Pilot Design and Allocation for Multi-Cell Massive MIMO Systems with Two-Tier Receiving

Anzhong Hu, *Member, IEEE*

***Abstract*—This paper investigates pilot design and allocation in multi-cell massive multiple-input multiple-output systems with two-tier receiving. As the length of pilot is limited, the pilots cannot be orthogonal and cause interference. The sum rate as well as the signal-to-interference-plus-noise ratio (SINR) of the system are firstly analyzed. The sum rate upper bound and the asymptotic SINR are derived. The analysis shows that the pilot correlation and the spatial correlation jointly affect the sum rate. Based on this observation, the discrete Fourier transform vectors are employed in designing pilots with various correlation values. Then, with the channel correlation information and the designed pilots, the pilots are allocated to the cells in a greedy way with the object of maximizing the sum rate upper bound. Moreover, the sum rate analysis, the pilot design, and the pilot allocation are extended to the scenario with grouped correlations. Simulation results show that the proposed pilot design and allocation performs better than other pilots and the derived upper bound is effective.**

***Index Terms*—Massive MIMO systems, sum rate, pilot, correlated channel.**

I. Introduction

As wireless communication technologies evolve toward the next generation, the massive multiple-input multiple-output (MIMO) is an important technology in improving the spectral efficiency [1]. The large number of antennas equipped at the base station (BS) in massive MIMO systems can provide higher multiplexing gain and diversity gain than traditional MIMO systems.

Among linear processing methods in MIMO systems such as zero forcing and minimum mean squared error (MMSE), the two-tier processing, sometimes named as the two-layer processing, has been investigated in [2]-[5]. In [2]-[4], two-tier precoding is investigated, where the outer precoding mitigates inter-cell or inter-group interference, and the inner precoding mitigates intra-cell or intra-group interference. Because the outer precoding is able to reduce the effective channel dimension, the computational complexity for signal processing at the BS is greatly reduced. Since the outer precoding can be employed with analog phase shifts network, the number of radio frequency (RF) chains can also be reduced. In [5], two-tier receiving is investigated, where the outer receiving mitigates intra-cell interference and the inner receiving mitigates inter-cell interference. These two-tier processing methods are different. The method in [2] processes in a single-cell scenario with one BS. In [3] and [4], each BS processes individually with perfect instantaneous channel state information (CSI). In [5], the signals are processed at each BS in the outer processing, but at a central unit in the inner processing.

With two-tier processing, the uplink pilots are necessary for uplink equivalent channel estimation. When the pilots are not orthogonal, the channel estimate suffers from inter-user interference and this will deteriorate the sum rate performance. For multi-cell systems, a pilot design that tries to maximize the minimal angle among the pilot vectors is proposed in [6] and [7]. In [8], pilot positions are optimized for improving channel estimation in single-cell systems with selective channels. In [9], pilots are designed within frequency-domain and arranged to user groups for keeping the pilots orthogonal in multi-cell massive MIMO systems. In [10], pilots are designed by combining orthogonal sequences and optimizing the combining coefficients for maximizing the minimal capacity in multi-cell massive MIMO systems. In [11], pilots are successively designed in each cell for minimizing the channel estimation error in multi-cell massive MIMO systems. In [12], the phase shifts of Zadoff-Chu (ZC) sequences are designed to minimize the sum of the pilot correlation norms in multi-cell massive MIMO systems. In [13], the authors utilize a similar criterion as [12] and design both phase shifts and the integer parameters of ZC sequences. The authors investigate pilot design for distributed massive MIMO systems and use deep learning to optimize the power allocation of the pilots basis in [14]. In [15], the channel estimation error in multi-cell massive MIMO systems is analyzed and is minimized by successively designing pilots in each BS with convex optimization. The authors analyze the capacity for multi-cell massive MIMO systems and design pilots to achieve capacity bound in [16]. However, these pilots are not designed for massive MIMO systems with two-tier processing.

Despite pilot design, pilot allocation has also been proposed to mitigate inter-cell interference. In [17], the pilots are allocated to minimize the channel estimation error, which is calculated with the channel covariances of the MSs. In [18], a greedy pilot allocation approach based on the large scale fading coefficients is proposed, in which the pilots are allocated in a sequential way and the MSs with worse channels are allocated with better pilots. In [19], an interference graph is constructed based on the large scale fading coefficients, and a

This research was supported by Project LY20F010007 supported by Zhejiang Provincial Natural Science Foundation of China, Project 61601152 and Project U1609216 supported by National Natural Science Foundation of China, and China Scholarship Council. The associate editor coordinating the review of this paper and approving it for publication was L. Dai. *(Corresponding author: Anzhong Hu).*

A. Hu is with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China, email: huaz@hdu.edu.cn.

graph coloring algorithm is proposed to mitigate inter-cell interference. In [20], the pilot allocation problem is constructed as a coalitional game, and a distributed algorithm in which each cell tries to gain pilots based on the ergodic capacity is proposed. In [21], the pilot allocation problem is also modeled as a game with the payoff function calculated with instantaneous channels. In [22], the pilots are allocated with location information for multi-cell massive MIMO systems with Rician fading. In [23], the spectral efficiency of multi-cell massive MIMO systems is analyzed and the pilots are allocated with single-branch searching for maximizing the spectral efficiency. However, the pilots are not allocated for massive MIMO systems with two-tier processing. Apart from pilot design and allocation, power control was also investigated for sum rate maximization in massive MIMO systems. In [24] and [25], deep learning is employed to control uplink transmission power for maximizing the sum rate.

Motivated by the superiority of two-tier processing in [2]-[5], in this paper, the sum rate of multi-cell massive MIMO systems with two-tier receiving is investigated. With MMSE employed in channel estimation and two-tier receiving, the signal-to-interference-plus-noise ratio (SINR) is first derived, and then the upper bound on the sum rate and the asymptotic SINR are analyzed. Based on the analysis, it is found that the spatial correlation and the pilot correlation jointly affect the sum rate upper bound. Then, motivated by the fact that the orthogonal sequences are useful in pilot design in [9], [10], [12], [13], the discrete Fourier transform (DFT) vectors with various shifts are designed. With the object of maximizing the sum rate upper bound, and motivated by the effect of pilot allocation in the existing papers, the pilots are allocated in a greedy way. Finally, the sum rate analysis is extended to the grouped correlation scenario. The pilot design and the pilot allocation are also modified for this scenario. The main contributions of this paper are three-fold.

1) The sum rate of the massive MIMO system with MMSE channel estimation and two-tier receiving with MMSE principle is analyzed. The analysis reveals that the pilot design should take the spatial correlation into consideration. The two-tier processing here employs outer processing and inner processing as [2]-[5]. However, the two-tier processing here is different from [5] in that it is completed at each BS without information sharing among the BSs. Moreover, the analysis here is different from [2]-[4] in that the multi-cell scenario with channel estimation error is considered here.
2) The pilots based on DFT vectors are designed, which can achieve various correlations. In the proposed pilot design, redundant pilots are designed for choosing, while the pilots are directly designed in the existing papers. Moreover, a greedy pilot allocation approach with the object of maximizing the sum rate upper bound is proposed. As only the channel correlation is needed, which is statistical, the proposed pilot allocation facilitates the system deployment. The greedy pilot allocation proposed is based on both the spatial correlation and the derived sum rate upper bound, which is different from the existing papers.
3) The sum rate analysis, the pilot design, and the pilot allocation are extended to the scenario with grouped correlations.

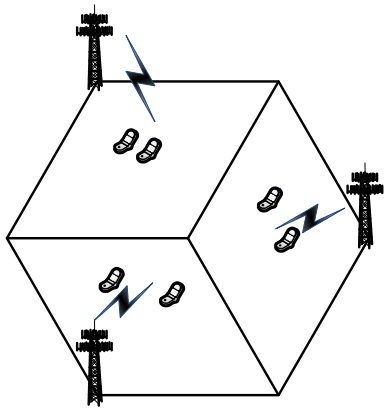

Fig. 1. A cellular layout example with $L = 3, K = 2$.

This paper is organized as follows. In Section II, the system model and the assumptions are given. Section III presents the analysis of the sum rate. In Section IV, the pilot is designed and allocated. In Section V, the investigation is extended to the grouped correlation scenario. Section VI gives the simulation parameters and the numerical results. Finally, conclusions are drawn in Section VII.

Notations: Lower-case (upper-case) boldface symbols denote vectors (matrices); $\mathbf{I}_K$ represents the $K \times K$ identity matrix; $\mathbf{0}_K$ represents the $K \times 1$ zero vector and $\mathbf{0}_{K\times L}$ represents the $K \times L$ zero matrix; $(\cdot)^H$ and $\mathbb{E}\{\cdot\}$ denote the conjugate transpose and the expectation, respectively; $\circ$ is the Hadamard product; $[\cdot]_j$ is the j-th element of a vector or the j-th column of a matrix; $[\cdot]_{j,k}$ is the element in the j-th row and k-th column of a matrix; $\mathbf{A}_{(-n)}$ is the matrix with the n-th row of $\mathbf{A}$ removed, $\mathbf{A}_{(-n,-n)}$ is the matrix with the n-th row and the n-th column of $\mathbf{A}$ removed; $\mathrm{tr}\{\cdot\}$ is the trace of a matrix; $|\cdot|$ is the absolute value of a variable, or the determinant of a matrix, or the cardinality of a set; i is the imaginary unit; $\delta(\cdot)$ is the delta function; $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\mathbf{z}$ is a complex circularly symmetric Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\exp(\cdot)$ is the exponential function.

II. System Model

Consider a massive MIMO cellular system which consists of L cells. In each cell, there is one BS and K MSs. The BS is equipped with M antennas and each mobile station (MS) has one antenna. An example of the cellular layout is shown in Fig. 1. We focus on the channel estimation and the uplink transmission. In the channel estimation stage, all the MSs transmit pilots, and the received pilots at the j-th BS are written as

$$\mathbf{Y}_j^{\mathrm{p}} = \sum_{l=1}^{L} \mathbf{H}_{jl}\mathbf{X}_l^{\mathrm{p}} + \mathbf{Z}_j^{\mathrm{p}} \in \mathbb{C}^{M\times L_{\mathrm{p}}}, \tag{1}$$

where $\mathbf{H}_{jl} \in \mathbb{C}^{M\times K}$ is the channel matrix from the MSs in the l-th cell to the j-th BS, $\mathbf{X}_l^{\mathrm{p}} \in \mathbb{C}^{K\times L_{\mathrm{p}}}$ is the pilot matrix of the l-th cell, and $\mathbf{Z}_j^{\mathrm{p}} \in \mathbb{C}^{M\times L_{\mathrm{p}}}$ is the uncorrelated noise matrix with $[\mathbf{Z}_j^{\mathrm{p}}]_n \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M), \forall n$. The BS processes the

pilots with beamforming and pilot correlation and this results into

$$\tilde{\mathbf{Y}}_j^{\mathrm{p}} = \mathbf{B}_j^H \mathbf{Y}_j^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} \in \mathbb{C}^{K\times K}, \tag{2}$$

where $\mathbf{B}_j \in \mathbb{C}^{M\times K}$ is the beamforming matrix for the j-th cell. Substituting (1) into (2) results into

$$\tilde{\mathbf{Y}}_j^{\mathrm{p}} = \sum_{l=1}^{L} \mathbf{B}_j^H \mathbf{H}_{jl} \mathbf{X}_l^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} + \mathbf{B}_j^H \mathbf{Z}_j^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H}. \tag{3}$$

In the uplink transmission stage, the MSs send data symbols, and the received signal at the BS is expressed as

$$\mathbf{y}_j^{\mathrm{d}} = \sum_{l=1}^{L} \mathbf{H}_{jl} \mathbf{x}_l^{\mathrm{d}} + \mathbf{z}_j^{\mathrm{d}} \in \mathbb{C}^{M\times 1}, \tag{4}$$

where $\mathbf{x}_l^{\mathrm{d}} \in \mathbb{C}^{K\times 1}$ is the transmitted data vector of the l-th cell, $\mathbf{z}_j^{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ is the noise vector. The BS processes the received signal with beamforming and this yields

$$\tilde{\mathbf{y}}_j^{\mathrm{d}} = \mathbf{B}_j^H \mathbf{y}_j^{\mathrm{d}} \in \mathbb{C}^{K\times 1}. \tag{5}$$

Substituting (4) into (5) results into

$$\tilde{\mathbf{y}}_j^{\mathrm{d}} = \sum_{l=1}^{L} \mathbf{B}_j^H \mathbf{H}_{jl} \mathbf{x}_l^{\mathrm{d}} + \mathbf{B}_j^H \mathbf{z}_j^{\mathrm{d}}. \tag{6}$$

A. System Assumptions

The useful assumptions and the corresponding preliminary results are listed as follows.

1)The channels are correlated with $[\mathbf{H}_{jl}]_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{W}_{jlk})$, where $\mathbf{W}_{jlk} = \mathbf{W}_{jl}, \forall k$ is assumed and

$$\mathbf{W}_{jl} = \mathbf{U}_{jl}\mathbf{D}_{jl}\mathbf{U}_{jl}^H + \bar{\mathbf{U}}_{jl}\bar{\mathbf{D}}_{jl}\bar{\mathbf{U}}_{jl}^H \in \mathbb{C}^{M\times M} \tag{7}$$

is positive semi-definite, $\mathbf{U}_{jl} \in \mathbb{C}^{M\times K}$ is composed of the partial eigenvectors of $\mathbf{W}_{jl}$, $\mathbf{D}_{jl} \in \mathbb{R}^{K\times K}$ is a matrix of the K largest eigenvalues of $\mathbf{W}_{jl}$; $\bar{\mathbf{U}}_{jl} \in \mathbb{C}^{M\times(M-K)}$ is composed of the remaining eigenvectors of $\mathbf{W}_{jl}$, $\bar{\mathbf{D}}_{jl} \in \mathbb{R}^{(M-K)\times(M-K)}$ is a matrix of the corresponding eigenvalues of $\mathbf{W}_{jl}$. Thus, the channel matrix can be written as

$$\mathbf{H}_{jl} = \mathbf{W}_{jl}^{\frac{1}{2}} \tilde{\mathbf{H}}_{jl}, \tag{8}$$

where $\tilde{\mathbf{H}}_{jl} \in \mathbb{C}^{M\times K}$ and $[\tilde{\mathbf{H}}_{jl}]_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. Note that this kind of channel model corresponds to short-range rich scattering transmission environments, in which the distributions of the path angles are similar, such as the urban micro cell with street canyon scenario. The case with $\mathbf{W}_{jlk}, \forall k$ being the same for groups and are different between groups will be considered in Section V.

