Assuming independent Rayleigh fading channels with arbitrary channel variance profile, closed-form lower bounds for the ergodic downlink BC sum capacity are derived explicitly, corresponding to different rate-achieving user scheduling schemes, with sufficiently large number of users.
- The optimal cluster size in the cell-free FDD L-DAS, in terms of maximizing the downlink sum achievable rate, regarding imperfect CSI, channel training and CSI feedback is investigated in depth. Both analog and digital feedback schemes are considered. Closed-form expressions of the optimal cluster size for one-dimensional antenna-topology systems and the corresponding achievable rate are derived for the first time to the best our knowledge. The scaling results of the optimal cluster size for two-dimensional systems are also given.

The rest of the paper is organized as follows. The related work is discussed in Section II. The system model and the closed-form expressions of the downlink sum rate are given in Section III. The optimal cluster size and system achievable rates considering imperfect CSI, channel training and feedback are analyzed in Section IV. The simulation results are given in Section V. Finally, we conclude our work in Section VI.

[2]In fact, we can cluster antennas which are associated with different central units, by connecting the central units with backhaul links. Thereby, possible dynamic antenna clustering schemes can be realized. Specifically, a clustering scheme decides which set of antennas should form a cluster and thus the central units perform the joint processing of the signals for those antennas, where the signals are shared via fiber links. It is noteworthy that in practice, the configurations of the links among antennas are quite flexible and diverse, while Fig. 1 only illustrates one possible configuration for brevity.

Notations : We use boldface uppercase letters, boldface lowercase letters and lowercase letters to designate matrices, column vectors and scalars respectively. $\boldsymbol{X}^\dagger$ are used to denote the complex conjugate transpose of matrix $\boldsymbol{X}$. $\boldsymbol{x}_i$ denotes the i-th element of vector $\boldsymbol{x}$. Notation $\det(\boldsymbol{X})$ and $\mathrm{Tr}(\boldsymbol{X})$ denote the determinant and the trace of matrix $\boldsymbol{X}$, respectively. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circularly symmetric complex Gaussian random vector of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathrm{diag}[\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n]$ denotes a block diagonal matrix with $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ on its diagonal. The logarithm $\log(x)$ denotes the binary logarithm.

II. Related Work

For the existing literature, extensive work has been done regarding the capacity region of general MU-MIMO downlink broadcast channels (BC) (see e.g. [9–12]). It is found that the sum capacity of BC is the solution to a min-max problem in which the optimum can be achieved using Costa's "dirty paper coding" [13]. An important duality is found between downlink BC and uplink MAC [12], which reduces the computational complexity of the sum capacity. There have been further works computing the sum capacity of BC under concrete channel conditions (see e.g. [14–16]). Ref. [14] gives an explicit solution to the min-max problem of the BC sum capacity considering a specific channel matrix. Ref. [15][16] apply the random matrix theory to this problem, and obtain some large system analysis results [17]. In our previous work [18] and [19], closed-form expressions of the sum capacity of CAS and DAS BC are derived respectively, which are exploited to study the optimal cluster size in this paper.

The CSIT is obtained by feedback from users in an FD-D system. Many papers have considered limited feedback schemes (see e.g. [20][21] and references therein), which show that it is of great importance to study the impact of channel training and feedback overhead and imperfect channel estimation on the system achievable rate of downlink BC. To this end, the training capacity is studied in the literature recently [22]. The training capacity refers to the system capacity considering imperfect CSI and overhead caused by channel estimation. Ref. [23] considers the impact of imperfect CSI and feedback on the MU-MIMO system, both analog and digital feedback schemes are studied. Ref. [24] considers imperfect CSI training and feedback of the MU-MIMO system, where a capacity region inner bound is given regarding the worst case uncorrelated noise. The training capacity serves as the achievable rate in this paper, in order to analyze the optimal cluster size of MU-MIMO systems.

The performance advantages of DAS, in terms of capacity improvements, coverage extension, energy efficiency (EE) and battery life are shown in [3, 4, 25–28]. A comprehensive survey on this subject can be found in [26]. In [25], the coverage extension enabled by DAS is verified by conducting an experiment using the 802.11g signals. It is further shown that the additional delay spread introduced by the DAS is not detrimental under certain circumstances. The EE optimization

in DAS is considered in [27] and [28], where in [28] the authors consider the joint power control and beamforming problem in a single-user setting. In [27] the authors consider the multi-user scenario, where they first formulate the EE optimization problem, then propose a decomposition-based algorithm to solve it.

In practice, given the physical parameters, e.g. coherence time, coherence bandwidth, and the number of users, the system should be able to decide the optimal cluster size accordingly. If the cluster size is too large, the "feedback storm" will significantly reduce the system throughput [23]. Yet in the literature, there have been few studies concerning the cluster size of MU-MIMO systems. In [29] [30], the optimal cluster size of uplink channel in DAS is analyzed, but ICLI is not considered. Recent studies (see [31] and references therein) analyze the BS cooperation assuming random antennas locations, where the relationship between cluster size and outage performance are given.

III. System Model and Achievable Sum Rate

The downlink channel of an antenna-cluster is considered, where M distributed antennas serves N single-antenna users in the intended area.[3] One channel use is expressed as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} + \boldsymbol{s}, \tag{1}$$

wherein $\boldsymbol{x} \in \mathcal{C}^M, \boldsymbol{y} \in \mathcal{C}^N$ are the transmit and receive signals of the antennas and users, with downlink transmit power of the m-th antenna $P_m = \mathbb{E}[\boldsymbol{x}_m \boldsymbol{x}_m^\dagger]$. $\boldsymbol{n}$ is the thermal noise with variance σ_{n}^2, and $\boldsymbol{s}$ denotes the ICLI. The thermal noise and the ICLI are assumed to be independently Gaussian distributed[4]. Denote

$$\sigma_i^2 = \sigma_{\mathrm{n}}^2 + s_i, \tag{2}$$

as the effective noise power of user i, where $s_i = \mathbb{E}[\boldsymbol{s}_i \boldsymbol{s}_i^\dagger]$. Denote $\boldsymbol{H}_{i,j}$ as the channel coefficient from antenna i to user j, including pathloss and small-scale fading. We assume the entries of $\boldsymbol{H}$ are non-identical independently Rayleigh distributed (with non-identical variance for different channel coefficients). For the ease of exposition, the channel is normalized as

$$\boldsymbol{y} = \tilde{\boldsymbol{H}}\boldsymbol{x} + \boldsymbol{z}, \tag{3}$$

where $\tilde{\boldsymbol{H}} = \boldsymbol{G}^{-1}\boldsymbol{H}$, $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$, and $\boldsymbol{G} = \mathrm{diag}[\sigma_i]$. The instantaneous channel gain is denoted by

$$||\boldsymbol{H}_{i,j}||^2 = \boldsymbol{V}_{i,j}. \tag{4}$$

In this section, we assume the transmitter and receiver know the CSI perfectly, i.e., genie-aided. Thus in this section the channel coherence time is irrelevant to the ergodic capacity. Imperfect CSI and the block-fading model are considered in Section IV.

Remark 1: One possible extension of this work is to consider different clustering configurations in different clusters, which may bring extra performance gain, compared with our fixed-antenna-size clustering scheme.

Remark 2: Note that we treat the ICLI as Gaussian noise, which is not a good approximation, especially, in the orthogonal-frequency-division-multiple-access (OFDMA) system [33], where the ICLI suffered by a given tone signal is dominated by the interference from a small number of users using the same tone signal. However, in the L-DAS system, after the downlink precoding, the interfering signals are actually the superposition of the signals for all the users. Therefore, the *interference-average* effect leads to the fact that the ICLI can be well approximated by the Gaussian noise.

Moreover, although treating the ICLI as Guassian noise is sub-optimum in the sense that analyzing the structure of the ICLI may lead to better performance. However, there is some recent work, see [34] and references therein, showing that when the desired signal strength is reasonably large enough (the exact conditions can be found in [34]), treating interference as Gaussian noise actually achieves all points in the capacity region to within a constant gap.

A. Sum Rate Lower Bound of BC

The generic BC capacity with arbitrary channel matrix and per-antenna power constraints, $P_m \leq P_{\mathrm{dl}}$, $\forall m$, is the solution to a min-max problem which is given in [12]. To be concrete, we consider a scenario where the users are divided into G groups. The users in the same group have *identical* large-scale fading coefficients, and each group has K users, thus $N = KG$.[5] See Fig. 1 for an example of system setting. When K is sufficiently large, the downlink ergodic sum capacity is lower bounded by the following theorem.