2)The BS uses $\mathbf{B}_j = \mathbf{U}_{jj}$ as the beamforming matrix, cf. (2) and (5). Since the channel correlation $\mathbf{W}_{jj}$ changes slowly in comparison to the instantaneous channel $\tilde{\mathbf{H}}_{jj}$, the MSs can transmit dedicated pilots for the BSs to estimate the channel correlations with affordable resource consumption. Thus, it is assumed that the channel correlation $\mathbf{W}_{jj}$ is known to the BS. This kind of assumption and the corresponding reason are also discussed in [3] and [17]. Note this beamforming matrix is used for the reason that the channel subspace correlations, i.e., $||\mathbf{U}_{jj}^H \mathbf{U}_{jl}||, \forall j \neq l$, may be low. We denote

$$\bar{\mathbf{H}}_{jl} \triangleq \mathbf{B}_j^H \mathbf{H}_{jl} = \mathbf{U}_{jj}^H \mathbf{H}_{jl} \in \mathbb{C}^{K\times K} \tag{9}$$

as the beamspace channel matrix. Substituting (7) and (8) into (9) yields

$$\bar{\mathbf{H}}_{jj} = \mathbf{D}_{jj}^{\frac{1}{2}} \mathbf{U}_{jj}^H \tilde{\mathbf{H}}_{jj}. \tag{10}$$

Substituting (9) into (3) results into

$$\tilde{\mathbf{Y}}_j^{\mathrm{p}} = \sum_{l=1}^{L} \bar{\mathbf{H}}_{jl} \mathbf{X}_l^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} + \mathbf{B}_j^H \mathbf{Z}_j^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} = \sum_{l=1}^{L} \bar{\mathbf{H}}_{jl} \mathbf{C}_{jl} + \tilde{\mathbf{Z}}_j^{\mathrm{p}}, \tag{11}$$

where $\mathbf{C}_{jl} = \mathbf{X}_l^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} \in \mathbb{C}^{K\times K}$ is the pilot correlation matrix and

$$\tilde{\mathbf{Z}}_j^{\mathrm{p}} = \mathbf{B}_j^H \mathbf{Z}_j^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} = \mathbf{U}_{jj}^H \mathbf{Z}_j^{\mathrm{p}} \mathbf{X}_j^{\mathrm{p}H} \in \mathbb{C}^{K\times K}. \tag{12}$$

Similarly, substituting (9) into (6), the received signal can be re-written as

$$\tilde{\mathbf{y}}_j^{\mathrm{d}} = \sum_{l=1}^{L} \bar{\mathbf{H}}_{jl} \mathbf{x}_l^{\mathrm{d}} + \tilde{\mathbf{z}}_j^{\mathrm{d}}, \tag{13}$$

where $\tilde{\mathbf{z}}_j^{\mathrm{d}} = \mathbf{B}_j^H \mathbf{z}_j^{\mathrm{d}} = \mathbf{U}_{jj}^H \mathbf{z}_j^{\mathrm{d}} \in \mathbb{C}^{K\times 1}$.

3)The pilot length is insufficient with $K \leq L_{\mathrm{p}} < 2K$, which means the pilots cannot be orthogonal for any two cells, i.e., $\sum_{k=1}^{K} ||[\mathbf{C}_{jl}]_k|| > 0, \forall j \neq l$. Moreover, pilot vectors in the same cell are orthogonal and of the same power, i.e.,

$$\mathbf{C}_{jj} = \rho \mathbf{I}_K. \tag{14}$$

This means that the pilot transmission powers are the same. Similarly, the uplink data transmission power is ρ, i.e., $\mathbb{E}\{\mathbf{x}_l^{\mathrm{d}} \mathbf{x}_l^{\mathrm{d}H}\} = \rho \mathbf{I}_K$, where the uplink data vector $\mathbf{x}_l^{\mathrm{d}}$ is defined below (4).

4) The BS employs MMSE to estimate the beamspace channels. For the k-th MS in the j-th cell, the corresponding received pilot vector, cf. the received pilot matrix in (11), is defined as

$$\mathbf{y}_{jk}^{\mathrm{p}} = \frac{1}{\rho}[\tilde{\mathbf{Y}}_j^{\mathrm{p}}]_k = \bar{\mathbf{h}}_{jk} + \sum_{l\neq j} \frac{1}{\rho} \bar{\mathbf{H}}_{jl} [\mathbf{C}_{jl}]_k + \frac{1}{\rho}[\tilde{\mathbf{Z}}_j^{\mathrm{p}}]_k \in \mathbb{C}^{K\times 1}, \tag{15}$$

where (14) is employed in deriving the second equation,

$$\bar{\mathbf{h}}_{jk} = [\bar{\mathbf{H}}_{jj}]_k \in \mathbb{C}^{K\times 1} \tag{16}$$

is the beamspace channel vector of the k-th MS in the j-th cell. The MMSE estimate of $\bar{\mathbf{h}}_{jk}$ is [26][27]

$$\hat{\bar{\mathbf{h}}}_{jk} = \mathbf{\Sigma}_{jk} \mathbf{y}_{jk}^{\mathrm{p}} \in \mathbb{C}^{K\times 1}, \tag{17}$$

where $\mathbf{\Sigma}_{jk} \in \mathbb{C}^{K\times K}$ is denoted as

$$\mathbf{\Sigma}_{jk} = \mathbf{D}_{jj} \left(\sum_{l=1}^{L} \mathbf{U}_{jj}^H \mathbf{W}_{jl} \mathbf{U}_{jj} \frac{||[\mathbf{C}_{jl}]_k||^2}{\rho^2} + \frac{1}{\rho} \mathbf{I}_K \right)^{-1}. \tag{18}$$

Then, the beamspace channel estimation error is denoted as

$$\boldsymbol{\xi}_{jk} = \bar{\mathbf{h}}_{jk} - \hat{\bar{\mathbf{h}}}_{jk} \in \mathbb{C}^{K\times 1}, \tag{19}$$

and the MMSE estimate of $\bar{\mathbf{H}}_j$ is denoted as

$$\hat{\bar{\mathbf{H}}}_j = [\hat{\bar{\mathbf{h}}}_{j1}, \hat{\bar{\mathbf{h}}}_{j2}, \cdots, \hat{\bar{\mathbf{h}}}_{jK}] \in \mathbb{C}^{K\times K}. \tag{20}$$

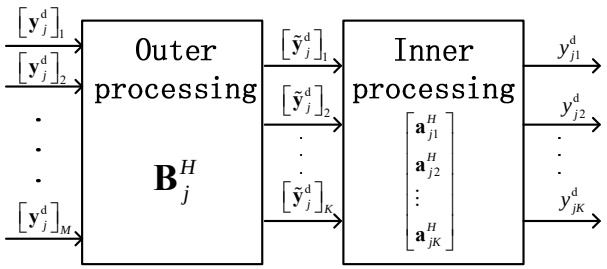

Fig. 2. An illustration of the two-tier processing.

5)The BS employs the MMSE receiving vector given in [26]. When MMSE receiving of (13) is employed, the resulting data symbol for the k-th MS in the j-th cell is denoted as

$$y_{jk}^{\mathrm{d}} = \mathbf{a}_{jk}^H \tilde{\mathbf{y}}_j^{\mathrm{d}}, \tag{21}$$

where

$$\mathbf{a}_{jk} = \tilde{\boldsymbol{\Xi}}_{jk}^{-1} \hat{\mathbf{h}}_{jk} \in \mathbb{C}^{K\times 1} \tag{22}$$

is the MMSE receiving vector, and

$$\tilde{\boldsymbol{\Xi}}_{jk} = \sum_{n\neq k} \hat{\mathbf{h}}_{jn}\hat{\mathbf{h}}_{jn}^H + \frac{1}{\rho}\boldsymbol{\Xi}_j \in \mathbb{C}^{K\times K}, \tag{23}$$

$$\boldsymbol{\Xi}_j = \mathbb{E}\{\tilde{\boldsymbol{\xi}}_j \tilde{\boldsymbol{\xi}}_j^H\} \in \mathbb{C}^{K\times K}, \tag{24}$$

$$\tilde{\boldsymbol{\xi}}_j = \sum_{k=1}^{K} \boldsymbol{\xi}_{jk}[\mathbf{x}_j^{\mathrm{d}}]_k + \sum_{l\neq j} \bar{\mathbf{H}}_{jl}\mathbf{x}_l^{\mathrm{d}} + \tilde{\mathbf{z}}_j^{\mathrm{d}} \in \mathbb{C}^{K\times 1}. \tag{25}$$

Note that the BS processes the received signal with beamforming, cf. (5), and MMSE detection, cf. (21). The beamforming in (5) is termed as the outer processing, which tackles the inter-cell interference. The MMSE detection in (21) is termed as the inner processing, which tackles the intra-cell interference. This kind of receiving method is named two-tier receiving here. An illustration of this two-tier processing is shown in Fig. 2. As the outer processing can be employed with analog circuits and the signal dimension for the inner processing is much smaller than the received signal dimension, the number of RF chains and the computational complexity can be reduced. The two-tier processing here is similar to other two-tier processing methods in [2]-[5]. However, there are some differences. The two-tier processing in [2] is analyzed for the single-cell scenario, but the two-tier processing here is analyzed for the multi-cell scenario. The two-tier processing in [3] and [4] is analyzed with perfect CSI, but channel estimation error is considered here. The two-tier processing in [5] requires information sharing among the BSs, which is not necessary in the two-tier processing here.

B. Problem Formulation

Substituting (13) into (21) yields

$$\begin{aligned} y_{jk}^{\mathrm{d}} &= \sum_{l=1}^{L} \mathbf{a}_{jk}^H \bar{\mathbf{H}}_{jl}\mathbf{x}_l^{\mathrm{d}} + \mathbf{a}_{jk}^H \tilde{\mathbf{z}}_j^{\mathrm{d}} \\ &= \mathbf{a}_{jk}^H \hat{\mathbf{h}}_{jk}[\mathbf{x}_j^{\mathrm{d}}]_k + \mathbf{a}_{jk}^H \sum_{n\neq k} \hat{\mathbf{h}}_{jn}[\mathbf{x}_j^{\mathrm{d}}]_n + \mathbf{a}_{jk}^H \tilde{\boldsymbol{\xi}}_j, \end{aligned}$$

where (16) and (19) are used to obtain the second equation. As can be seen, the SINR of the k-th MS in the j-th cell is

$$\begin{aligned} \acute{\gamma}_{jk} &= \frac{\rho|\mathbf{a}_{jk}^H \hat{\mathbf{h}}_{jk}|^2}{\rho\sum_{n\neq k}|\mathbf{a}_{jk}^H\hat{\mathbf{h}}_{jn}|^2 + \mathbf{a}_{jk}^H\mathbb{E}\{\tilde{\boldsymbol{\xi}}_j\tilde{\boldsymbol{\xi}}_j^H|\boldsymbol{\xi}_{jk},\bar{\mathbf{H}}_{jl}\}\mathbf{a}_{jk}} \\ &\approx \frac{\rho|\mathbf{a}_{jk}^H \hat{\mathbf{h}}_{jk}|^2}{\rho\sum_{n\neq k}|\mathbf{a}_{jk}^H\hat{\mathbf{h}}_{jn}|^2 + \mathbf{a}_{jk}^H\boldsymbol{\Xi}_j\mathbf{a}_{jk}} \\ &= \frac{|\mathbf{a}_{jk}^H \hat{\mathbf{h}}_{jk}|^2}{\mathbf{a}_{jk}^H\tilde{\boldsymbol{\Xi}}_{jk}\mathbf{a}_{jk}} \triangleq \gamma_{jk}, \end{aligned} \tag{26}$$

(27)

where (24) is used to derive the approximation, (23) is used in deriving the third equation, and γ_{jk} is the approximated SINR. Then, the ergodic sum rate of the system is approximated as

$$R = \sum_{j=1}^{L}\sum_{k=1}^{K} \mathbb{E}\{\log_2(1+\gamma_{jk})\} \tag{28}$$

$$\leq \sum_{j=1}^{L}\sum_{k=1}^{K} \log_2(1+\mathbb{E}\{\gamma_{jk}\}) \triangleq \bar{R}, \tag{29}$$

where the inequality is based on the Jensen's inequality. According to (27), we know that part of the signal is treated as noise in the SINR. Moreover, the expectation in (29) makes the channel estimate "forgotten" in the signal detection, which means the bound in (29) belongs to the use-and-then-forget bound in [28]. The problem investigated is how to design and allocate the pilot sequences to maximize the sum rate upper bound, i.e., to maximize the expectation of the SINR, and is expressed as

$$\max_{\mathbf{x}_j^{\mathrm{p}}, j=1,2,\cdots,L} \bar{R} \ \text{ s.t. } (14). \tag{30}$$

With the system model, the sum rate will be analyzed in the following section.

III. Sum Rate Analysis

In this section, the SINR is analyzed. First, the SINR expression is simplified. Then, the expectation of the matrix that determines the SINR is analyzed. Finally, the upper bound on the expectation of the SINR is derived, which is related to the expectation of the matrix that determines the SINR. Moreover, the asymptotic results of the SINR are also derived to show the factors that influence the SINR.

First, the SINR can be simplified, and is shown in Proposition 1.