Theorem 1: With users divided into G groups, each of which has $K \gg 1$ users, identical large-scale fading and i.i.d. small-scale fading coefficients, the downlink ergodic sum capacity of one antenna-cluster is lower bounded by

$$\bar{C}_{\mathrm{WB}} = \log\left(1 + P_{\mathrm{eq}}^M \prod_{i=1}^{M} c_i\right), \tag{5}$$

where $P_{\mathrm{eq}} = \frac{P_{\mathrm{dl}} M}{G}$, $c_i = \sum_{j=1}^{G} u_{i,j}^2/\sigma_j^2$, and $u_{i,j}$ denotes the channel gain magnitude from antenna i to the users in the j-th group.

Proof: See Appendix A. ∎

Remark 3: The system setting in Theorem 1 represents a heavy-loaded ($K \gg 1$) scenario, which is a typical scenario for the L-DAS. The lower bound in Theorem 1, which can be viewed as an achievable sum rate, is treated as the optimization objective in Section IV when we consider the optimal cluster size of the cell-free L-DAS. The *Wide-bandwidth* (WB)

[3]The large scale considered in this paper means that the total number of antennas in the intended area is large, i.e., M times the number of clusters. To illustrate, we are considering using a L-DAS network to cover an intended area, where letting all antennas cooperate with each other is impractical due to the CSIT acquisition overhead. Therefore, antenna clustering is considered, which divides the L-DAS network into clusters, each of which consists of M antennas.

[4]Analyzing the structure of the interference leads to the study of the interference channel [32], and the channel estimation error leads to the study of the specific CSI training and feedback schemes. Those are out of the scope of this paper

[5]Dividing the users into groups w.r.t. their large-fading coefficients is reasonable in the cellular system considering the users usually form hot spots. Similar techniques can be found in [36]

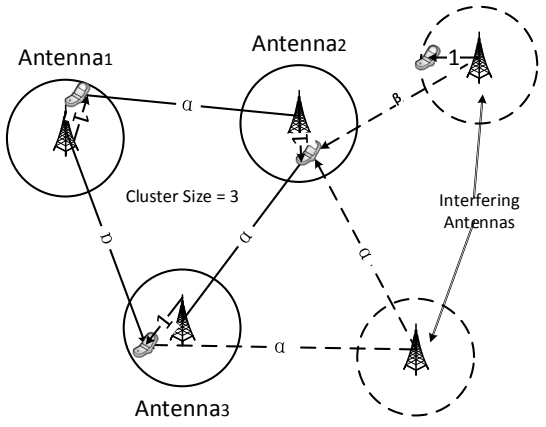


Fig. 2. A DAS example with 5 antennas. Three of them form a cluster, and the other two antennas are interfering antennas, wherein α and β are the respective channel gains.

scheme is assumed in Theorem 1, i.e., equal power is allocated to all users (26). However, in the presence of channel fading, we can always allocate more power to the users with better channel condition to achieve the *multi-user diversity gain*. Thus A *Group-TDMA* (GT) scheme is considered. The GT scheduling scheme uniformly divides the K users of each group into L sub-groups and transmits to the users with the largest channel gain in each group at each time instance. The case where K is not divided exactly by L is ignored as K is sufficiently large, thus the division remainder can be neglected. Hence, each sub-group has $\frac{K}{L}$ users. With the GT scheme, we have the following corollary.

Corollary 1: The downlink ergodic sum capacity of one antenna-cluster is lower bounded by

$$\bar{C}_{\mathrm{GT}} = \log\left(1 + P_{\mathrm{eq}}^{M}\prod_{i=1}^{M} c_i\right), \tag{6}$$

where $P_{\mathrm{eq}} = \frac{P_{\mathrm{dl}}M}{G}$, $c_i = \sum\limits_{j\neq i}^{G}\frac{u_{i,j}^2}{\sigma_j^2} + \frac{u_{i,i}^2}{\sigma_i^2}(1-\varepsilon)\log_e K$, and ε is a parameter satisfying $L = K^{\varepsilon}$, where $0 < \varepsilon < 1$.

Proof: See Appendix B. ∎

Remark 4: Qualitatively, since Corollary 1 selects the users with the largest channel gains to corresponding antennas in each sub-group, the so-called *multiuser-diversity gain* is achieved. Basically, it stems from the fact that it is more likely to have larger channel gains and better channel orthogonality, and in consequence, better rates, when there are more users assuming user-channels are independent. In fact, extensive work, see [35] and references therein for example, has been done showing that the optimum capacity scaling law is of the order $\mathcal{O}(\log\log(K))$, which is exactly what we have in Corollary 1.

Remark 5: Since the rate-achieving user-scheduling schemes are given in Theorem 1 and Corollary 1, they essentially serve as achievable rates of downlink BC, and all achievable rates are lower bounds of the capacity by definition. Notice that Theorem 1 and Corollary 1 are applicable to any multi-antenna topology as long as the large-scale fading coefficients are given. For example, consider a multi-antenna topology described in Fig. 2. The signal and ICLI strength are depicted in the figure. Applying Theorem 1, the sum capacity lower bound is

$$\bar{C}_{\mathrm{WB}} = \log(1 + P_{\mathrm{dl}}^{3}c_1c_2c_3), \tag{7}$$

where $c_1 = \alpha^2/(\sigma^2+\alpha^2P_{\mathrm{dl}}+\beta^2P_{\mathrm{dl}}) + \alpha^2/(\sigma^2+\alpha^2P_{\mathrm{dl}}) + 1/\sigma^2$, $c_2 = 1/(\sigma^2+\alpha^2P_{\mathrm{dl}}+\beta^2P_{\mathrm{dl}}) + \alpha^2/(\sigma^2+\alpha^2P_{\mathrm{dl}}) + \alpha^2/\sigma^2$, and $c_3 = \alpha^2/(\sigma^2+\alpha^2P_{\mathrm{dl}}+\beta^2P_{\mathrm{dl}}) + 1/(\sigma^2+\alpha^2P_{\mathrm{dl}}) + \alpha^2/\sigma^2$.

For the ease of exposition, the sum capacity lower bounds for the Wyner model are calculated below, where only the ICLI of neighboring antenna-sites with strength α is considered. The number of user groups in each antenna-cluster is assumed to be identical with the cluster size, M.

$$\bar{C}_{\mathrm{WB,Wyner}} = \log\left(1 + P_{\mathrm{dl}}^{M}\prod_{i=1}^{M} c_i^{\mathrm{WB}}\right), \tag{8}$$

where $c_1^{\mathrm{WB}} = c_2^{\mathrm{WB}} = \frac{1}{\sigma^2+\alpha^2P_{\mathrm{dl}}} + \alpha^2/\sigma^2$, $c_3^{\mathrm{WB}} = c_4^{\mathrm{WB}} = \frac{\alpha^2}{\sigma^2+\alpha^2P_{\mathrm{dl}}} + (1+\alpha^2)/\sigma^2$, and $c_i^{\mathrm{WB}} = (1+2\alpha^2)/\sigma^2, i > 4$. Similarly, using the GT scheme and Corollary 1,

$$\bar{C}_{\mathrm{GT,Wyner}} = \log\left(1 + P_{\mathrm{dl}}^{M}\prod_{i=1}^{M} c_i^{\mathrm{GT}}\right), \tag{9}$$

where $c_1^{\mathrm{GT}} = c_2^{\mathrm{GT}} = \frac{(1-\varepsilon)\log_e K}{\sigma^2+\alpha^2P_{\mathrm{dl}}} + \alpha^2/\sigma^2$, $c_3^{\mathrm{GT}} = c_4^{\mathrm{GT}} = \frac{\alpha^2}{\sigma^2+\alpha^2P_{\mathrm{dl}}} + ((1-\varepsilon)\log_e K + \alpha^2)/\sigma^2$, $c_i^{\mathrm{GT}} = ((1-\varepsilon)\log_e K + 2\alpha^2)/\sigma^2, i > 4$, and $0 < \varepsilon < 1$.

IV. Optimal Cluster Size with Imperfect CSI

In this section, the optimal antenna cluster size in the cell-free L-DAS is considered, in the sense of maximizing the downlink achievable rate when considering ICLI, imperfect channel estimation and the overhead of channel training and feedback. Both analog and digital feedback schemes are considered.

A. Training-based System Achievable Rates

To obtain the achievable sum rates w.r.t. the CSI estimation overhead and imperfect CSI, we consider a typical training-based system, where the channel training and feedback procedures are specified in the following subsections. We assume the block fading model, where channel gain matrix $\boldsymbol{H}$ is constant over each frame of length T channel uses, and evolves from frame to frame according to an ergodic stationary spatially white jointly Gaussian process. Here we assume the training and feedback processes of different clusters occupy orthogonal time-frequency resources, i.e., the pilot contamination [1] is not considered since our focus is on the CSI acquisition overhead.

1) Channel Training: Antenna i transmits τ_{c} shared pilots ($\tau_{\mathrm{c}} \geq 1$) on the downlink. User j estimates its channel from antenna i based on the downlink channel output $s_{i,j}$

$$s_{i,j} = \sqrt{\tau_{\mathrm{c}}P_{\mathrm{dl}}}\boldsymbol{H}_{i,j} + z_{i,j}, \tag{10}$$

where $z_{i,j} \sim \mathcal{CN}(0, \sigma_{\rm n}^2)$. The MMSE estimation $\boldsymbol{H}_{i,j}^{\rm est,UE}$ of $\boldsymbol{H}_{i,j}$ given $s_{i,j}$ is

$$\begin{aligned}\boldsymbol{H}_{i,j}^{\rm est,UE}&=\mathbb{E}[\boldsymbol{H}_{i,j}s_{i,j}]\mathbb{E}[|s_{i,j}|^2]^{-1}s_{i,j}\\&=\frac{\boldsymbol{V}_{i,j}\sqrt{\tau_c P_{\rm dl}}}{\boldsymbol{V}_{i,j}\tau_c P_{\rm dl}+\sigma_{\rm n}^2}s_{i,j},\end{aligned} \tag{11}$$

with estimation error variance [37]

$$\sigma_c^2 = \frac{\boldsymbol{V}_{i,j}}{\boldsymbol{V}_{i,j}\tau_c P_{\rm dl}/\sigma_{\rm n}^2+1}. \tag{12}$$

2) Channel Feedback:

a) Analog Feedback: Each user feeds back $\tau_{\rm fb}$ symbols of its channel immediately after it finishes receiving the training symbols from the BS using unquantized quadrature-amplitude modulation (QAM)[23]. The resulting channel estimation and error variance are[6]

$$\begin{aligned}\boldsymbol{H}_{i,j}&=\boldsymbol{H}_{i,j}^{\rm est,BS}+e_{i,j}^{\rm Ana},\\\sigma^2_{e_{i,j}^{\rm Ana}}&=\boldsymbol{V}_{i,j}\left(\frac{1}{\boldsymbol{V}_{i,j}\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1}\right.\\&\quad\left.+\frac{\boldsymbol{V}_{i,j}\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2}{(\boldsymbol{V}_{i,j}\tau_c P_{\rm dl}/\sigma_{\rm n}^2+1)(\boldsymbol{V}_{i,j}\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1)}\right),\end{aligned} \tag{13}$$

where $P_{\rm ul}$ is defined as the transmit power of each user in the channel feedback precess.

b) Digital Feedback: Each user feeds back a quantized version of the channel estimates by B bits, where we assume an error-free AWGN channel[7]. The channel estimation of the i-th row of the channel matrix, i.e., $\boldsymbol{h}_i = [\boldsymbol{H}_{i,1}, \boldsymbol{H}_{i,2}, ..., \boldsymbol{H}_{i,M}]$ is

$$\boldsymbol{h}_i = \boldsymbol{h}_i^{\rm est,BS} + \underbrace{\boldsymbol{n}_i - \boldsymbol{w}_i^{\rm Dig}}_{\boldsymbol{e}_i^{\rm Dig}}, \tag{14}$$

where $\boldsymbol{n}_i$ and $\boldsymbol{w}_i^{\rm Dig}$ are uncorrelated estimation error related to channel training (12) and digital feedback, respectively. Based on the random vector quantization theory (RVQ) [38],

$$\begin{aligned}\sum_{j=1}^{M}\sigma^2_{e_{i,j}^{\rm Dig}}&=\mathbb{E}(||\boldsymbol{w}_i^{\rm Dig}||^2)+\mathbb{E}(||\boldsymbol{n}_i||^2)\\&=\sum_{j=1}^{M}\frac{2^{-\frac{B_{i,j}}{M-1}}\boldsymbol{V}_{i,j}^2\tau_c P_{\rm dl}/\sigma_{\rm n}^2+\boldsymbol{V}_{i,j}}{\boldsymbol{V}_{i,j}\tau_c P_{\rm dl}/\sigma_{\rm n}^2+1},\end{aligned} \tag{15}$$

where $B_{i,j} = \tau_{\rm fb}(M-1)\log(1+\boldsymbol{V}_{i,j}P_{\rm ul}/\sigma_{\rm n}^2)$. Here, each user uses one symbol to transmit the norm information for fair comparison with the analog feedback scheme [23].

3) Data Transmission: During the data transmission phase, the BS uses the feedback estimate $\boldsymbol{H}_{i,j}^{\rm est,BS}$ as the known channel. The transmission can be expressed as, affected by the ICLI,

$$\boldsymbol{y} = \boldsymbol{H}^{\rm est,BS}\boldsymbol{x} + \boldsymbol{E}\boldsymbol{x} + \boldsymbol{n} + \boldsymbol{s}, \tag{16}$$

where $\boldsymbol{E}$ is the estimation error matrix, and

$$\boldsymbol{H} = \boldsymbol{H}^{\rm est,BS} + \boldsymbol{E}. \tag{17}$$

[6]The uplink channel SNR is $P_{\rm ul}/\sigma_{\rm n}^2$.

[7]The assumption of a error-free unfaded AWGN feedback channel is unrealistic in the sense that the block length prevents us from coding over a long interval, and the feedback delay is not considered. Whereas it is still a reasonable and useful simplification mathematically in our analysis.

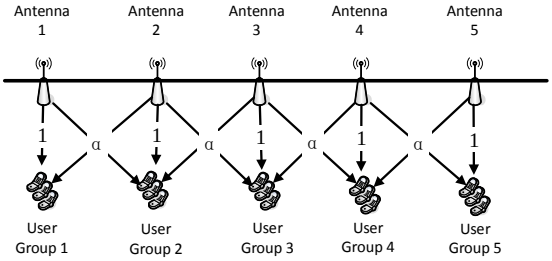


Fig. 3. Specifications of the Wyner channel model. Users in the same group have identical large-scale fading coefficients, i.e., pathloss, to the antennas. And only the ICLI of the *neighboring antennas* are considered, the channel gain magnitude from the neighboring antenna to this user group is α.

The entries of $\boldsymbol{E}$ are Gaussian distributed with variance $\sigma^2_{e_{i,j}^{\rm Ana}}$ or $\sigma^2_{e_{i,j}^{\rm Dig}}$, corresponding to analog (AF) or digital feedback (DF) scheme. According to [22], the achievable rate of the channel (16) is lower bounded by observing the transmission with uncorrelated Gaussian noise. Therefore,

$$e_i = \{\mathbb{E}[\boldsymbol{E}\boldsymbol{x}\boldsymbol{x}^\dagger\boldsymbol{E}^\dagger]\}_{i,i} = \begin{cases}\displaystyle\sum_{j=1}^{M}\sigma^2_{e_{i,j}^{\rm Ana}}P_{\rm dl}, & \text{for AF,}\\ \displaystyle\sum_{j=1}^{M}\sigma^2_{e_{i,j}^{\rm Dig}}P_{\rm dl}, & \text{for DF,}\end{cases} \tag{18}$$

The ICLI, i.e., the variance profile of $\boldsymbol{s}$ in (16), is determined by the user locations and antenna locations, which will be specified in the following subsection.

The achievable sum rate with optimum number of training symbols and feedback bits is

$$\begin{aligned}&R_{\rm opt,tf}(M) = \underset{\tau_c,\tau_{\rm fb}}{\rm maximize}\, R(\tau_c,\tau_{\rm fb})\\&s.t.\ \ \tau_c \geq 1, \tau_{\rm fb} \geq 1, \tau_{\rm fb}MN + \tau_c M \leq T,\end{aligned} \tag{19}$$

where $R(\tau_c, \tau_{\rm fb})$ is the sum rate regarding imperfect CSI and ICLI. In our analysis, $R(\tau_c, \tau_{\rm fb})$ corresponds to the capacity lower bound derived in Theorem 1. The optimal training and feedback problem, i.e., the optimal $\tau_{\rm c}$ and $\tau_{\rm fb}$, of downlink multi-user systems is studied in [39]. Our focus is on the optimal cluster size, therefore an exhaustive search is used here to find the optimal $\tau_{\rm c}$ and $\tau_{\rm fb}$.