Proposition 1. *The SINR can be written as*

$$\gamma_{jk} = \frac{1}{[(\mathbf{I}_K + \mathbf{V}_j)^{-1}]_{k,k}} - 1, \tag{31}$$

where

$$\mathbf{V}_j = \rho\hat{\mathbf{H}}_j^H \boldsymbol{\Xi}_j^{-1}\hat{\mathbf{H}}_j \in \mathbb{C}^{K\times K}. \tag{32}$$

Proof: Refer to Appendix A. ■

Then, the expectation of the matrix $\mathbf{V}_j$ is shown in Proposition 2.

Proposition 2. *Define the expectation of* $\mathbf{V}_j$ *as*

$$\mathbb{E}\{\mathbf{V}_j\} \triangleq \mathbf{R}_j, \tag{33}$$

we have its expression as

$$\mathbf{R}_j = \sum_{l\neq j} \mathbf{X}_{jl} \circ \mathbf{C}_{jl}^H \mathbf{C}_{jl} + \mathbf{Y}_j, \tag{34}$$

where $\mathbf{X}_{jl} \in \mathbb{C}^{K\times K}$, $\mathbf{Y}_j \in \mathbb{C}^{K\times K}$,

$$[\mathbf{X}_{jl}]_{s,t} = \frac{1}{\rho}\mathrm{tr}\{\mathbf{U}_{jj}\mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1}\mathbf{\Sigma}_{jt}\mathbf{U}_{jj}^H \mathbf{W}_{jl}\}, \tag{35}$$

$$\begin{aligned}[\mathbf{Y}_j]_{s,t} &= \delta(s-t)\rho\mathrm{tr}\{\mathbf{D}_{jj}\mathbf{\Sigma}_{js}^H\mathbf{\Xi}_j^{-1}\mathbf{\Sigma}_{js}\} \\ &+\delta(s-t)\mathrm{tr}\{\mathbf{\Sigma}_{js}^H\mathbf{\Xi}_j^{-1}\mathbf{\Sigma}_{js}\},\end{aligned} \tag{36}$$

and

$$\mathbf{\Xi}_j = \rho\mathbf{D}_{jj}(K\mathbf{I}_K - \sum_{k=1}^{K}\mathbf{\Sigma}_{jk}^H) + \rho K\sum_{l\neq j}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj} + \mathbf{I}_K. \tag{37}$$

Proof: Refer to Appendix B. ∎

Based on Proposition 1 and Proposition 2, the bound on the expectation of the SINR can be derived, as shown in Lemma 1.

Lemma 1. *The expectation of the SINR is upper bounded as*

$$\begin{aligned}\mathbb{E}\{\gamma_{jk}\} \leq \bar{\gamma}_{jk} &= [\mathbf{R}_j]_{k,k} \\ &-\frac{\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{K[\mathbf{R}_j^{-1}]_{k,k}(1+\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\})}.\end{aligned} \tag{38}$$

Proof: Refer to Appendix C. ∎

Remark 1. *According to the sum rate upper bound in (38) and the definition of* $\mathbf{R}_j$ *in Proposition 2, it can be seen that the pilot correlation* $\mathbf{C}_{jl}$ *and the space correlation* $\mathbf{U}_{jj}^H\mathbf{W}_{jl}$ *jointly impact on the sum rate upper bound.*

By substituting the upper bound in (38) to the upper bound in (29), another upper bound on the sum rate is derived, which is

$$\tilde{R} \triangleq \sum_{j=1}^{L}\sum_{k=1}^{K}\log_2(1+\bar{\gamma}_{jk}), \tag{39}$$

and the pilot design problem in (30) is changed into

$$\max_{\mathbf{X}_j^{\mathrm{p}}} \tilde{R} \ \text{ s.t. } (14). \tag{40}$$

Based on Proposition 1 and Proposition 2, the limit of the SINR can also be derived, as shown in Lemma 2.

Lemma 2. *In the asymptotic region, we have*

$$\lim_{\substack{\mathbf{C}_{jl}/\rho\to\mathbf{0}_{K\times K},\forall l\neq j \\ \rho\to\infty}} \gamma_{jk} = \frac{1}{[(\mathbf{I}_K+\bar{\mathbf{V}}_j)^{-1}]_{k,k}} - 1, \tag{41}$$

where

$$\bar{\mathbf{V}}_j \to \bar{\mathbf{H}}_{jj}^H(K\sum_{l\neq j}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj})^{-1}\bar{\mathbf{H}}_{jj}.$$

Moreover, we have

$$\lim_{\mathbf{U}_{jj}^H\mathbf{W}_{jl}\to\mathbf{0}_{K\times M},\forall l\neq j} \gamma_{jk} = \frac{1}{[(\mathbf{I}_K+\bar{\mathbf{V}}_j^{\mathrm{a}})^{-1}]_{k,k}} - 1, \tag{42}$$

where $\bar{\mathbf{H}}_j^{\mathrm{a}} \in \mathbb{C}^{K\times K}$, $\mathbf{\Xi}_j^{\mathrm{a}} \in \mathbb{C}^{K\times K}$, $\bar{\mathbf{V}}_j^{\mathrm{a}} \in \mathbb{C}^{K\times K}$, *and are defined as*

$$\begin{aligned}\bar{\mathbf{H}}_j^{\mathrm{a}} &= \mathbf{D}_{jj}\left(\mathbf{D}_{jj}+\frac{1}{\rho}\mathbf{I}_K\right)^{-1}(\bar{\mathbf{H}}_{jj}+\frac{1}{\rho}\tilde{\mathbf{Z}}_j^{\mathrm{p}}), \\ \mathbf{\Xi}_j^{\mathrm{a}} &= \rho K\mathbf{D}_{jj}\left(\rho\mathbf{D}_{jj}+\mathbf{I}_K\right)^{-1}+\mathbf{I}_K, \\ \bar{\mathbf{V}}_j^{\mathrm{a}} &= \rho(\bar{\mathbf{H}}_j^{\mathrm{a}})^H(\mathbf{\Xi}_j^{\mathrm{a}})^{-1}\bar{\mathbf{H}}_j^{\mathrm{a}}.\end{aligned}$$

Proof: Refer to Appendix D. ∎

Remark 2. *The results shows that even the pilots are orthogonal, the inter-cell interference cannot be completely eliminated for the channel subspace correlation. However, when the channel subspace are orthogonal, the inter-cell interference can be completely eliminated no matter what value the pilot correlation is. Thus, in general conditions, the channel subspace correlation and the pilot correlation jointly impact on the SINR. Moreover, the two-tier receiving can lessen the requirement on the pilot correlations in low channel subspace correlation conditions.*

Also, a corollary can be derived with Lemma 2, which is shown as follows.

Corollary 1. *When the BS is equipped with a linear array and the direction-of-arrival (DOA) region of MSs in one cell is non-overlapping with that in any other cell, we have*

$$\lim_{M\to\infty}\gamma_{jk} = \frac{1}{[(\mathbf{I}_K+\bar{\mathbf{V}}_j^{\mathrm{a}})^{-1}]_{k,k}} - 1. \tag{43}$$

Proof: According to the results in [17], when the BS is equipped with a linear array and the direction-of-arrival region of MSs in one each cell is non-overlapping with that in any other cell, we have $\lim_{M\to\infty}\mathbf{U}_{jj}^H\mathbf{W}_{jl} = \mathbf{0}_{K\times M}, \forall l\neq j$. Then, with (42) in Lemma 2, we have (43). ∎

Remark 3. *This corollary shows that the inter-cell interference tends to be completely eliminated when the number of BS antennas tends to infinity for systems with linear BS array and non-overlapping DOA regions. Thus, the two-tier receiving can lessen the requirement on the pilot correlations when there are a large number of BS antennas. Moreover, this result can be easily extended to the case with a rectangular array at the BS.*

With the derived results on the SINR, we will continue to design and allocate pilots in the next section.

IV. Pilot Design and Allocation

Based on the prior analysis, it can be seen that the pilots should be designed to solve (40), in which the sum rate upper bound depends on $\mathbf{R}_j$. According to the analysis in the last section, the pilot optimization problem should jointly consider the pilot correlation and the space correlation. In this section, we propose to first design the pilot codebook which consists of pilots with various correlations, then we propose a greedy algorithm to allocate pilots with regarding to the space correlation.

A. DFT Vector Based Pilot Design

Since the pilot correlation directly influences the sum rate upper bound, a pilot codebook of various correlations is preferable. As known to all, DFT vectors are orthogonal and the orthogonality is kept when phase shifted. Moreover, DFT vectors of different phase shifts are of different correlations. Thus, we propose to use these vectors to form our pilot codebook. More specifically, the pilot codebook is denoted as

$$\mathcal{X} = \{\mathbf{F}_n, n = 1, 2, \cdots, N\},$$

where N is the size of the whole codebook, $\mathbf{F}_n \in \mathbb{C}^{K \times L_\mathrm{p}}$ is the matrix of DFT vectors, and is defined as

$$[\mathbf{F}_n]_{k,m} = \frac{\sqrt{\rho}}{\sqrt{L_\mathrm{p}}} \exp(i\frac{2\pi}{L_\mathrm{p}}(k + a_n)m),$$
$$m \in \{1, 2, \cdots, L_\mathrm{p}\}, k \in \{1, 2, \cdots, K\},$$

where a_n is the phase shift and is a normal distributed random variable.

Then, we have

$$[\mathbf{F}_n^H]_k^H [\mathbf{F}_{n'}^H]_{k'} = \frac{\rho}{L_\mathrm{p}} \sum_{m=1}^{L_\mathrm{p}} \exp(i\frac{2\pi}{L_\mathrm{p}}(k' - k + a_{n'} - a_n)m).$$

For the case of $n = n'$, we have

$$[\mathbf{F}_n^H]_k^H [\mathbf{F}_n^H]_{k'} = \begin{cases} 0, & \forall k \neq k', \\ \rho, & \forall k = k', \end{cases}$$

which means

$$\mathbf{F}_n \mathbf{F}_n^H = \rho \mathbf{I}_K, \forall n.$$

Moreover, for the correlations with $n \neq n'$, we plot an example here, in which $N = 100$, $K = L_\mathrm{p} = 10$. In Fig. 2, the histogram of $1/\rho[\mathbf{F}_1^H]_1^H[\mathbf{F}_{n'}^H]_{k'}, \forall n' > 1, \forall k'$ is shown. We can see that most of the correlations are small. Thus, by selecting pilot vectors from this codebook in the way that

$$\mathbf{X}_j^\mathrm{p} \in \mathcal{X},$$

the restriction of (14) in (40) can be satisfied, and small pilot correlation values are beneficial for pilot allocation.

With this pilot design, the optimization problem of (40) is changed into a pilot allocation problem as

$$\max_{\mathbf{X}_j^\mathrm{p} \in \mathcal{X}} \tilde{R}. \tag{44}$$

With the pilot design and allocation for two-tier receiving in systems with the same channel correlation for MSs in each cell, we will continue to extend to systems with the same channel correlation for MSs in the same group.

B. Greedy Pilot Allocation

From the pilot allocation problem in (44), we can see that the choice of the pilot sub-codebook index n influences the pilot correlations. Thus, we need to allocation the pilots based on the pilot correlations. As both the spatial correlation and the pilot correlation have impact on the sum rate, we first give a proposition about their relations with the sum rate.

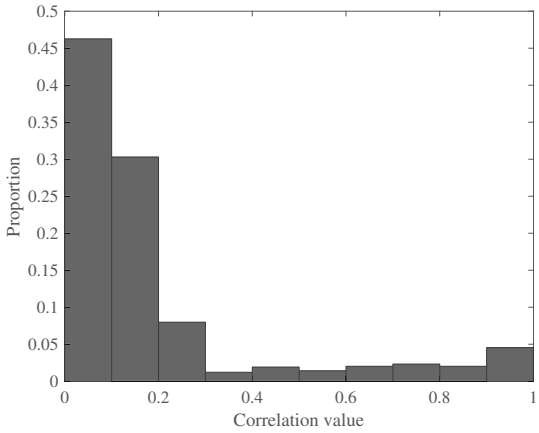

Fig. 3. An example of the proportions of the pilot correlation values $1/\rho[\mathbf{F}_1^H]_1^H[\mathbf{F}_{n'}^H]_{\tilde{m}}, \forall n' > 1, \forall \tilde{m}$ with $N = 100$, $K = L_\mathrm{p} = 10$.

Proposition 3. *When the rank of $\mathbf{W}_{jl}$ is not larger than K, and the spatial correlation tends to zero, i.e. $\mathbf{U}_{jj}^H \mathbf{U}_{jl} \to \mathbf{0}_{K \times K}, \forall j \neq l$, the impacts of pilots on the SINR γ_{jk} and the SINR upper bound $\bar{\gamma}_{jk}$ tend to diminish.*

Proof: Refer to Appendix E. ■

Remark 4. *This proposition shows that in the ideal case of the rank of $\mathbf{W}_{jl}$ being K, pilot design is of less importance when the spatial correlation decreases. Even in the non-ideal case with the rank larger than K, the impact of pilots still depends on the spatial correlations $\mathbf{U}_{jj}^H \mathbf{U}_{jl}, \forall j \neq l$.*

According to this proposition, we propose to allocate pilots in a greedy way that the cells with higher spatial correlation will be given priority in allocating pilots. More specifically, the spatial correlation between the j-th cell and the l-th cell is measured by the spatial correlation factor

$$\eta_{jl} = ||\mathbf{U}_{jj}^H \mathbf{U}_{jl}||^2. \tag{45}$$

The j_0-th cell and the j_1-th cell are allocated pilots first if $\eta_{j_0 j_1} \geq \eta_{jl}, \forall j_0 \neq j, \forall j_1 \neq l$. Then, the j_2-th cell is allocated pilots if $\max\{\eta_{j_2 j_0}, \eta_{j_2 j_1}\} \geq \max\{\eta_{j j_0}, \eta_{j j_1}\}, \forall j \neq j_0, j_1$. This continues until the last cell is allocated with pilots. For the pilot allocation of the j_n-th cell, assume n cells with indices $j_0, j_1, \cdots, j_{n-1}$ have been allocated pilots, the pilot allocation problem in (44) is transformed into

$$\max_{\mathbf{X}_{j_n}^\mathrm{p} \in \mathcal{X}} \check{R}_n, \tag{46}$$

where $\check{R}_n$ is the sum rate upper bound with the pilots of cells other than $j_0, j_1, \cdots, j_{n-1}$ being zeros but with (14) kept. The problem (46) can be easily solved with an exhaustive search. The sketch of the proposed pilot allocation algorithm is as follows.

Algorithm 1 Greedy pilot allocation with descending order of spatial correlations

1: calculate η_{jl} as (45)
2: sort the cells with η_{jl} as $j_0, j_1, \cdots, j_{L-1}$
3: **for** $n = 0 \rightarrow L-1$
4: allocate pilots as (46)
5: **end for**

V. Extension To Grouped Correlation

In this section, we first derive the basic results with general grouping. Then, we discuss the special case with ideal grouping.

A. General Grouping

When the scattering environments vary dramatically inside the cell, such as in urban macro cell scenarios, only MSs close to each other have the same scattering environments. Then, the first assumption in Section II should be modified as follows. The channel correlations satisfy $\mathbf{W}_{jlk} = \tilde{\mathbf{W}}_{jlg}, \forall k \in \mathcal{G}_{lg}$ is assumed, where $\mathcal{G}_{lg}$ is the set of the indices of the MSs in the g-th group in the l-th cell that have the same scattering environments, and there are G groups in each cell. Moreover, we have

$$\tilde{\mathbf{W}}_{jlg} = \mathbf{U}_{jlg}\mathbf{D}_{jlg}\mathbf{U}_{jlg}^H + \bar{\mathbf{U}}_{jlg}\bar{\mathbf{D}}_{jlg}\bar{\mathbf{U}}_{jlg}^H \in \mathbb{C}^{M\times M}$$

is positive semi-definite, $\mathbf{U}_{jlg} \in \mathbb{C}^{M\times|\mathcal{G}_{lg}|}$ is composed of the partial eigenvectors of $\tilde{\mathbf{W}}_{jlg}$, $\mathbf{D}_{jlg} \in \mathbb{R}^{|\mathcal{G}_{lg}|\times|\mathcal{G}_{lg}|}$ is a matrix of the $|\mathcal{G}_{lg}|$ largest eigenvalues of $\tilde{\mathbf{W}}_{jlg}$; $\bar{\mathbf{U}}_{jlg} \in \mathbb{C}^{M\times(M-|\mathcal{G}_{lg}|)}$ is composed of the remaining eigenvectors of $\tilde{\mathbf{W}}_{jlg}$, $\bar{\mathbf{D}}_{jlg} \in \mathbb{R}^{(M-|\mathcal{G}_{lg}|)\times(M-|\mathcal{G}_{lg}|)}$ is a matrix of the corresponding eigenvalues of $\tilde{\mathbf{W}}_{jlg}$. Assuming that the channel vectors are sorted by groups in the channel matrix $\mathbf{H}_{jl}$, the channel matrix can be written as

$$\mathbf{H}_{jl} = [\tilde{\mathbf{W}}_{jl1}^{\frac{1}{2}}\tilde{\mathbf{H}}_{jl1}, \cdots, \tilde{\mathbf{W}}_{jlG}^{\frac{1}{2}}\tilde{\mathbf{H}}_{jlG}], \tag{47}$$

where $\tilde{\mathbf{H}}_{jlg} \in \mathbb{C}^{M\times|\mathcal{G}_{lg}|}$ and $[\tilde{\mathbf{H}}_{jlg}]_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$.

Then, the received pilots and data signals can be processed with per-group receiving similar to that in [2]. More specifically, for the g-th group in the j-th cell, the BS processes the pilots as

$$\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}} = \mathbf{U}_{jjg}^H\mathbf{Y}_j^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}, \tag{48}$$

where $\mathbf{U}_{jjg}$ is the beamforming matrix for the g-th group in the j-th cell, $\mathbf{X}_{jg}^{\mathrm{p}} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times L_{\mathrm{P}}}$ is the pilot matrix corresponding to the MSs in this group. Substituting (1) into (48) results into

$$\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}} = \sum_{l=1}^{L}\mathbf{U}_{jjg}^H\mathbf{X}_l^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H} + \mathbf{U}_{jjg}^H\mathbf{Z}_j^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H}. \tag{49}$$

Substituting (47) into (49) yields

$$\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}} = \sum_{l=1}^{L}\sum_{g'=1}^{G}\bar{\mathbf{H}}_{jlgg'}\mathbf{C}_{jlgg'} + \mathbf{U}_{jjg}^H\mathbf{Z}_j^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H}, \tag{50}$$

where

$$\bar{\mathbf{H}}_{jlgg'} = \mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jlg'}^{\frac{1}{2}}\tilde{\mathbf{H}}_{jlg'} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{lg'}|}, \tag{51}$$

$\mathbf{C}_{jlgg'} = \mathbf{X}_{lg'}^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H} \in \mathbb{C}^{|\mathcal{G}_{lg'}|\times|\mathcal{G}_{jg}|}$. According to (14), we have $\mathbf{C}_{jjgg} = \rho\mathbf{I}_{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$ and $\mathbf{C}_{jlgg'} = \mathbf{0}_{|\mathcal{G}_{lg'}|\times|\mathcal{G}_{jg}|}$. Thus, (50) can be re-written as

$$\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}} = \rho\bar{\mathbf{H}}_{jjgg} + \sum_{\substack{l=1\\l\neq j}}^{L}\sum_{g'=1}^{G}\bar{\mathbf{H}}_{jlgg'}\mathbf{C}_{jlgg'} + \mathbf{U}_{jjg}^H\mathbf{Z}_j^{\mathrm{p}}\mathbf{X}_{jg}^{\mathrm{p}\ H}.$$

With per-group processing, the sum rate upper bound is

$$R_{\mathrm{group}} \triangleq \sum_{j=1}^{L}\sum_{g=1}^{G}\sum_{k=1}^{|\mathcal{G}_{jg}|}\log_2(1+\bar{\gamma}_{jgk}). \tag{52}$$

The details of the derivation are shown in Appendix F. The proposed pilot design in Section IV can still be used, but the pilot allocation should be changed. More specifically, the pilot allocation object in (46) should be changed into

$$\max_{\mathbf{X}_l^{\mathrm{p}}\in\mathcal{X}} \acute{R}_l, \tag{53}$$

where $\acute{R}_l$ is the sum rate upper bound with the pilots of unallocated cells being zeros but with (14) kept. Then, the greedy pilot allocation is as Algorithm 2 below.

Algorithm 2 Greedy pilot allocation for general grouping

1:**for** $l = 1 \rightarrow L$
2: allocate pilots as (53)
3:**end for**

B. Ideal Grouping

The rank of $\tilde{\mathbf{W}}_{jlg}$ depends on the angle range, the array aperture, and the number of antennas. According to the results in [31] and [32], with limited scattering, the sum of a limited number of the eigenvalues of $\tilde{\mathbf{W}}_{jlg}$ approximates the trace of $\tilde{\mathbf{W}}_{jlg}$, i.e., the rank of $\tilde{\mathbf{W}}_{jlg}$ can be approximately taken as a limited value. According to the results in [2], with appropriate grouping, $\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jjg'} \approx \mathbf{0}_{|\mathcal{G}_{jg}|\times M}, \forall g' \neq g$ holds with high probability. Thus, in the single-cell case, the two-tier receiving can approximately achieve the effect of BD; in the multi-cell case, only inter-cell interference deteriorates the system sum rate. More specifically, we have the following proposition.

Proposition 4. *With ideal grouping, i.e.,* $\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jjg'} = \mathbf{0}_{|\mathcal{G}_{jg}|\times M}, \forall g' \neq g$, *the inter-group interference inside each cell can be eliminated.*

Proof: When $\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jjg'} = \mathbf{0}_{|\mathcal{G}_{jg}|\times M}, \forall g' \neq g$, according to (51), we have $\bar{\mathbf{H}}_{jjgg'} = \mathbf{0}_{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg'}|}$. Accordingly, (98) can be changed into

$$\tilde{\mathbf{y}}_{jg}^{\mathrm{d}} = \bar{\mathbf{H}}_{jjgg}\mathbf{x}_{jg}^{\mathrm{d}} + \sum_{\substack{l=1\\l\neq j}}^{L}\mathbf{U}_{jjg}^H\mathbf{H}_{jl}\mathbf{x}_l^{\mathrm{d}} + \mathbf{U}_{jjg}^H\mathbf{z}_j^{\mathrm{d}}.$$

As can be seen, the inter-group interference inside each cell disappears when beamforming is applied to the received data signals. Similarly, the pilots in (50) are changed into

$$\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}} = \bar{\mathbf{H}}_{jjgg}\mathbf{C}_{jjgg} + \sum_{\substack{l=1\\l\neq j}}^{L}\sum_{g'=1}^{G}\bar{\mathbf{H}}_{jlgg'}\mathbf{C}_{jlgg'} + \tilde{\mathbf{Z}}_{jg}^{\mathrm{p}},$$

where $\tilde{\mathbf{Z}}^{\mathrm{p}}_{jg} = \mathbf{U}^{H}_{jjg}\mathbf{Z}^{\mathrm{p}}_{j}{\mathbf{X}^{\mathrm{p}}_{jg}}^{H} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$. As can be seen, without the condition that the pilots in each cell are orthogonal, the inter-group interference is eliminated with beamforming. Since $\mathbf{C}_{jjgg} = \rho\mathbf{I}_{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, the pilots in the above equation are changed into

$$\tilde{\mathbf{Y}}^{\mathrm{p}}_{jg} = \rho\bar{\mathbf{H}}_{jjgg} + \sum_{\substack{l=1\\ l\neq j}}^{L}\sum_{g'=1}^{G}\bar{\mathbf{H}}_{jlgg'}\mathbf{C}_{jlgg'} + \tilde{\mathbf{Z}}^{\mathrm{p}}_{jg}.$$

Apparently, the intra-group interference in the pilots is eliminated, and only inter-cell interference exists. ■

Since each group can be taken as a virtual cell, the asymptotic result in Corollary 1 can also be employed here, which means the inter-group interference tends to be completely eliminated as the number of BS antennas tends to infinity when the DOA ranges of the groups are non-overlapping and the BS array is linear or rectangular. More specifically, we have the corollary as follows.

Corollary 2. *When the BS is equipped with a linear array and the direction-of-arrival (DOA) region of MSs in one group is non-overlapping with that in any other group, we have*

$$\lim_{M\to\infty}\gamma_{jgk} = \frac{1}{[(\mathbf{I}_{|\mathcal{G}_{jg}|} + \bar{\mathbf{V}}^{\mathrm{a}}_{jg})^{-1}]_{k,k}} - 1, \tag{54}$$

where $\bar{\mathbf{H}}^{\mathrm{a}}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, $\mathbf{\Xi}^{\mathrm{a}}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, $\bar{\mathbf{V}}^{\mathrm{a}}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, *and are defined as*

$$\begin{aligned}
\bar{\mathbf{H}}^{\mathrm{a}}_{jg} &= \mathbf{D}_{jjg}\left(\mathbf{D}_{jjg} + \frac{1}{\rho}\mathbf{I}_{|\mathcal{G}_{jg}|}\right)^{-1}(\bar{\mathbf{H}}_{jjgg} + \frac{1}{\rho}\tilde{\mathbf{Z}}^{\mathrm{p}}_{jg}),\\
\mathbf{\Xi}^{\mathrm{a}}_{jg} &= \rho|\mathcal{G}_{jg}|\mathbf{D}_{jjg}\left(\rho\mathbf{D}_{jjg} + \mathbf{I}_{|\mathcal{G}_{jg}|}\right)^{-1} + \mathbf{I}_{|\mathcal{G}_{jg}|},\\
\bar{\mathbf{V}}^{\mathrm{a}}_{jg} &= \rho(\bar{\mathbf{H}}^{\mathrm{a}}_{jg})^{H}(\mathbf{\Xi}^{\mathrm{a}}_{jg})^{-1}\bar{\mathbf{H}}^{\mathrm{a}}_{jg}.
\end{aligned}$$

Remark 5. *As the proof is similar to Proposition 1, Lemma2, and Corollary 1, it is omitted here. Also, it can be extended to the scenario with rectangular arrays. This corollary shows that the large number of BS antennas is beneficial for two-tier processing and lessens the requirement on pilots.*

With the results of the group correlations, we will continue to show simulation results in the next section which can verify the effect of the pilot design and allocation and the analysis on the SINR and the sum rate.

VI. Numerical Results

In this section, numerical results are provided to compare the performance of the proposed approach and other approaches. There are three approaches for comparison. One is the reused pilot design, which means $\mathbf{X}^{\mathrm{p}}_{l} = \mathbf{X}^{\mathrm{p}}_{j}, \forall j, l$, and is denoted as "Reuse". One is the pilot design with minimal phase difference between the pilots proposed in [6] and [7], and is denoted as "Min-phase [6][7]". The proposed pilot design and allocation is denoted as "Proposed". Moreover, the ideal pilots, i.e., orthogonal pilots with $\mathbf{C}_{jl} = \mathbf{0}_{K\times K}, \forall l \neq j$ are also compared, and are denoted as "Ideal". It should be mentioned that the ideal pilots cannot be realized unless $L_{\mathrm{p}} \geq KL$. As we assume $L_{\mathrm{p}} < 2K$ in this paper and set $L = 3$, the sum rate with the ideal pilots is taken as an unattainable upper bound on the sum rate achieved by the pilot design. Note that the sum rate is calculated with (28) and by replacing the expectation with the average of the trials. The upper bounds in (39) and (52) are also shown here, and are denoted as "Upper bound (39)" and "Upper bound (52)". Note that the sum rate upper bound (39) corresponds to the SINR upper bound in (38). Note that the upper bounds are calculated with the proposed pilot design. Moreover, in the simulations with changing number of BS antennas, the asymptotic sum rates based on the SINRs in (43) and (54) are also compared, which are defined as

$$R_{\mathrm{a}} = \sum_{j=1}^{L}\sum_{k=1}^{K}\log_2\left(\frac{1}{[(\mathbf{I}_K + \bar{\mathbf{V}}^{\mathrm{a}}_{j})^{-1}]_{k,k}}\right), \tag{55}$$

$$R^{\mathrm{group}}_{\mathrm{a}} = \sum_{j=1}^{L}\sum_{g=1}^{G}\sum_{k=1}^{|\mathcal{G}_{jg}|}\log_2\left(\frac{1}{[(\mathbf{I}_K + \bar{\mathbf{V}}^{\mathrm{a}}_{jg})^{-1}]_{k,k}}\right), \tag{56}$$

and are denoted as "Asymptotic (55)" and "Asymptotic (56)".