Generally speaking, given the CSIT, we consider the dirty-paper-coding (DPC) [11][13] scheme, which is proved to be the capacity-achieving coding scheme for Gaussian broadcast channels, and does not necessarily require that the number of users N is less than the number of BS antennas M. In particular, we derived a lower bound of the sum rate achieved by the DPC scheme in Section III, and use it as the objective in our achievable rate maximization analysis in this section.

B. The Optimal Cluster Size Analysis

In general, the explicit analysis for the optimal cluster size in DAS is intractable, for the reason that the channel matrix variance profile can be arbitrary, depending on the user and antenna locations. For the ease of exposition simplicity, we consider the scenario where the number of user groups in Theorem 1 equals to the cluster size, i.e., $G = M$, and each

group has K users. Denote the k-th user in the g-th group as the user g_k, and denote the user set of the m-th group as $\mathcal{G}_m$. The Wyner model is adopted[8], see Fig. 3, which suggests that

- Users in the same group have identical large-scale fading coefficients, i.e., pathloss, to the antennas. And only the ICLI of the *neighboring antennas* is considered, the channel gain magnitude from the neighboring antenna to this user group is α.
- The antennas are evenly distributed on an infinite line (one-dimensional). Later in this section, in Corollary 2, this assumption is removed.

The fixed training and feedback schemes are assumed, i.e., $\tau_{\rm c}$ and $\tau_{\rm fb}$ are fixed in the analysis. In general, this is suboptimal since they should be optimized accordingly. However, according to [39], its performance is quite close to the optimal scheme especially in the high-SNR regime, since high-SNR training and feedback leads to negligible CSI estimation error. The analysis is based on the analog feedback scheme. Nonetheless, the analysis for digital feedback scheme is similar, thus omitted for brevity.

Under this system setting, we consider the problem: What is the optimal cluster size, i.e., the optimal number of cooperating distributed antennas, that can maximize the per-user-group sum rate[9]. To this end, we obtain the following theorem.

Theorem 2: Define

$$\begin{aligned} a=&\frac{1}{\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1}+\frac{\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2}{(\tau_c P_{\rm dl}/\sigma_{\rm n}^2+1)(\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1)}\\ b=&\alpha^2\left(\frac{1}{\alpha^2\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1}\right.\\ &\left.+\frac{\alpha^2\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2}{(\alpha^2\tau_c P_{\rm dl}/\sigma_{\rm n}^2+1)(\alpha^2\tau_{\rm fb}P_{\rm ul}/\sigma_{\rm n}^2+1)}\right). \end{aligned} \tag{20}$$

The optimal cluster size $M_{\rm opt}$ for one-dimensional systems in the high-SNR regime is

$$M_{\rm opt}=\left[\frac{\log\frac{(c_5^{\rm Ana})^4}{(c_1^{\rm Ana})^2(c_3^{\rm Ana})^2}}{(3\tau_{\rm fb}K+\tau_{\rm c})\log(P_{\rm dl}c_5^{\rm Ana})}T\right]^{\frac{1}{2}}, \tag{21}$$

where $c_1^{\rm Ana}=c_2^{\rm Ana}=\frac{1}{\sigma_{\rm n}^2+\alpha^2P_{\rm dl}+(a+b)P_{\rm dl}}+\alpha^2/(\sigma_{\rm n}^2+(a+2b)P_{\rm dl})$, $c_3^{\rm Ana}=c_4^{\rm Ana}=\frac{\alpha^2}{\sigma_{\rm n}^2+\alpha^2P_{\rm dl}+(a+b)P_{\rm dl}}+(1+\alpha^2)/(\sigma_{\rm n}^2+(a+2b)P_{\rm dl})$, $c_i^{\rm Ana}=(1+2\alpha^2)/(\sigma_{\rm n}^2+(a+2b)P_{\rm dl})$, $\forall i>4$. The achievable per-group sum rate with the optimal cluster size is

$$R_{\rm opt,cs}=\left[\sqrt{\log(P_{\rm dl}c_5^{\rm Ana})}-\sqrt{\frac{3\tau_{\rm fb}K+\tau_{\rm c}}{T}\log\frac{(c_5^{\rm Ana})^4}{(c_1^{\rm Ana})^2(c_3^{\rm Ana})^2}}\right]^2. \tag{22}$$

Proof: See Appendix C. ∎

Remark 1: In practice,

$$T=T_{\rm c}W_{\rm c}\sim\frac{1}{v}, \tag{23}$$

where $T_{\rm c}$ and $W_{\rm c}$ denote the channel coherence time and coherence bandwidth respectively, and v is the movement speed of the users. Hence it is observed that if averagely, the users move 2 times faster, the cluster size should be $1/\sqrt{2}$ times smaller. Similarly, if T is fixed, as

$$M_{\rm opt}\sim\sqrt{\frac{1}{3\tau_{\rm fb}K+\tau_{\rm c}}}, \tag{24}$$

if the number of users in each group increases by 2 times, the optimal cluster size shrinks by approximately $1/\sqrt{2}$, because the channel estimation overhead increases as the users.

Remark 6: When $T\to\infty$, $R_{\rm opt,cs}\to\log(P_{\rm dl}c_5^{\rm Ana})$, which is the ICLI-free per-group capacity with perfect CSIT. There is an intuitive explanation that when T goes to infinity, the overhead of channel estimation is negligible and since the optimal cluster size is large ($M_{\rm opt}\sim\sqrt{T}$), the ICLI vanishes. Remark that the convergence rate to the ICLI-free sum rate is proportional to $\sqrt{\frac{1}{T}}$ according to (22).

Remark 7: : If the ICLI signal strength, α is zero, i.e., substitute $\alpha=0$ into (43), then $M_{\rm opt}=1$. In this case no cooperation is needed since the channel matrix is inherently diagonal, forming parallel subchannels in space.

Now we extend our results to two-dimensional topology, and we have the following corollary:

Corollary 2: The optimal cluster size $M_{\rm opt}$ for the two-dimensional systems is asymptotically

$$M_{\rm opt}\sim T^{\frac{2}{3}}. \tag{25}$$

Proof: See Appendix D. ∎

Remark 8: For two-dimensional antenna-topology systems, the closed-form expression of the optimal cluster size is not found because the multi-antenna topology can be very different in two-dimensional systems (e.g. hexagon cells, square cells, or even random-locations [31]). Nevertheless, the scaling result given in Corollary 2 is irrelevant to the multi-antenna topology, because from the proof it is observed that this result is based on the fact that there are roughly $\mathcal{O}(\sqrt{M})$ groups of users on the cluster edge in two-dimensional systems, regardless of the multi-antenna topology.

V. Numerical Results

Fig. 4 shows the ergodic per-group capacity and the lower bounds of the DAS with one-dimensional multi-antenna topology. The lower bounds correspond to the WB and GT (with $\varepsilon=0.2$) rate-achieving scheduling schemes, respectively. The ergodic capacity Monte-Carlo simulation result and the

[8]Admittedly, mathematical simplicity is the reason for both assumptions as follows. The Wyner model is first studied by [8] to provide a tractable model for cellular networks. There are some works [40][16] considering the accuracy of the Wyner model since it depends solely on one parameter α which characterizes the inter-cell signal strength, showing that the Wyner model is inaccurate when considering user outage probability. But [16] shows the Wyner model can be adopted to handle metrics like sum rate or average rate, which is our concern in this work.

[9]The per-group rate is maximized here for the reason that we want to maximize the sum rate of all the antennas, which belong to various clusters. Notice that the number of groups equals the cluster size by assumption. Therefore maximizing the sum rate is equivalent to maximizing the per-group rate.

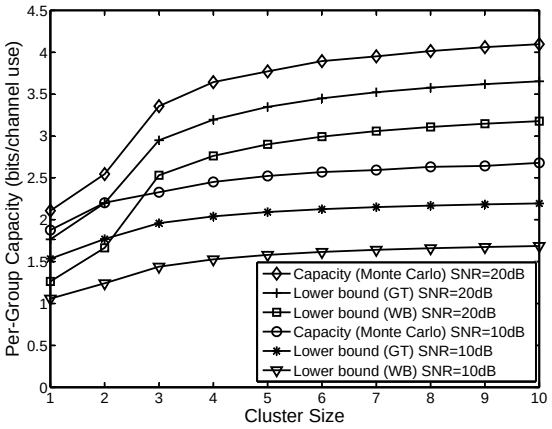


Fig. 4. Downlink per-group capacity and bounds with independent Rayleigh fading channels, where inter-cell signal strength $\alpha = \sqrt{0.1}, \epsilon = 0.2$, number of users in each cell $K = 16$.