Without other statements, the simulation parameters are as follows. The number of MSs in each cell is $K = 7$, the length of the pilots is $L_{\mathrm{p}} = 10$, the number of cells is $L = 3$, the codebook size is $N = 100$. The signal power is $\rho = 10^2$, and the SNR is defined as $10\log_{10}\rho = 20$ dB.

A. Results with $\mathbf{W}_{jlk} = \mathbf{W}_{jl}, \forall k$

The correlation $\mathbf{W}_{jl}$ is calculated based on numerical sampling of the MS positions. For $\mathbf{W}_{jl}$ in (7), it is generated based on the spatial channel model as

$$\mathbf{W}_{jl} = \frac{1}{N_{\mathrm{s}}}\sum_{n=1}^{N_{\mathrm{s}}}\mathbf{a}(\theta_n, \phi_n)\mathbf{a}^{H}(\theta_n, \phi_n),$$

where θ_n is a random vertical DOA, ϕ_n is a random horizontal DOA, and they are the line-of-sight DOAs from one random position in the l-th cell to the j-th BS, N_{s} is the number sampling points, $\mathbf{a}(\theta, \phi) \in \mathbb{C}^{M\times 1}$ is the array steering vector and is defined as

$$\begin{aligned}
[\mathbf{a}(\theta,\phi)]_{(m-1)N_{\mathrm{V}}+n} = \frac{1}{\sqrt{M}}\exp(i\pi(&-(n-1)\sin(\theta)\\
&+(m-1)\cos(\theta)\sin(\phi))),
\end{aligned}$$

where $m = 1, 2, \cdots, N_{\mathrm{H}}$, $n = 1, 2, \cdots, N_{\mathrm{V}}$. N_{H} is the number of horizontal antennas at the BS, and N_{V} is the number of vertical antennas at the BS, thus we have $M = N_{\mathrm{H}}N_{\mathrm{V}}$. Here we set $N_{\mathrm{H}} = 5$, $N_{\mathrm{V}} = 20$, $N_{\mathrm{s}} = 100$. The detailed illustration of the array and the DOAs can be found in [33]. The cells are in diamond shape and they constitute a hexagon, as shown in Fig. 1. The antenna array at the BS points towards the center of the cell in the horizontal plane and is vertical to the ground. The urban micro cell with street canyon setting defined by 3rd generation partnership project (3GPP) in [34] is also employed here. The distance between the BSs is 200 m. The minimal distance between the MS and the BS is 10 m. The BS height is 10 m. In the following simulations, two kinds of MS distributions are considered. One is the uniform distribution, and the other is the nonuniform distribution with MSs only distribute in the 60° sector with maximal range 50 m and is symmetric to the BS array boresight.

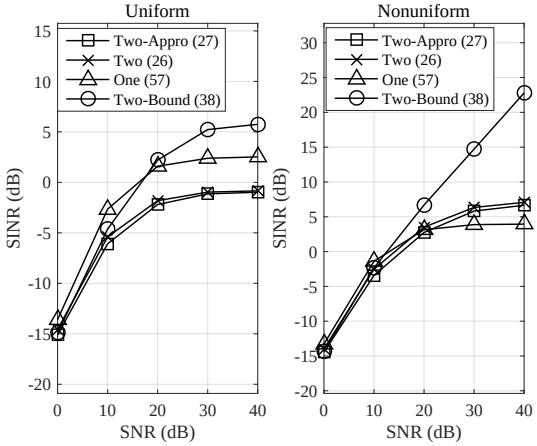

Fig. 4. The average of the SINRs in (26) and (27), the upper bound in (38), and the average of the SINR without two-tier receiving in (57) versus the SNR.

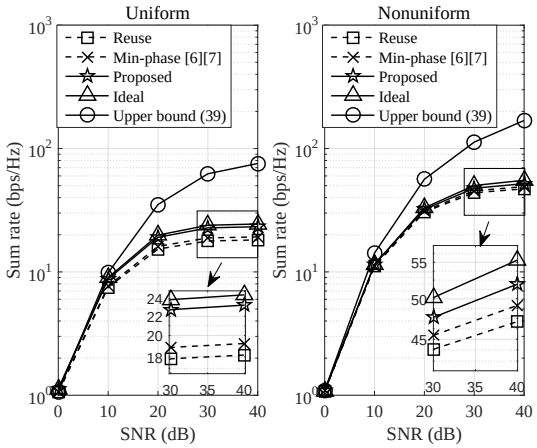

Fig. 5. The achievable sum rates of the three pilot designs versus the SNR.

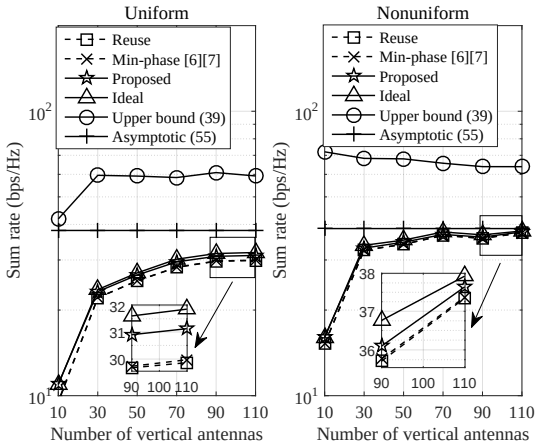

Fig. 6. The achievable sum rates of the three pilot designs versus the number of vertical antennas.

In Fig. 4, the SINR values are compared and are obtained with reused pilots. The SINRs are also compared here. The SINR upper bound for the two-tier processing in (38) is denoted as "Two-bound (38)". The SINR and the approximate

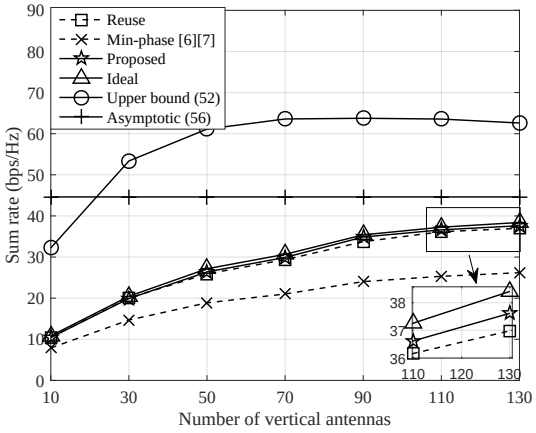

Fig. 7. The achievable sum rates of the three pilot designs with grouped correlation versus the number of vertical antennas.

SINR for the two-tier processing in (26) and (27) are denoted as "Two (26)" and "Two-Appro (27)". Moreover, the SINR achieved without two-tier processing, i.e., with MMSE channel estimation and MMSE receiving is also compared, and is denoted as "One (57)". Following the analyses in Section III, the SINR achieved without two-tier processing can be written as

$$\check{\gamma}_{jk} = \frac{\rho|\mathbf{a}_{jk}^H \check{\mathbf{h}}_{jk}|^2}{\rho \sum_{n\neq k} |\mathbf{a}_{jk}^H \check{\mathbf{h}}_{jn}|^2 + \mathbf{a}_{jk}^H \mathbb{E}\{\tilde{\boldsymbol{\xi}}_j \tilde{\boldsymbol{\xi}}_j^H | \boldsymbol{\xi}_{jk}, \bar{\mathbf{H}}_{jl}\} \mathbf{a}_{jk}}, \quad (57)$$

where

$$\check{\mathbf{h}}_{jk} = \check{\boldsymbol{\Sigma}}_{jk} \mathbf{y}_{jk}^{\mathrm{p}} \in \mathbb{C}^{M\times 1},$$

$$\check{\boldsymbol{\Sigma}}_{jk} = \mathbf{W}_{jj} \left(\sum_{l=1}^{L} \mathbf{W}_{jl} \frac{||[\mathbf{C}_{jl}]_k||^2}{\rho^2} + \frac{1}{\rho} \mathbf{I}_M \right)^{-1} \in \mathbb{C}^{M\times M},$$

$$\check{\mathbf{a}}_{jk} = \check{\boldsymbol{\Xi}}_{jk}^{-1} \check{\mathbf{h}}_{jk} \in \mathbb{C}^{K\times 1},$$

$$\check{\boldsymbol{\Xi}}_{jk} = \sum_{n\neq k} \check{\mathbf{h}}_{jn} \check{\mathbf{h}}_{jn}^H + \frac{1}{\rho} \check{\boldsymbol{\Xi}}_j \in \mathbb{C}^{K\times K},$$

$$\check{\boldsymbol{\Xi}}_j = \rho \mathbf{W}_{jj} (K\mathbf{I}_M - \sum_{k=1}^{K} \check{\boldsymbol{\Sigma}}_{jk}^H) + \rho K \sum_{l\neq j} \mathbf{W}_{jl} + \mathbf{I}_M \in \mathbb{C}^{M\times M}.$$

It can be seen that the derived upper bound in Lemma 1 is effective and is closer to the averaged SINR when the signal-to-noise ratio (SNR) is lower. This is because part of the interference is omitted in the SINR upper bound and the effect of the omitting error is more obvious when the SNR is higher. Moreover, as the spatial correlation with nonuniform distribution is smaller than that with uniform distribution, which means the interference in the former is lower than the interference in the latter, the averaged SINR and the upper bound with the nonuniform distribution are higher than that with uniform distribution. Besides, the SINR of the two-tier processing in (26) and the approximate SINR of the two-tier processing in (27) are almost the same, which shows that the approximation in (27) is reasonable. Additionally, it can be seen that the SINR achieved with the two-tier receiving is 3

dB higher than that without two-tier receiving when the SNR is higher than 20 dB and the MSs distribute in a nonuniform way, but the former is lower than the latter when the MSs distribute uniformly. The reason is as follows. As the MS distribution gets more nonuniformly, the overlap of the intra-cell channel subspace and the inter-cell channel subspace decreases, the outer processing gets more efficient in mitigating interference. Thus, despite the merits of reducing computational complexity and the number of RF chains, the two-tier receiving is of better performance in favorable scenarios, i.e., high SNR and nonuniform MS distribution.

In Fig. 5, the sum rate versus the SNR is demonstrated. The sum rates achieved by the proposed pilots are higher than that achieved by the existing pilots. Additionally, taking the ideal pilots as the reference, the gap between the proposed pilots and the existing pilots is large. For the uniform distribution, the proposed pilots can achieve 3 bps/Hz higher sum rate than other pilots, and the gap between the proposed pilots and the ideal pilots is 1 bps/Hz. For the nonuniform distribution, the proposed pilots can achieve 2 bps/Hz higher sum rate than other pilots, and the gap between the proposed pilots and the ideal pilots is 2 bps/Hz. These results show that the proposed pilots in Section IV are effective for massive MIMO systems with two-tier processing. It can also be seen that the gap between the proposed pilots and other pilots is smaller than the gap between the proposed pilots and the upper bound. In fact, the sum rate upper bound is not the bound on the sum rate with pilot design, which corresponds to the the sum rate with ideal pilots in the simulations, but the sum rate upper bound on the sum rate with the proposed pilots.

In Fig. 6, the sum rate versus the number of vertical BS antennas N_{V} is demonstrated. In this simulation, $N_{\mathrm{H}} = 1$, and N_{V} changes from 10 to 110, $N_{\mathrm{s}} = N_{\mathrm{H}}N_{\mathrm{V}}$. The sum rate of the proposed pilot design is higher than that of other pilots and is close to that of the ideal pilots. For the uniform distribution with $N_{\mathrm{V}} = 110$, the sum rate of the proposed pilots is 1.5 bps/Hz higher than that of other pilots, and is 0.7 bps/Hz lower than that of the ideal pilots. For the nonuniform distribution with $N_{\mathrm{V}} = 110$, the sum rate of the proposed pilots is 0.2 bps/Hz higher than that of other pilots, and is 0.3 bps/Hz lower than that of the ideal pilots. The upper bound is always effective, which verifies Lemma 1. As the number of antennas increases, the sum rate of the proposed pilots increases and gets close to the asymptotic value. This is because the channel subspaces tend to be orthogonal as the number of antennas increases. This result verifies Corollary 1.

In Fig. 5 and Fig. 6, it can also be seen that the gaps between the sum rates of the pilots with nonuniform MS distribution are larger than that with uniform MS distribution. When the MS distribution changes from the uniform distribution to the nonuniform distribution, the channel subspace correlation decreases, and the inter-cell interference lessens. Thus, the effect of pilot design lessens. This result verifies Lemma 2.

B. Results with Grouped $\mathbf{W}_{jlk}$

The urban macro cell setting defined by 3GPP in [34] is employed here. The distance between the BSs is 500 m. The minimal distance between the MS and the BS is 35 m. The BS height is 25 m. The number of MSs is $K = 6$, the number of groups is $G = 2$, and the number of MSs in each group is $|\mathcal{G}_{jg}| = 3, \forall j, g$. The maximal range of the scattering in each group is 40 m. The center of each group is also fixed. The channel correlation matrix of each group is calculated in a similar way as the last subsection and the BS array is the same.

In Fig. 7, the sum rate versus the number of vertical BS antennas N_{V} is demonstrated. In this simulation, $N_{\mathrm{H}} = 1$, and N_{V} changes from 10 to 130, $N_{\mathrm{s}} = N_{\mathrm{H}}N_{\mathrm{V}}$. The results manifest that the proposed pilots can achieve higher sum rates than other pilots in systems with grouped correlation. For the uniform distribution with $N_{\mathrm{V}} = 110$, the sum rate of the proposed pilots is 0.5 bps/Hz higher than that of other pilots, and is 0.7 bps/Hz lower than that of the ideal pilots. Also, the derived sum rate upper bound is effective, which verifies Section V. As the number of antennas increases, the sum rate of the proposed pilots gets close to the asymptotic sum rate, which is because the channel subspaces tend to be orthogonal as the number of antennas increases. This result verifies Corollary 2.

VII. Conclusions

In this paper, pilot design in massive MIMO systems with two-tier receiving is investigated. The SINR and the sum rate upper bound are analyzed. The pilots with various correlations are designed and allocated in a greedy based on the derived upper bound. The discussions show that the two-tier receiving can reduce the number of RF chains and the computational complexity in comparison to traditional one-tier receiving. Additionally, simulation results show that the two-tier receiving can achieve 3 dB gain in SINR in comparison to traditional one-tier receiving when the SNR is higher than 30 dB and the MSs distribute nonuniformly. The simulation results show that the proposed pilots benefit from the increase of the SNR and the increase of the number of antennas. The sum rate of the proposed pilots is 2 to 3 bps/Hz higher than the sum rate of other pilots with the 5×10 rectangular array and the SNR higher than 30 dB. The sum rate of the proposed pilots is 0.2 to 1.5 bps/Hz higher than the sum rate of other pilots with the 1×110 linear array and the SNR being 20 dB.

Appendix A
Proof of Proposition 1

Substituting (22) into (27) yields

$$\gamma_{jk} = \hat{\tilde{\mathbf{h}}}_{jk}^H \tilde{\boldsymbol{\Xi}}_{jk}^{-1} \hat{\tilde{\mathbf{h}}}_{jk}. \tag{58}$$

According to (23), there is

$$\tilde{\boldsymbol{\Xi}}_{jk} = \tilde{\boldsymbol{\Xi}}_j - \hat{\tilde{\mathbf{h}}}_{jk} \hat{\tilde{\mathbf{h}}}_{jk}^H, \tag{59}$$

where

$$\tilde{\boldsymbol{\Xi}}_j = \sum_{k=1}^{K} \hat{\tilde{\mathbf{h}}}_{jk} \hat{\tilde{\mathbf{h}}}_{jk}^H + \frac{1}{\rho} \boldsymbol{\Xi}_j \in \mathbb{C}^{K \times K}. \tag{60}$$

Applying the Sherman-Morrison formula to (59), we have

$$\tilde{\boldsymbol{\Xi}}_{jk}^{-1} = \tilde{\boldsymbol{\Xi}}_j^{-1} + \frac{1}{1 - z_{jk}} \tilde{\boldsymbol{\Xi}}_j^{-1} \hat{\tilde{\mathbf{h}}}_{jk} \hat{\tilde{\mathbf{h}}}_{jk}^H \tilde{\boldsymbol{\Xi}}_j^{-1}, \tag{61}$$

where

$$z_{jk} = \hat{\tilde{\mathbf{h}}}_{jk}^H \tilde{\mathbf{\Xi}}_j^{-1} \hat{\tilde{\mathbf{h}}}_{jk}. \quad (62)$$

Substituting (61) into (58) results into

$$\gamma_{jk} = \frac{z_{jk}}{1 - z_{jk}}. \quad (63)$$

According to (20) and substituting (60) into (62) yields

$$z_{jk} = \left[\hat{\tilde{\mathbf{H}}}_j^H \left(\hat{\tilde{\mathbf{H}}}_j \hat{\tilde{\mathbf{H}}}_j^H + \frac{1}{\rho} \mathbf{\Xi}_j \right)^{-1} \hat{\tilde{\mathbf{H}}}_j \right]_{k,k}. \quad (64)$$

Applying the Sherman-Morrison-Woodbury (SMW) formula to (64), we have

$$\begin{aligned} z_{jk} &= \left[\mathbf{V}_j (\mathbf{I}_K - (\mathbf{I}_K + \mathbf{V}_j)^{-1} \mathbf{V}_j) \right]_{k,k} \\ &= 1 - [(\mathbf{I}_K + \mathbf{V}_j)^{-1}]_{k,k}, \end{aligned} \quad (65)$$

where the second equation is also derived with the SMW formula. Substituting (65) into (63) results into (31). Note that the derivations in the proof have some similarities with those in [26] but the details and the results are not the same.

Appendix B
Proof of Proposition 2

According to (19), we have $\mathbb{E}\{\boldsymbol{\xi}_{jk}\boldsymbol{\xi}_{jk}^H\} = \mathbb{E}\{\bar{\mathbf{h}}_{jk}\bar{\mathbf{h}}_{jk}^H\} + \mathbb{E}\{\hat{\bar{\mathbf{h}}}_{jk}\hat{\bar{\mathbf{h}}}_{jk}^H\} - \mathbb{E}\{\bar{\mathbf{h}}_{jk}\hat{\bar{\mathbf{h}}}_{jk}^H\} - \mathbb{E}\{\hat{\bar{\mathbf{h}}}_{jk}\bar{\mathbf{h}}_{jk}^H\}$. Substituting (10) into (16) yields $\bar{\mathbf{h}}_{jk} = \mathbf{D}_{jj}^{\frac{1}{2}} \mathbf{U}_{jj}^H [\tilde{\mathbf{H}}_{jj}]_k$. Thus, we have $\mathbb{E}\{\bar{\mathbf{h}}_{jk}\bar{\mathbf{h}}_{jk}^H\} = \mathbf{D}_{jj}$. Substituting (15) into (17) results into $\hat{\bar{\mathbf{h}}}_{jk} = \mathbf{\Sigma}_{jk}(\bar{\mathbf{h}}_{jk} + \sum_{l \neq j} \frac{1}{\rho} \bar{\mathbf{H}}_{jl} [\mathbf{C}_{jl}]_k + \frac{1}{\rho} [\tilde{\mathbf{Z}}_j^{\mathrm{p}}]_k)$. Then, we have $\mathbb{E}\{\bar{\mathbf{h}}_{jk}\hat{\bar{\mathbf{h}}}_{jk}^H\} = \mathbb{E}\{\bar{\mathbf{h}}_{jk}\bar{\mathbf{h}}_{jk}^H\}\mathbf{\Sigma}_{jk}^H = \mathbf{D}_{jj}\mathbf{\Sigma}_{jk}^H$, $\mathbb{E}\{\hat{\bar{\mathbf{h}}}_{jk}\bar{\mathbf{h}}_{jk}^H\} = \mathbf{\Sigma}_{jk}\mathbf{D}_{jj}$, $\mathbb{E}\{\hat{\bar{\mathbf{h}}}_{jk}\hat{\bar{\mathbf{h}}}_{jk}^H\} = \mathbf{\Sigma}_{jk}(\sum_{l=1}^{L} \mathbf{U}_{jj}^H \mathbf{W}_{jl} \mathbf{U}_{jj} \frac{||[\mathbf{C}_{jl}]_k||^2}{\rho^2} + \frac{1}{\rho}\mathbf{I}_K)\mathbf{\Sigma}_{jk}^H = \mathbf{\Sigma}_{jk}\mathbf{D}_{jj}$, where (18) is used to derive the second equation. Consequently, we have $\mathbb{E}\{\boldsymbol{\xi}_{jk}\boldsymbol{\xi}_{jk}^H\} = \mathbf{D}_{jj} - \mathbf{D}_{jj}\mathbf{\Sigma}_{jk}^H$. Then, substituting (25) into (24), we have $\mathbf{\Xi}_j = \sum_{k=1}^{K} \mathbb{E}\{\boldsymbol{\xi}_{jk}\boldsymbol{\xi}_{jk}^H |[\mathbf{x}_j^{\mathrm{d}}]_k|^2\} + \sum_{l \neq j} \mathbb{E}\{\bar{\mathbf{H}}_{jl}\mathbf{x}_l^{\mathrm{d}}{}^H \bar{\mathbf{H}}_{jl}^H\} + \mathbb{E}\{\tilde{\mathbf{z}}_j^{\mathrm{d}}(\tilde{\mathbf{z}}_j^{\mathrm{d}})^H\}$, which means

$$\mathbf{\Xi}_j = \rho \mathbf{D}_{jj}(K\mathbf{I}_K - \sum_{k=1}^{K} \mathbf{\Sigma}_{jk}^H) + \rho K \sum_{l \neq j} \mathbf{U}_{jj}^H \mathbf{W}_{jl} \mathbf{U}_{jj} + \mathbf{I}_K.$$

According to (32), we have

$$\begin{aligned} \mathbb{E}\{[\mathbf{V}_j]_{s,t}\} &= \rho \mathbb{E}\{\hat{\bar{\mathbf{h}}}_{js}^H \mathbf{\Xi}_j^{-1} \hat{\bar{\mathbf{h}}}_{jt}\} \\ &= \rho \mathbb{E}\{\mathbf{y}_{js}^{\mathrm{p}}{}^H \mathbf{\Sigma}_{js} \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{jt} \mathbf{y}_{jt}^{\mathrm{p}}\} \\ &= \rho(\mathbb{E}\{\bar{\mathbf{h}}_{js}^H \mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{jt} \bar{\mathbf{h}}_{jt}\} \\ &\quad + \frac{1}{\rho^2} \mathbb{E}\{[\tilde{\mathbf{Z}}_j^{\mathrm{p}}]_s^H \mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{jt} [\tilde{\mathbf{Z}}_j^{\mathrm{p}}]_t\} \\ &\quad + \mathbb{E}\{\sum_{l \neq j} \frac{1}{\rho^2} [\mathbf{C}_{jl}]_s^H \bar{\mathbf{H}}_{jl}^H \mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{jt} \bar{\mathbf{H}}_{jl} [\mathbf{C}_{jl}]_t\}) \\ &= \rho(\delta(s-t)\mathrm{tr}\{\mathbf{D}_{jj}\mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{js}\} \\ &\quad + \frac{1}{\rho}\delta(s-t)\mathrm{tr}\{\mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{js}\} \\ &\quad + \frac{1}{\rho^2} \sum_{l \neq j} \mathrm{tr}\{\mathbf{U}_{jj}\mathbf{\Sigma}_{js}^H \mathbf{\Xi}_j^{-1} \mathbf{\Sigma}_{jt} \mathbf{U}_{jj}^H \mathbf{W}_{jl}\} \\ &\quad \times [\mathbf{C}_{jl}]_s^H [\mathbf{C}_{jl}]_t), \end{aligned}$$

where (15) and (17) are used to derive the second and the third equations, the definition of $\bar{\mathbf{h}}_{jk}$ in (16) and (10), the definition of $\bar{\mathbf{H}}_{jl}$ in (9), as well as the definition of $\tilde{\mathbf{Z}}_j^{\mathrm{p}}$ in (12) are used to derive the forth equation. Then, according to the definition in (33), we have (34).

Appendix C
Proof of Lemma 1

1)Denote

$$\mathbf{P}_j = \sqrt{\rho} \mathbf{\Xi}_j^{-1/2} \hat{\tilde{\mathbf{H}}}_j \in \mathbb{C}^{K \times K}, \quad (66)$$

$$(\mathbf{P}_j^H)_{(-k)}^H = \tilde{\mathbf{U}}_{jk} \tilde{\mathbf{D}}_{jk} \tilde{\mathbf{V}}_{jk}^H, \quad (67)$$

where the rank of $(\mathbf{P}_j^H)_{(-k)}^H \in \mathbb{C}^{K \times (K-1)}$ is $K-1$, which is because that $\hat{\tilde{\mathbf{H}}}_j$ is composed of independent random variables and the relation between $\mathbf{P}_j$ and $\hat{\tilde{\mathbf{H}}}_j$ is as (66). Moreover, $\tilde{\mathbf{U}}_{jk} \in \mathbb{C}^{K \times (K-1)}$ and $\tilde{\mathbf{V}}_{jk} \in \mathbb{C}^{(K-1) \times (K-1)}$ are matrices of the singular vectors that correspond to non-zero singular values, $\tilde{\mathbf{D}}_{jk} \in \mathbb{R}^{(K-1) \times (K-1)}$ is the matrix of non-zero singular values.