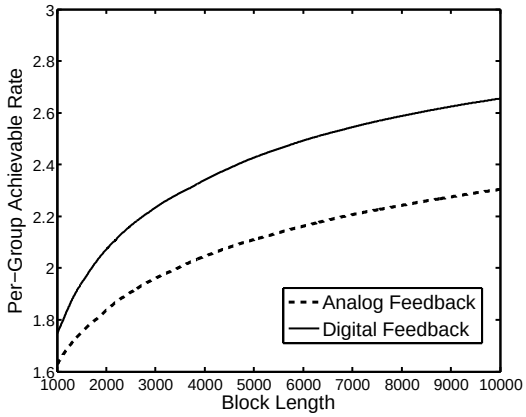


Fig. 7. The impact of block length on per-group achievable rate with optimal cluster size and optimal $\tau_{\rm c}$, $\tau_{\rm fb}$, the number of users in each group $K = 4$, $\mathsf{SNR} = 20$ dB, and ICLI signal strength $\alpha = \sqrt{0.1}$.

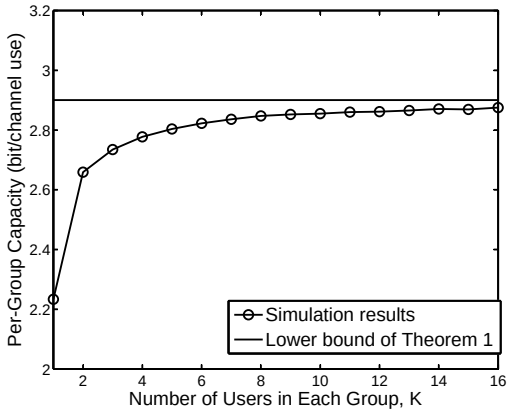


Fig. 5. The impact of finite number of users in each cell, K, on per-group capacity Monte-Carlo simulation results compared with the lower bound of Theorem 1, which is derived assuming $K \to \infty$. Inter-cell signal strength $\alpha = \sqrt{0.1}$, $\mathsf{SNR} = 20$ dB.

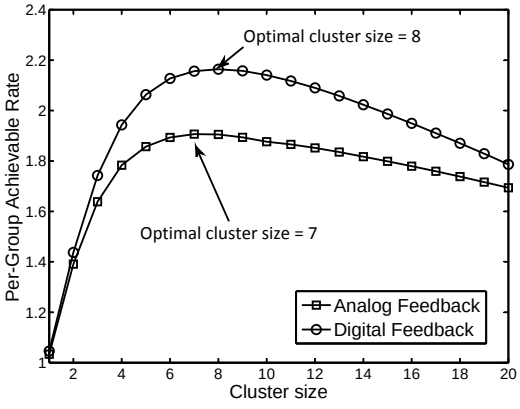


Fig. 6. The impact of cluster size on downlink per-group achievable rate with optimal $\tau_{\rm c}$, $\tau_{\rm fb}$, block length $T = 2500$, the number of users in each cell $K = 4$, $\mathsf{SNR} = 20$ dB, and inter-cell signal strength $\alpha = \sqrt{0.1}$. In this figure, the optimal cluster size is 7 and 8 for analog feedback scheme and digital feedback scheme, respectively.

lower bound of the GT scheme is calculated with $K = 16$. Examining the difference between two lower bounds reveals the multi-user diversity gain as we discussed in Theorem 1.

Fig. 5 shows the impact of the number of users per cell K on the capacity lower bound from Theorem 1. The figure illustrates the appropriate value of K in our analysis (since we assume $K \to \infty$ to get the closed-form expression). The Monte-Carlo simulation gives the lower bound in (26) with finite K. The lower bound of Theorem 1 is computed by (8) with $K \to \infty$. It is observed that the ergodic capacity of $K = 4$ is already quite close (with an about 0.1 bits per channel use difference) to the lower bound in (8), indicating although Theorem 1 is derived assuming $K \to \infty$, it is applicable even if the number of users is quite limited.

Fig. 6 shows the impact of cluster size of the distributed antenna system on the per-group achievable rate (PGR), under the one-dimensional system topology. The PGR is obtained by calculating the achievable rate derived in Theorem 1. Here $\alpha = \sqrt{0.1}$, block length $T = 2500$, downlink $\mathsf{SNR} = 20$ dB, uplink $\mathsf{SNR} = 10$ dB, and the number of users in each group $K = 4$. According to [18], the sum capacity of MU-MIMO systems with $K = 4$ is quite close to that with $K \to \infty$. The digital feedback scheme achieves higher achievable rate than analog feedback. This is because based on (15), the estimation error decays exponentially with the number of feedback bits for digital feedback, comparing with in (13), where the estimation error only decays as $1/\tau_{\rm fb}$ for analog feedback. Besides, from (15), the estimation error related to feedback goes to zero if the feedback link SNR is large in digital feedback scheme. It is also observed that when the cluster size is large, the performance of analog feedback is as good as digital feedback, because as the cluster size becomes larger, there will be more channels to estimate. Therefore the optimal $\tau_{\rm c}$ and $\tau_{\rm fb}$ are both almost 1, in which case the analog feedback scheme is optimal.

Figs. 7-9 show the impact of block length on the achievable rate and the optimal cluster size. Here $\alpha = \sqrt{0.1}$, $K = 4$, downlink $\mathsf{SNR} = 20$ dB, uplink $\mathsf{SNR} = 10$ dB. Fig. 7 is the simulation result of the achievable rate with optimal cluster size versus the block length with the one-dimensional Wyner model. The digital feedback scheme obtains higher achievable rate. Moreover, shown in Fig. 8, as the digital

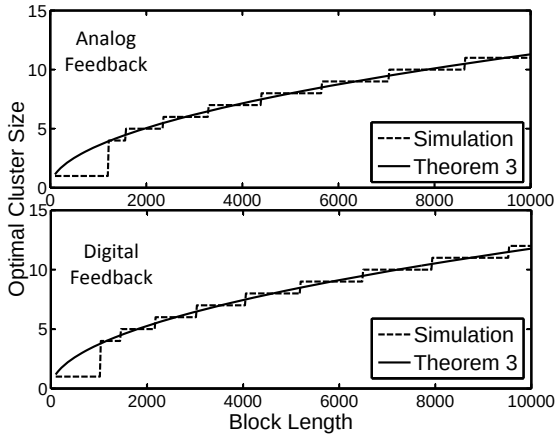


Fig. 8. Optimal cluster size versus block length of one-dimensional systems with $\tau_{\rm c} = 1$, $\tau_{\rm fb} = 4$, the number of users in each group $K = 4$, SNR $= 20$ dB, and ICLI signal strength $\alpha = \sqrt{0.1}$.

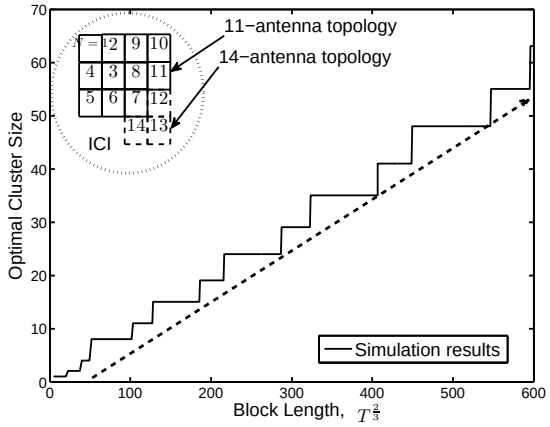


Fig. 9. Optimal cluster size versus block length, $T^{\frac{2}{3}}$, of two-dimensional systems for analog feedback scheme with $\tau_{\rm c} = 1$, $\tau_{\rm fb} = 4$, the number of users in each group $K = 4$, SNR $= 20$dB, and ICLI signal strength $\alpha = \sqrt{0.1}$. The multi-antenna topology of a antenna cluster by M cells is zigzag, and shown in the figure, where each square represents one antenna. Only one cluster is depicted for brevity.

feedback scheme can better suppress channel feedback error, less feedback bits are needed. Thus, the optimal cluster size of digital feedback is larger. The optimal cluster size versus the block length T of one-dimensional systems with fixed training and feedback is shown in Fig. 8, we can observe that the closed-form expression derived in Theorem 2 gives accurate approximation of the optimal cluster size for both analog and digital feedback schemes. Fig. 9 shows the simulation result of the optimal cluster size of a two-dimensional system. The multi-cell topology is shown in the figure. The signal of one BS affects the neighboring 4 adjacent cells with signal strength α. The sum achievable rate is calculated by Theorem 1. Only the analog feedback scheme is shown since the case for the digital feedback scheme is similar. The power law between the optimal cluster size and the block length, $N_{opt} \sim T^{\frac{2}{3}}$, is observed, which is expected by Corollary 2. It is important to notice that, in practice, the optimal cluster size is also limited by other factors, such as backhaul capacity. Therefore, the curves in Fig. 8 and Fig. 9 cannot scale with the block length T all the way up. Eventually, the curves will flatten, which is the consequences of other cluster size limitations. However it is still of great importance to study the scaling behavior with the block length.