2)Denote $\mathbf{G}_{jk} \in \mathbb{C}^{K \times K}$,

$$\mathbf{G}_{jk} = \mathbb{E}\{[[\mathbf{P}_j]_k \ \ (\mathbf{P}_j^H)_{(-k)}^H]^H [[\mathbf{P}_j]_k \ \ (\mathbf{P}_j^H)_{(-k)}^H]\}, \quad (68)$$

$$\zeta_{jk} = [\mathbf{G}_{jk}]_{1,1} - [\mathbf{G}_{jk}]_{1(-1)}^H \mathbf{G}_{jk(-1,-1)}^{-1} [\mathbf{G}_{jk}]_{1(-1)}. \quad (69)$$

3)Denote

$$\mathbf{B}_j = \mathbf{I}_K + \mathbf{V}_j \in \mathbb{C}^{K \times K}. \quad (70)$$

From (32) and (66), we have

$$\mathbf{V}_j = \mathbf{P}_j^H \mathbf{P}_j. \quad (71)$$

Moreover, from (33), we have

$$\mathbf{R}_j = \mathbb{E}\{\mathbf{P}_j^H \mathbf{P}_j\}. \quad (72)$$

According to (68), (69), and (72), we have

$$\begin{aligned} \zeta_{jk} &= [\mathbf{R}_j]_{k,k} - [\mathbf{R}_j]_{k(-k)}^H \mathbf{R}_{j(-k,-k)}^{-1} [\mathbf{R}_j]_{k(-k)} \\ &= \frac{1}{[\mathbf{R}_j^{-1}]_{k,k}}. \end{aligned} \quad (73)$$

Based on (31), (70), and (71), we have [30]

$$\begin{aligned} \gamma_{jk} &= [\mathbf{B}_j]_{k,k} - [\mathbf{B}_j]_{k(-k)}^H \mathbf{B}_{j(-k,-k)}^{-1} [\mathbf{B}_j]_{k(-k)} - 1 \\ &= [\mathbf{V}_j]_{k,k} - [\mathbf{P}_j]_k^H (\mathbf{P}_j^H)_{(-k)}^H (\mathbf{I}_{K-1} \\ &\quad + (\mathbf{P}_j^H)_{(-k)} (\mathbf{P}_j^H)_{(-k)}^H)^{-1} (\mathbf{P}_j^H)_{(-k)} [\mathbf{P}_j]_k. \end{aligned} \quad (74)$$

Then, γ_{jk} in (74) can be written as

$$\begin{aligned} \gamma_{jk} &= [\mathbf{P}_j]_k^H [\tilde{\mathbf{U}}_{jk} \tilde{\mathbf{u}}_{jk}^{\mathrm{c}}][\tilde{\mathbf{U}}_{jk} \tilde{\mathbf{u}}_{jk}^{\mathrm{c}}]^H [\mathbf{P}_j]_k \\ &\quad - [\mathbf{P}_j]_k^H \tilde{\mathbf{U}}_{jk} \tilde{\mathbf{D}}_{jk} \tilde{\mathbf{V}}_{jk}^H (\mathbf{I}_{K-1} + \tilde{\mathbf{V}}_{jk} \tilde{\mathbf{D}}_{jk}^2 \tilde{\mathbf{V}}_{jk}^H)^{-1} \\ &\quad \times \tilde{\mathbf{V}}_{jk} \tilde{\mathbf{D}}_{jk} \tilde{\mathbf{U}}_{jk}^H [\mathbf{P}_j]_k \\ &= |s_{jk}|^2 + \mathbf{t}_{jk}^H \mathbf{t}_{jk} - \mathbf{t}_{jk}^H \tilde{\mathbf{D}}_{jk} \tilde{\mathbf{V}}_{jk}^H \\ &\quad \times (\mathbf{I}_{K-1} + \tilde{\mathbf{V}}_{jk} \tilde{\mathbf{D}}_{jk}^2 \tilde{\mathbf{V}}_{jk}^H)^{-1} \tilde{\mathbf{V}}_{jk} \tilde{\mathbf{D}}_{jk} \mathbf{t}_{jk} \\ &= |s_{jk}|^2 + \mathbf{t}_{jk}^H (-\tilde{\mathbf{D}}_{jk}^2 (\mathbf{I}_{K-1} + \tilde{\mathbf{D}}_{jk}^2)^{-1} + \mathbf{I}_{K-1}) \mathbf{t}_{jk} \\ &= |s_{jk}|^2 + \mathbf{t}_{jk}^H (\mathbf{I}_{K-1} + \tilde{\mathbf{D}}_{jk}^2)^{-1} \mathbf{t}_{jk}, \end{aligned} \quad (75)$$

where $\tilde{\mathbf{u}}_{jk}^{\mathrm{c}} \in \mathbb{C}^{K\times 1}$ is the null space of $\tilde{\mathbf{U}}_{jk}$, i.e., $(\tilde{\mathbf{u}}_{jk}^{\mathrm{c}})^H\tilde{\mathbf{u}}_{jk}^{\mathrm{c}}=1$, $(\tilde{\mathbf{u}}_{jk}^{\mathrm{c}})^H\tilde{\mathbf{U}}_{jk}=\mathbf{0}_{1\times(K-1)}$;

$$s_{jk} = (\tilde{\mathbf{u}}_{jk}^{\mathrm{c}})^H[\mathbf{P}_j]_k, \tag{76}$$

$$\mathbf{t}_{jk} = \tilde{\mathbf{U}}_{jk}^H[\mathbf{P}_j]_k \in \mathbb{C}^{(K-1)\times 1}. \tag{77}$$

From (76) and (77), we have

$$\begin{aligned}\mathbb{E}\{|s_{jk}|^2+\mathbf{t}_{jk}^H\mathbf{t}_{jk}\} &= \mathbb{E}\{||[\tilde{\mathbf{u}}_{jk}^{\mathrm{c}}\tilde{\mathbf{U}}_{jk}]^H[\mathbf{P}_j]_k||^2\}\\ &= \mathbb{E}\{||[\mathbf{P}_j]_k||^2\}\\ &= \mathbb{E}\{[\mathbf{V}_j]_{k,k}\} \quad (78)\\ &= [\mathbf{R}_j]_{k,k}, \quad (79)\end{aligned}$$

where (78) is based on (71), and (79) is based on (33).

Then, the SINR in (75) satisfies

$$\gamma_{jk} = |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\mathbf{t}_{jk}^H\tilde{\mathbf{D}}_{jk}^2(\mathbf{I}_{K-1}+\tilde{\mathbf{D}}_{jk}^2)^{-1}\mathbf{t}_{jk} \tag{80}$$

$$\leq |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\frac{||\tilde{\mathbf{D}}_{jk}\mathbf{t}_{jk}||^2}{1+\mathrm{tr}\{\tilde{\mathbf{D}}_{jk}^2\}} \tag{81}$$

$$= |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\frac{||\tilde{\mathbf{V}}_{jk}\tilde{\mathbf{D}}_{jk}\tilde{\mathbf{U}}_{jk}^H[\mathbf{P}_j]_k||^2}{1+\mathrm{tr}\{\tilde{\mathbf{V}}_{jk}\tilde{\mathbf{D}}_{jk}^2\tilde{\mathbf{V}}_{jk}^H\}} \tag{82}$$

$$= |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\frac{||(\mathbf{P}_j^H)_{(-k)}[\mathbf{P}_j]_k||^2}{1+\mathrm{tr}\{(\mathbf{P}_j^H)_{(-k)}(\mathbf{P}_j^H)_{(-k)}^H\}} \tag{83}$$

$$= |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\frac{||[\mathbf{V}_j]_{k(-k)}||^2}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}, \tag{84}$$

where (80) is derived based on $(\mathbf{I}_{K-1}+\tilde{\mathbf{D}}_{jk}^2)^{-1} = \mathbf{I}_{K-1}-\tilde{\mathbf{D}}_{jk}^2(\mathbf{I}_{K-1}+\tilde{\mathbf{D}}_{jk}^2)^{-1}$, (81) is based on $\mathrm{tr}\{\tilde{\mathbf{D}}_{jk}^2\} \geq \max_n[\tilde{\mathbf{D}}_{jk}^2]_{n,n}$. Moreover, (82) is based on the fact that $\tilde{\mathbf{V}}_{jk}$ is unitary and by substituting (77), (83) is based on the definition in (67), (84) is based on the definition in (71).

According to the correlation property in (68) and the properties of the multivariate normal distributions [29][30], we have

$$[\mathbf{P}_j]_k-\boldsymbol{\mu}_{jk} \sim \mathcal{CN}(\mathbf{0}_K, \frac{\zeta_{jk}}{K}\mathbf{I}_K),$$

where

$$\boldsymbol{\mu}_{jk} = (\mathbf{P}_j^H)_{(-k)}^H\mathbf{G}_{jk(-1,-1)}^{-1}[\mathbf{G}_{jk}]_{1(-1)} \in \mathbb{C}^{K\times 1}. \tag{85}$$

Correspondingly, we have $(\mathbf{P}_j^H)_{(-k)}[\mathbf{P}_j]_k-(\mathbf{P}_j^H)_{(-k)}\boldsymbol{\mu}_{jk} \sim \mathcal{CN}(\mathbf{0}_K, \frac{\zeta_{jk}}{K}\mathbb{E}\{(\mathbf{P}_j^H)_{(-k)}(\mathbf{P}_j^H)_{(-k)}^H\})$. According to the definition of $\bar{\mathbf{R}}_j$ in (72), we have

$$(\mathbf{P}_j^H)_{(-k)}[\mathbf{P}_j]_k-(\mathbf{P}_j^H)_{(-k)}\boldsymbol{\mu}_{jk} \sim \mathcal{CN}(\mathbf{0}_K, \frac{\zeta_{jk}}{K}\mathbf{R}_{j(-k,-k)}).$$

According to the definition of $\mathbf{V}_j$ in (71) and the definition of $\boldsymbol{\mu}_{jk}$ in (85), we further have

$$[\mathbf{V}_j]_{k(-k)}-\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk} \triangleq \boldsymbol{\nu}_{jk} \sim \mathcal{CN}(\mathbf{0}_K, \frac{\zeta_{jk}}{K}\mathbf{R}_{j(-k,-k)}), \tag{86}$$

where $\mathbf{g}_{jk}=\mathbf{G}_{jk(-1,-1)}^{-1}[\mathbf{G}_{jk}]_{1(-1)} \in \mathbb{C}^{(K-1)\times 1}$. Note that the derivations in this paragraph are similar to that in [30], but the details are not the same.

Substituting (86) into (84), we have

$$\gamma_{jk} \leq |s_{jk}|^2+||\mathbf{t}_{jk}||^2-\frac{||\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk}+\boldsymbol{\nu}_{jk}||^2}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}. \tag{87}$$

For the expectation of the last part, we have

$$\begin{aligned}&\mathbb{E}\{\frac{||\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk}+\boldsymbol{\nu}_{jk}||^2}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}\}\\ &= \mathbb{E}_{\mathbf{V}_j}\{\mathbb{E}_{\boldsymbol{\nu}_{jk}}\{\frac{||\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk}||^2}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}\\ &+\frac{||\boldsymbol{\nu}_{jk}||^2+2\mathrm{Re}\{\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk}\boldsymbol{\nu}_{jk}^H\}}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}\}\} \quad (88)\\ &\geq \mathbb{E}_{\mathbf{V}_j}\{\frac{\zeta_{jk}/K\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}\} \quad (89)\\ &\geq \frac{\zeta_{jk}/K\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{1+\mathbb{E}_{\mathbf{V}_j}\{\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}\}} \quad (90)\\ &= \frac{\zeta_{jk}/K\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{1+\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}, \quad (91)\end{aligned}$$

where in (88) we temporally use the notations $\mathbb{E}_{\mathbf{V}_j}$, $\mathbb{E}_{\boldsymbol{\nu}_{jk}}$, and $\mathrm{Re}\{\cdot\}$ to denote the expectation of $\mathbf{V}_j$, the expectation of $\boldsymbol{\nu}_{jk}$, and the real part, respectively; (89) is based on the zero mean property of $\boldsymbol{\nu}_{jk}$ in (86), (90) is based on the Jensen's inequality, and (91) is based on (33).

By taking the expectation of γ_{jk} in (87), we have

$$\begin{aligned}\mathbb{E}\{\gamma_{jk}\} &\leq \mathbb{E}\{|s_{jk}|^2+||\mathbf{t}_{jk}||^2\}\\ &\quad -\mathbb{E}\{\frac{||\mathbf{V}_{j(-k,-k)}\mathbf{g}_{jk}+\boldsymbol{\nu}_{jk}||^2}{1+\mathrm{tr}\{\mathbf{V}_{j(-k,-k)}\}}\}\\ &\leq [\mathbf{R}_j]_{k,k}-\frac{\zeta_{jk}}{K}\frac{\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{1+\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}} \quad (92)\\ &= \bar{\gamma}_{jk}, \quad (93)\end{aligned}$$

where (93) is based on (79) and (91). Substituting (73) into (92), we have

$$\bar{\gamma}_{jk} = [\mathbf{R}_j]_{k,k}-\frac{\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\}}{K[\mathbf{R}_j^{-1}]_{k,k}(1+\mathrm{tr}\{\mathbf{R}_{j(-k,-k)}\})}.$$

Appendix D
Proof of Lemma 2

According to (18), we have

$$\lim_{\rho\to\infty}\boldsymbol{\Sigma}_{jk} = \mathbf{D}_{jj}\left(\sum_{l=1}^{L}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj}\frac{||[\mathbf{C}_{jl}]_k||^2}{\rho^2}\right)^{-1}. \tag{94}$$

According to (15)-(17), we have $\check{\mathbf{H}}_j \in \mathbb{C}^{K\times K}$ and

$$[\check{\mathbf{H}}_j]_k \triangleq \lim_{\rho\to\infty}\hat{\bar{\mathbf{h}}}_{jk} = \bar{\boldsymbol{\Sigma}}_{jk}(\bar{\mathbf{h}}_{jk}+\sum_{l\neq j}\frac{1}{\rho}\bar{\mathbf{H}}_{jl}[\mathbf{C}_{jl}]_k). \tag{95}$$

According to (37), we have

$$\lim_{\rho\to\infty}\frac{1}{\rho}\boldsymbol{\Xi}_j = \mathbf{D}_{jj}(K\mathbf{I}_K-\sum_{k=1}^{K}\boldsymbol{\Sigma}_{jk}^H)+K\sum_{l\neq j}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj}. \tag{96}$$

Substituting (96) into (32) yields

$$\begin{aligned}\lim_{\rho\to\infty}\mathbf{V}_j &= \check{\mathbf{H}}_j^H(\mathbf{D}_{jj}(K\mathbf{I}_K-\sum_{k=1}^{K}\bar{\boldsymbol{\Sigma}}_{jk}^H)+\\ &K\sum_{l\neq j}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj})^{-1}\check{\mathbf{H}}_j. \quad (97)\end{aligned}$$

When $\mathbf{C}_{jl}/\rho \to \mathbf{0}_{K\times K}, \forall l \neq j$, according to (94), we have

$$\lim_{\substack{\mathbf{C}_{jl}/\rho\to\mathbf{0}_{K\times K},\forall l\neq j \\ \rho\to\infty}} \mathbf{\Sigma}_{jk} = \mathbf{I}_K.$$

According to (95), we have

$$\lim_{\substack{\mathbf{C}_{jl}/\rho\to\mathbf{0}_{K\times K},\forall l\neq j \\ \rho\to\infty}} [\check{\mathbf{H}}_j]_k = \bar{\mathbf{h}}_{jk},$$

which means

$$\lim_{\substack{\mathbf{C}_{jl}/\rho\to\mathbf{0}_{K\times K},\forall l\neq j \\ \rho\to\infty}} \check{\mathbf{H}}_j = \bar{\mathbf{H}}_{jj}.$$

According to (97), we have

$$\lim_{\substack{\mathbf{C}_{jl}/\rho\to\mathbf{0}_{K\times K},\forall l\neq j \\ \rho\to\infty}} \mathbf{V}_j = \bar{\mathbf{H}}_{jj}^H (K\sum_{l\neq j}\mathbf{U}_{jj}^H\mathbf{W}_{jl}\mathbf{U}_{jj})^{-1}\bar{\mathbf{H}}_{jj}.$$

Then, according to (31), we have (41).