Fig. 10 shows the impact of downlink SNR on the optimal cluster size. The uplink SNR is set 10 dB lower than the downlink SNR. Here $\alpha = \sqrt{0.1}$, $K = 4$, and $\tau_{\rm c}$ and $\tau_{\rm fb}$ are chosen to optimize the achievable rate. When the SNR is low, the optimal cluster size is small, i.e., the benefit of large cluster size is limited. As the SNR gets higher, the optimal cluster size is larger to better suppress the ICLI.

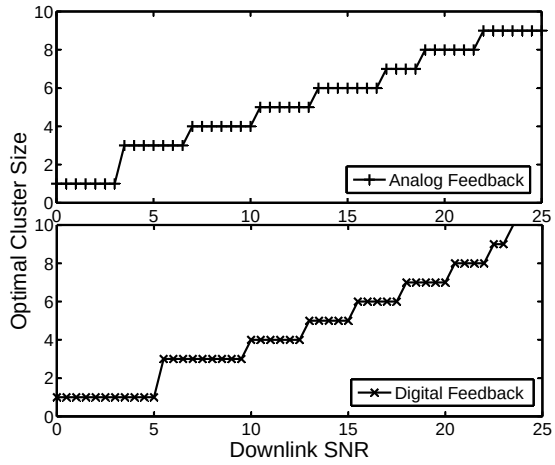


Fig. 10. The impact of downlink SNR on the optimal cluster size with optimal $\tau_{\rm c}$, $\tau_{\rm fb}$, where the number of users in each group $K = 4$, block length $T = 2500$, and the ICLI signal strength $\alpha = \sqrt{0.1}$.

TABLE I
SYSTEM PARAMETERS

Carrier frequency f_c	2.4 GHz
Bandwidth	10 MHz
Thermal noise	-174 dBm/Hz
Pathloss model	Pathloss $= 26.7 \log_{10}(d) + 22.7$ $+26 \log_{10}(f_c)$
Antenna-spacing	100 m

A. Impact of Random User Locations

In our analysis, we assume the users in each group have identical large-scale fading coefficients. To be more general, we consider a scenario where MK users are uniformly distributed on a line, i.e., the one-dimensional system is considered. The antennas are located with constant distances between each other. The simulation parameters are shown in Table I. The achievable sum rates are calculated by running Monte-carlo simulations on the capacity lower bound in (26). The impact of the number of users on the optimal cluster size is shown in Fig. 11. In particular, from Fig. 11, the square-root scaling law proved in Theorem 1 is well observed, which shows that the identical pathloss assumption not only simplifies our analysis, but also is of good practical accuracy.

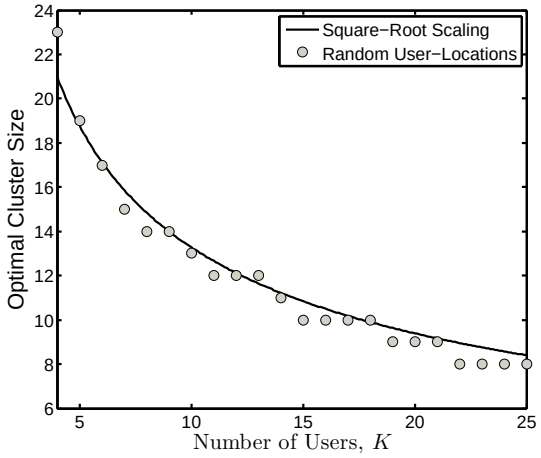


Fig. 11. The optimal cluster size versus the number of users in each group, K, with random user-locations.

VI. Conclusions

Closed-form lower bounds of BC downlink ergodic sum capacity are derived, corresponding to different rate-achieving user-scheduling schemes, assuming sufficiently large number of users. The theorem is applicable to arbitrary multi-antenna topology. The lower bound corresponding to the GT scheduling scheme achieves the multi-user diversity gain, thus is tighter than the WB scheduling scheme.

In what follows, considering the interplay between the ICLI and the imperfect CSI acquisition in FDD cell-free L-DAS, closed-form expressions of the optimal cluster size and the corresponding achievable rate are derived concerning the one-dimensional Wyner model. We further show that in the asymptotic regime where the block length T is large, the optimal cluster size for two-dimensional systems should be $M_{\text{opt}} \sim T^{\frac{2}{3}}$. Therefore, it is evident that the optimal cluster size is strongly related to the multi-antenna topology, due to the various ICLI environments. Numerical results show the impact of block length and SNR on the system achievable rate and the optimal cluster size, which validates our analysis of Theorem 2 and Corollary 2. The results also show the digital feedback scheme has better achievable rate and larger optimal cluster size than the analog feedback scheme because the digital feedback scheme can better suppress the error of imperfect channel estimation. Possible future directions include considering more sophisticated channel statistics, where the user-channels have non-identical channel fading distributions.

Appendix A
Proof of Theorem 1

Proof: Based on our previous work [18], the ergodic capacity is lower bounded by

$$C_{\text{dl}} \geq \mathbb{E}_{\tilde{\boldsymbol{H}}}\left[\log\left(1+\det\left(\frac{P_{\text{eq}}}{K}\tilde{\boldsymbol{H}}^{\dagger}\tilde{\boldsymbol{H}}\right)\right)\right]. \tag{26}$$

The expectation is taken over all realizations of $\tilde{\boldsymbol{H}}$. With $K \gg 1$, applying the strong law of large numbers (SLLN), the diagonal entries of $\frac{1}{K}\tilde{\boldsymbol{H}}^{\dagger}\tilde{\boldsymbol{H}}$ are deterministic and can be derived as follows

$$\begin{aligned}\left[\frac{1}{K}\tilde{\boldsymbol{H}}^{\dagger}\tilde{\boldsymbol{H}}\right]_{i,i} &= \frac{1}{K}\sum_{j=1}^{GK}\frac{1}{\sigma_j^2}\boldsymbol{H}_{j,i}\boldsymbol{H}_{j,i}^{\dagger}\\ &= \sum_{n=1}^{G}\frac{1}{\sigma_j^2}\left(\frac{1}{K}\sum_{m=1}^{K}\boldsymbol{H}_{m+(n-1)K,i}\boldsymbol{H}_{m+(n-1)K,i}^{\dagger}\right)\\ &\xrightarrow{\text{SLLN}} \sum_{j=1}^{G}\frac{1}{\sigma_j^2}u_{i,j}^2,\end{aligned} \tag{27}$$

where $u_{i,j}$ denotes the channel gain magnitude from antenna i to the users in the j-th group . As the entries of $\boldsymbol{H}$ are independent and have zero means, applying SLLN again, the off-diagonal entries of $\frac{1}{K}\tilde{\boldsymbol{H}}^{\dagger}\tilde{\boldsymbol{H}}$ are all zeros. Therefore, instead of taking the expectation in (26), the term inside converges to a deterministic constant, which can be calculated explicitly, i.e.,

$$\log\left(1+\det\left(\frac{P_{\text{eq}}}{K}\tilde{\boldsymbol{H}}^{\dagger}\tilde{\boldsymbol{H}}\right)\right) \overset{K\to\infty}{\longrightarrow} \log\left(1+P_{\text{eq}}^{M}\prod_{i=1}^{M}c_i\right), \tag{28}$$

where the definitions of the c_i's are shown below (5). Then we complete the proof. ■

Appendix B
Proof of Corollary 1

Proof: The proof is almost the same with Theorem 1. We have

$$C_{\text{dl}} \geq \mathbb{E}_{\tilde{\boldsymbol{H}}_{\text{GT}}}\left[\log\left(1+\det\left(\frac{P_{\text{eq}}}{L}\tilde{\boldsymbol{H}}_{\text{GT}}^{\dagger}\tilde{\boldsymbol{H}}_{\text{GT}}\right)\right)\right], \tag{29}$$

where $\tilde{\boldsymbol{H}}_{\text{GT}} \in \mathcal{C}^{LG\times M}$ denotes the channel matrix of the selected users in GT. And

$$\begin{aligned}\left[\frac{1}{L}\tilde{\boldsymbol{H}}_{\text{GT}}^{\dagger}\tilde{\boldsymbol{H}}_{\text{GT}}\right]_{i,i} &= \frac{1}{L}\left[\frac{1}{\sigma_i^2}\sum_{m\in\mathcal{S}_i}\boldsymbol{H}_{m,i}\boldsymbol{H}_{m,i}^{\dagger}\right]\\ &+\frac{1}{L}\left[\sum_{n\neq i}^{G}\frac{1}{\sigma_n^2}\sum_{m\in\mathcal{S}_n}\boldsymbol{H}_{m,i}\boldsymbol{H}_{m,i}^{\dagger}\right],\end{aligned} \tag{30}$$

where $\mathcal{S}_n$ denotes the user-index set of selected users from the n-th group by GT. The first term on the right-hand-side of the equality (30) is denoted by $\mathcal{L}_1$, and the second $\mathcal{L}_2$. By the design of GT, the selected users of $\mathcal{S}_i$ are those with the largest channel gains to the i-th antenna in each sub-group. Based on a result of [42], according to which the maximum of n i.i.d. $\chi^2(2)$ random variables $a_{\max}$ satisfies

$$\begin{aligned}&\Pr\{-\log\log n+\mathcal{O}(\log\log\log n)\\ &\leq a_{\max}-\log n \leq\\ &\log\log n+\mathcal{O}(\log\log\log n)\}\\ &>1-\mathcal{O}\left(\frac{1}{\log n}\right).\end{aligned} \tag{31}$$

As in the GT scheme, the scheduler chooses the users with the strongest channel gains among K/L users in the sub-group.

Let $n = K/L = K^{1-\varepsilon}$ (omitting little orders of $\log n$), we have

$$\mathcal{L}_1 \xrightarrow{\text{SLLN}} \frac{u_{i,i}^2}{\sigma_i^2}(1-\varepsilon)\log_e K. \tag{32}$$

Based on the fact that the channels of different antennas to the user are independent, we have

$$\mathcal{L}_2 \xrightarrow{\text{SLLN}} \sum_{j\neq i}^{G} \frac{u_{i,j}^2}{\sigma_j^2}. \tag{33}$$

And the off-diagonal entries are zero when $L \to \infty$. For the SLLN to hold in (30), the constant L should be chosen so that $L \to \infty$ as $K \to \infty$. In particular, let

$$L = K^{\varepsilon} \Rightarrow \frac{K}{L} = K^{1-\varepsilon}, 0 < \varepsilon < 1. \tag{34}$$

Following the same arguments as in Theorem 1, we obtain

$$\log\left(1+\det\left(\frac{P_{\text{eq}}}{L}\tilde{\boldsymbol{H}}_{\text{GT}}^{\dagger}\tilde{\boldsymbol{H}}_{\text{GT}}\right)\right) \overset{K\to\infty}{\longrightarrow} \log\left(1+P_{\text{eq}}^{M}\prod_{i=1}^{M}c_i\right), \tag{35}$$

where $P_{\text{eq}} = \frac{P_{\text{dl}}M}{G}$, $c_i = \sum_{j\neq i}^{G}\frac{u_{i,j}^2}{\sigma_j^2} + \frac{u_{i,i}^2}{\sigma_i^2}(1-\varepsilon)\log_e K$. Then we complete the proof. ∎

Appendix C
Proof of Theorem 2

Proof: Based on the Wyner model, for the analog feedback scheme, the ICLI and the channel estimation error are

$$s_i = \begin{cases} \alpha^2 P_{\text{dl}}, \text{ user } i \in G_1 \bigcup G_M \\ 0, \text{ else} \end{cases},$$

$$e_i^{\text{Ana}} = \{\mathbb{E}[\boldsymbol{E}\boldsymbol{x}\boldsymbol{x}^{\dagger}\boldsymbol{E}^{\dagger}]\}_{i,i} = \sum_{j=1}^{N}\sigma_{e_{i,j}^{\text{Ana}}}^2 P_{\text{dl}}, \tag{36}$$

and

$$e_i^{\text{Ana}} = \begin{cases} (a+b)P_{\text{dl}}, \text{user } i \in G_1 \bigcup G_M \\ (a+2b)P_{\text{dl}}, \text{else} \end{cases} \tag{37}$$

Define

$$\begin{aligned} d &= \frac{(c_5^{\text{Ana}})^4}{(c_1^{\text{Ana}})^2(c_3^{\text{Ana}})^2}, \\ e &= \log(P_{\text{dl}}c_5^{\text{Ana}}) + \frac{3\tau_{\text{fb}}K+\tau_{\text{c}}}{T}\log d, \\ f &= \frac{3\tau_{\text{fb}}K+\tau_{\text{c}}}{T}\log(P_{\text{dl}}c_5^{\text{Ana}}),\ g = \log d, \end{aligned} \tag{38}$$

where f and g are irrelevant with M. Notice, the users located in the center of the user groups (far from the interfering antennas) experience smaller ICLI, thus $c_5^{\text{Ana}} > c_3^{\text{Ana}}$, and $c_5^{\text{Ana}} > c_1^{\text{Ana}}$. Thus $f > 0, g > 0$. Then the per-group achievable rate is as follows with c_i^{Ana} given in (21). We have

$$\begin{aligned} &R^{\text{Ana}}(M) \\ &= \frac{1}{M}\left(1-\frac{3\tau_{\text{fb}}MK+\tau_{\text{c}}M}{T}\right) \\ &\quad \log\left(1+P_{\text{dl}}^{M}\prod_{i=1}^{M}c_i^{\text{Ana}}\right) &(39)\\ &= \frac{1}{M}\left(1-\frac{3\tau_{\text{fb}}MK+\tau_{\text{c}}M}{T}\right) \\ &\quad \log\left(1+\left(P_{\text{dl}}c_5^{\text{Ana}}\right)^{M}\Big/\frac{\left(c_5^{\text{Ana}}\right)^4}{\left(c_1^{\text{Ana}}\right)^2\left(c_3^{\text{Ana}}\right)^2}\right) &(40)\\ &\overset{SNR\to\infty}{\approx} \frac{1}{M}\left(1-\frac{3\tau_{\text{fb}}MK+\tau_{\text{c}}M}{T}\right) \\ &\quad \log\left[\frac{\left(P_{\text{dl}}c_5^{\text{Ana}}\right)^{M}}{d}\right] &(41)\\ &= e - fM - g\frac{1}{M} &(42)\\ &\overset{(a)}{\leq} e - 2\sqrt{fg} \triangleq R_{\text{opt,cs}}, &(43) \end{aligned}$$

where inequality (43) is because the algebraic mean is no less than the geometry mean. When $M_{\text{opt}} = \sqrt{\frac{g}{f}}$, the equality holds. Thus M_{opt} is given in (21), along with $R_{\text{opt,cs}}$ given in (22). ∎

Appendix D
Proof of Corollary 2

Proof: In light of the proof of Theorem 2, the most important factor that affects the scaling result is the number of user-groups located at the cell edge. As in the one-dimensional (1D) system, 2 out of M user-groups are at cell edge. While in the two-dimensional (2D) system, roughly $\mathcal{O}(\sqrt{M})$ out of M user-groups are at the cell edge. Let c_{c} and c_{e} denote the ICLI coefficients of cell-center and cell-edge user-groups respectively. The per-group achievable rate for 2D system is

$$\begin{aligned} R_{\text{2D}} &\approx \frac{1}{M}\left(1-\frac{3\tau_{\text{fb}}MK+\tau_{\text{c}}M}{T}\right)\log\left(P_{\text{dl}}^{M}c_{\text{c}}^{M}/\frac{c_{\text{c}}^{\sqrt{M}}}{c_{\text{e}}^{\sqrt{M}}}\right) \\ &= \log\left(P_{\text{dl}}c_{\text{c}}\right) - \frac{3\tau_{\text{fb}}K+\tau_{\text{c}}}{T}\log\left(P_{\text{dl}}c_{\text{c}}\right)M \\ &+ \frac{3\tau_{\text{fb}}K+\tau_{\text{c}}}{T}\log\left(\frac{c_{\text{c}}}{c_e}\right)\sqrt{M} - \log\left(\frac{c_{\text{c}}}{c_{\text{e}}}\right)\frac{1}{\sqrt{M}}. \end{aligned} \tag{44}$$

By calculating the second derivatives of (44), it is proved that (44) is concave in M. Thus the optimal cluster size M_{opt} can be found by solving $\frac{\partial R_{\text{2D}}}{\partial M} = 0$, which gives

$$M_{\text{opt}} \approx \left[\frac{\log(\frac{c_{\text{c}}}{c_{\text{e}}})T}{(6\tau_{\text{fb}}K+2\tau_{\text{c}})\log(P_{\text{dl}}c_{\text{c}})}\right]^{\frac{2}{3}}. \tag{45}$$

Hence the scaling result of Corollary 2 is proved. ∎

REFERENCES

[1] 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.

[2] H. Yang and T. L. Marzetta, "Capacity performance of multicell large-scale antenna systems," *Allerton Conference on Communication, Control, and Computing 2013*, to appear.

[3] K.J. Kerpez, "A radio access system with distributed antennas," *IEEE Trans. Veh. Technol.*," vol. 45, no. 2, pp. 265-275, May 1996

[4] W. Choi and J. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69-73, Jan. 2007.

[5] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983-2989, Jun. 2009.

[6] P. Marsch and G. Fettweis, "A framework for optimizing the uplink performance of distributed antenna systems under a constrained backhaul," in 2007 *IEEE International Conference on Communications*, 2007, pp. 975-979.

[7] J. Zhang, R. Chen, J. G. Andrews, and R. W. Heath, "Coordinated multi-cell MIMO systems with cellular block diagonalization," in 2007 Conference Record of the Forty-First *Asilomar Conference on Signals*, Systems and Computers, 2007, pp. 1669-1673.

[8] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713-1727, 1994.

[9] P. Viswanath and D. N. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel," in *Proc.Int. Symp. Information Theory*, June 2002, p. 497.

[10] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875-1892, Sep. 2004.

[11] 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.

[12] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361-374, Feb. 2006.

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

[14] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, pp. 1-19, 2008.

[15] O. Somekh, B. M. Zaidel, and S. Shamai (Shitz), "Sum rate characterization of joint multiple cell-site processing," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4473-4497, Dec. 2007.

[16] J. Xu, J. Zhang, and J. G. Andrews, "On the accuracy of the Wyner model in downlink cellular networks," in *IEEE International Conference on Communications (ICC)*, 2011, pp. 1-5.

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

[18] Z. Jiang, S. Zhou, and Z. Niu, "Capacity bounds of downlink network MIMO systems with inter-cluster interference," *IEEE GLOBECOM 2012*, pp.4612-4617,

[19] Z. Jiang, S. Zhou, and Z. Niu, "Minimum power consumption of a base station with large-scale antenna array," in Proc. *IEEE APCC 2013*,

[20] S. Zhou, J. Gong, and Z. Niu, "Distributed adaptation of quantized feedback for downlink network MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 61-67, Jan. 2011.

[21] R. Bhagavatula and R. W. Heath, Jr., "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800-811, Feb. 2011.

[22] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[23] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845-2866, Jun. 2010.

[24] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *IEEE GLOBECOM*, 2009, pp. 1-6.

[25] M. J. Crisp, L. Sheng, A. Watts, R. V. Penty, and I. H. White, "Uplink and Downlink Coverage Improvements of 802.11g Signals Using a Distributed Antenna Network," *J. Lightwave Technology*, vol. 25, pp. 3388 - 3395, Nov. 2007

[26] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Communications Magazine*, vol. 51, pp. 161 - 167, Apr. 2013.

[27] J. Joung, Y. K. Chia, S. Sun, "Energy-Efficient, Large-scale Distributed-Antenna System (L-DAS) for Multiple Users," *IEEE J. Select. Topics, Signal Processing*, Mar. 2014 (early access articles)

[28] H. Kim, E. Park, H. Park, I. Lee, "Beamforming and Power Allocation Designs for Energy Efficiency Maximization in MISO Distributed Antenna Systems," *IEEE Communications Letters*, vol. 17, pp. 2100 - 2103, Nov. 2013.

[29] J. Hoydis, M. Kobayashi, and M. Debbah, "On the optimal number of cooperative base stations in network MIMO systems," 2010, preprint: arXiv:1003.0332v1.

[30] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824-2833, Jun. 2011.

[31] K. Huang, and J.G. Andrews, "An analytical framework for multi-cell cooperation via stochastic geometry and

large deviations," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2501-2516, Apr. 2013.

[32] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Annals of Probability*, vol. 2, no. 5, pp. 805-814, 1974.

[33] J. Li, X. Wu, and R. Laroia, "*OFDMA mobile broadband communications: A systems approach*," Cambridge University Press, 2013

[34] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," in Proc. *Annual Allerton Conference on Communication, Control, and Computing*, pp. 1166-1173, Oct. 2013.

[35] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528-541, Mar. 2006.

[36] 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.

[37] H. Poor, "*An introduction to signal detection and estimation.*" New York: Springer-Verlag, 1994.

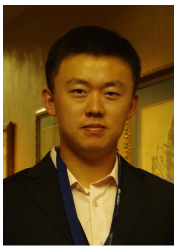
[38] C. Au-yeung and D. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458-462, Feb. 2007.

[39] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 59, no. 8, pp. 2228-2240, Aug. 2011.

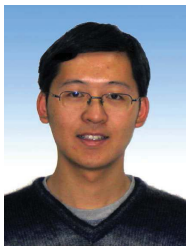
[40] J. Xu, J. Zhang, and J. G. Andrews, "When does the Wyner model accurately describe an uplink cellular network?" in Proc. *IEEE GLOBECOM*, Miami, FL, Dec. 2010.

[41] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information-theoretic statistics of gram random matrices with a given variance profile," *Ann. Appl. Probab.*, vol. 18, no. 6, pp. 2071-2130, 2008.

[42] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506-522, Feb. 2005.



Zhiyuan Jiang (S'12) was born at Beipiao, Liaoning Province, China in 1987. He received his B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China in 2010. Since September 2010, he has been a graduate student in Niulab of Electronic Engineering Department of Tsinghua University, where he is currently pursuing his doctoral studies. His main research interests include multiuser MIMO systems, green wireless networks, queuing theory and Lyapunov optimization.



Sheng Zhou (S'06, M'12) received his B.E. and Ph.D. degrees in Electronic Engineering from Tsinghua University, Beijing, China, in 2005 and 2011, respectively. He is now an assistant professor in Electronic Engineering Department at Tsinghua University, Beijing, China. From January to June 2010, he was a visiting student at Wireless System Lab, Electrical Engineering Department of Stanford University. He is a co-recipient of the Best Paper Award from the Asia-Pacific Conference on Communication (APCC) in 2009 and 2013, the 23th IEEE International Conference on Communication Technology (ICCT) in 2011, and the 25th Intl. Tele-traffic Cong. (ITC) in 2013. His research interests include cross-layer design for multiple antenna systems, cooperative transmission in cellular systems, and green wireless communications.

Zhisheng Niu (M'98-SM'99-F'12) graduated from Beijing Jiaotong University, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. During 1992-94, he worked for Fujitsu Laboratories Ltd., Japan, and in 1994 joined with Tsinghua University, Beijing, China, where he is now a professor at the Department of Electronic Engineering and deputy dean of the School of Information Science and Technology. He is also a guest chair professor of Shandong University, China. His major research interests include queueing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communication and networks.

Dr. Niu has been an active volunteer for various academic societies, including Director for Conference Publications (2010-11) and Director for Asia-Pacific Board (2008-09) of IEEE Communication Society, Membership Development Coordinator (2009-10) of IEEE Region 10, Councilor of IEICE-Japan (2009-11), and council member of Chinese Institute of Electronics (2006-11). He is now a distinguished lecturer (2012-15) and Chair of Emerging Technology Committee (2014-15) of IEEE Communication Society, a distinguished lecturer (2014-16) of IEEE Vehicular Technologies Society, a member of the Fellow Nomination Committee of IEICE Communication Society (2013-14), standing committee member of Chinese Institute of Communications (CIC, 2012-16), and associate editor-in-chief of IEEE/CIC joint publication China Communications.

Dr. Niu received the Outstanding Young Researcher Award from Natural Science Foundation of China in 2009 and the Best Paper Award from IEEE Communication Society Asia-Pacific Board in 2013. He also co-received the Best Paper Awards from the 13th, 15th and 19th Asia-Pacific Conference on Communication (APCC) in 2007, 2009, and 2013, respectively, International Conference on Wireless Communications and Signal Processing (WCSP'13), and the Best Student Paper Award from the 25th International Teletraffic Congress (ITC25). He is now the Chief Scientist of the National Basic Research Program (so called "973 Project") of China on "Fundamental Research on the Energy and Resource Optimized Hyper-Cellular Mobile Communication System" (2012-2016), which is the first national project on green communications in China. He is a fellow of both IEEE and IEICE.