When $\mathbf{U}_{jj}^H\mathbf{W}_{jl} \to \mathbf{0}_{K\times M}, \forall l \neq j$, according to (8) and (9), we have $\bar{\mathbf{H}}_{jl} \to \mathbf{0}_{K\times K}, \forall l \neq j$. According to (18), we have $\mathbf{\Sigma}_{jk} \to \mathbf{D}_{jj}\left(\mathbf{D}_{jj} + \frac{1}{\rho}\mathbf{I}_K\right)^{-1}$. According to (15)-(17), we have

$$\hat{\bar{\mathbf{h}}}_{jk} \to \mathbf{D}_{jj}\left(\mathbf{D}_{jj} + \frac{1}{\rho}\mathbf{I}_K\right)^{-1}(\bar{\mathbf{h}}_{jk} + \frac{1}{\rho}[\tilde{\mathbf{Z}}_j^{\mathrm{p}}]_k),$$

which means $\hat{\bar{\mathbf{H}}}_j \to \bar{\mathbf{H}}_j^{\mathrm{a}}$. According to (37), we have

$$\begin{aligned}\mathbf{\Xi}_j &\to \rho K\mathbf{D}_{jj}\left(\mathbf{I}_K - \mathbf{D}_{jj}\left(\mathbf{D}_{jj} + \frac{1}{\rho}\mathbf{I}_K\right)^{-1}\right) + \mathbf{I}_K \\ &= \mathbf{\Xi}_j^{\mathrm{a}}.\end{aligned}$$

According to (32), we have $\mathbf{V}_j \to \bar{\mathbf{V}}_j^{\mathrm{a}}$. According to (31), we have (42).

Appendix E
Proof of Proposition 3

When the rank of $\mathbf{W}_{jl}$ is not larger than K, according to (7), we have $\mathbf{W}_{jl} = \mathbf{U}_{jl}\mathbf{D}_{jl}\mathbf{U}_{jl}^H$. When $\mathbf{U}_{jj}^H\mathbf{U}_{jl} \to \mathbf{0}_{K\times K}, \forall j \neq l$, which means $\mathbf{U}_{jj}^H\mathbf{W}_{jl} \to \mathbf{0}_{K\times M}, \forall j \neq l$, according to Lemma 2, we know that the impact of pilots on the SINR γ_{jk} tends to diminish. On the other hand, according to (37), we have $\mathbf{\Xi}_j \to \rho\mathbf{D}_{jj}(K\mathbf{I}_K - \sum_{k=1}^K \mathbf{\Sigma}_{jk}) + \mathbf{I}_K$. According to (35), we have $\mathbf{X}_{jl} \to \mathbf{0}_{K\times K}, \forall j \neq l$. According to (34), we have $\mathbf{R}_j \to \mathbf{Y}_j$. According to (36), we know that the pilot correlation $\mathbf{C}_{jl}$ does not exist in the asymptotic expression of $\mathbf{Y}_j$. As a result, the impact of pilots on $\mathbf{R}_j$ tends to diminish as $\mathbf{U}_{jj}^H\mathbf{U}_{jl} \to \mathbf{0}_{K\times K}, \forall j \neq l$. According to Lemma 1, we know that the impact of pilots on the sum rate upper bound $\bar{\gamma}_{jk}$ tends to diminish as $\mathbf{U}_{jj}^H\mathbf{U}_{jl} \to \mathbf{0}_{K\times K}, \forall j \neq l$.

Appendix F
Derivation of The Sum Rate Upper bound

For the k-th MS in this group, the corresponding processed pilot is

$$\begin{aligned}\mathbf{y}_{jgk}^{\mathrm{p}} &= \frac{1}{\rho}[\tilde{\mathbf{Y}}_{jg}^{\mathrm{p}}]_k = \bar{h}_{jgk} \\ &+\frac{1}{\rho}\sum_{\substack{l=1\\l\neq j}}^{L}\sum_{g'=1}^{G}\bar{\mathbf{H}}_{jlgg'}[\mathbf{C}_{jlgg'}]_k + \frac{1}{\rho}\mathbf{U}_{jjg}^H\mathbf{Z}_j^{\mathrm{p}}[{\mathbf{X}_{jg}^{\mathrm{p}}}^H]_k,\end{aligned}$$

where $\bar{h}_{jgk} = [\bar{\mathbf{H}}_{jjgg}]_k \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1}$, and the MMSE channel estimate is [26][27]

$$\hat{\bar{\mathbf{h}}}_{jgk} = \mathbf{\Sigma}_{jgk}\mathbf{y}_{jgk}^{\mathrm{p}} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1},$$

where $\mathbf{\Sigma}_{jgk} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$ is denoted as

$$\mathbf{\Sigma}_{jgk} = \mathbf{D}_{jjg}\left(\sum_{l=1}^{L}\sum_{g'=1}^{G}\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jlg'}\mathbf{U}_{jjg} \times\frac{||[\mathbf{C}_{jlgg'}]_k||^2}{\rho^2} + \frac{1}{\rho}\mathbf{I}_{|\mathcal{G}_{jg}|}\right)^{-1},$$

and the beamspace channel estimation error is denoted as $\boldsymbol{\xi}_{jgk} = \bar{\mathbf{h}}_{jgk} - \hat{\bar{\mathbf{h}}}_{jgk} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1}$.

For the received signal in (4), after beamforming it is transformed into

$$\begin{aligned}\tilde{\mathbf{y}}_{jg}^{\mathrm{d}} &= \mathbf{U}_{jjg}^H\mathbf{y}_j^{\mathrm{d}} = \bar{\mathbf{H}}_{jjgg}\mathbf{x}_{jg}^{\mathrm{d}} + \sum_{\substack{g'=1\\g'\neq g}}^{G}\bar{\mathbf{H}}_{jjgg'}\mathbf{x}_{jg'}^{\mathrm{d}} \\ &+\sum_{\substack{l=1\\l\neq j}}^{L}\mathbf{U}_{jjg}^H\mathbf{H}_{jl}\mathbf{x}_l^{\mathrm{d}} + \mathbf{U}_{jjg}^H\mathbf{z}_j^{\mathrm{d}} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1},\end{aligned} \tag{98}$$

where $\mathbf{x}_{jg}^{\mathrm{d}} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1}$ is the corresponding data vector of the g-th group. When MMSE receiving of $\tilde{\mathbf{y}}_{jg}^{\mathrm{d}}$ is employed, the resulting data symbol for the k-th MS in the g-th group in the j-th cell is denoted as $y_{jgk}^{\mathrm{d}} = \mathbf{a}_{jgk}^H\tilde{\mathbf{y}}_{jg}^{\mathrm{d}}$, where $\mathbf{a}_{jgk} = \tilde{\mathbf{\Xi}}_{jgk}^{-1}\hat{\bar{\mathbf{h}}}_{jgk} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1}$ is the MMSE receiving vector, and

$$\begin{aligned}\tilde{\mathbf{\Xi}}_{jgk} &= \sum_{n\neq k}\hat{\bar{\mathbf{h}}}_{jgn}\hat{\bar{\mathbf{h}}}_{jgn}^H + \frac{1}{\rho}\mathbf{\Xi}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}, \\ \mathbf{\Xi}_{jg} &= \mathbb{E}\{\tilde{\boldsymbol{\xi}}_{jg}\tilde{\boldsymbol{\xi}}_{jg}^H\} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}, \\ \tilde{\boldsymbol{\xi}}_{jg} &= \sum_{k=1}^{|\mathcal{G}_{jg}|}\boldsymbol{\xi}_{jgk}[\mathbf{x}_{jg}^{\mathrm{d}}]_k + \sum_{\substack{g'=1\\g'\neq g}}^{G}\bar{\mathbf{H}}_{jjgg'}\mathbf{x}_{jg'}^{\mathrm{d}} \\ &\quad+\sum_{l\neq j}\mathbf{U}_{jjg}^H\mathbf{H}_{jl}\mathbf{x}_l^{\mathrm{d}} + \mathbf{U}_{jjg}^H\mathbf{z}_j^{\mathrm{d}} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times 1}.\end{aligned}$$

By comparing the results in this section to that in Section II, it can be seen that they are similar. In fact, with per-group processing, each group can be taken as a separate cell. Correspondingly, the sum rate upper bound can be adjusted from (39) in Section III according to this per-group processing and is denoted as (52), where

$$\begin{aligned}\bar{\gamma}_{jgk} &= [\mathbf{R}_{jg}]_{k,k} - \frac{1}{|\mathcal{G}_{jg}|[\mathbf{R}_{jg}^{-1}]_{k,k}} \\ &\quad\times\frac{\mathrm{tr}\{\mathbf{R}_{jg(-k,-k)}\}}{1+\mathrm{tr}\{\mathbf{R}_{jg(-k,-k)}\}}, \\ \mathbf{R}_{jg} &= \sum_{(l,g')\neq(j,g)}\mathbf{X}_{jlgg'}\circ\mathbf{C}_{jlgg'}^H\mathbf{C}_{jlgg'} + \mathbf{Y}_{jg},\end{aligned}$$

$$[\mathbf{X}_{jlgg'}]_{s,t} = \frac{1}{\rho}\mathrm{tr}\{\mathbf{U}_{jjg}\mathbf{\Sigma}_{jgs}^H\mathbf{\Xi}_{jg}^{-1}\mathbf{\Sigma}_{jgt}\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jlg'}\},$$
$$[\mathbf{Y}_{jg}]_{s,t} = \delta(s-t)\rho\mathrm{tr}\{\mathbf{D}_{jjg}\mathbf{\Sigma}_{jgs}^H\mathbf{\Xi}_{jg}^{-1}\mathbf{\Sigma}_{jgs}\}$$
$$+\delta(s-t)\mathrm{tr}\{\mathbf{\Sigma}_{jgs}^H\mathbf{\Xi}_{jg}^{-1}\mathbf{\Sigma}_{jgs}\},$$
$$\mathbf{\Xi}_{jg} = \rho\mathbf{D}_{jjg}(|\mathcal{G}_{jg}|\mathbf{I}_{|\mathcal{G}_{jg}|} - \sum_{k=1}^{|\mathcal{G}_{jg}|}\mathbf{\Sigma}_{jgk}^H)$$
$$+\rho|\mathcal{G}_{jg}|\sum_{(l,g')\neq(j,g)}\mathbf{U}_{jjg}^H\tilde{\mathbf{W}}_{jlg'}\mathbf{U}_{jjg} + \mathbf{I}_{|\mathcal{G}_{jg}|},$$

where $\mathbf{X}_{jlgg'} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, $\mathbf{Y}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$, $\mathbf{R}_{jg} \in \mathbb{C}^{|\mathcal{G}_{jg}|\times|\mathcal{G}_{jg}|}$.

ACKNOWLEDGMENT

The author would like to thank Dr. Zhirui Hu for helpful discussions.

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] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing–the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[3] A. Liu and V. Lau, "Two-stage subspace constrained precoding in massive MIMO cellular systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 16, pp. 3271–3279, Jun. 2015.

[4] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, Jun. 2014.

[5] T. V. Chien, C. Mollén, and E. Björnson, "Large-Scale-Fading Decoding in Cellular Massive MIMO Systems With Spatially Correlated Channels," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2746–2762, Apr. 2019.

[6] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, "On design of non-orthogonal pilot signals for a multi-cell massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 129–132, Apr. 2015.

[7] W. Zhang and W. Zhang, "On optimal training in massive MIMO systems with insufficient pilots," in *Proc. of IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.

[8] B. Gong, L. Gui, Q. Qin, X. Ren, and W. Chen, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9149–9161, Oct. 2017.

[9] X. Guo, S. Chen, J. Zhang, X. Mu, and L. Hanzo, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[10] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.

[11] H. Al-Salihi, T. V. Chien, T. A. Le, and M. R. Nakhai, "A successive optimization approach to pilot design for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1086–1089, May 2018.

[12] A. Hu, T. Lv, H. Gao, Y. Lu, and E. Liu, "Pilot design for large-scale multi-cell multiuser MIMO systems," in *Proc. 2013 IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 1–5.

[13] S. Ni, J. Zhao, and Y. Gong, "Optimal pilot design in massive MIMO systems based on channel estimation," *IET Commun.*, vol. 11, no. 7, pp. 975–984, Apr. 2017.

[14] J. Xu, P. Zhu, J. Li, and X. You, "Deep learning-based pilot design for multi-user distributed massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1016–1019, Aug. 2019.

[15] H. Al-Salihi, T. V. Chien, T. A. Le, and M. R. Nakhai, "A successive optimization approach to pilot design for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1086–1089, May 2018.

[16] N. Akbar, N, Yang, P. Sadeghi, and R. A. Kennedy, "Multi-cell multiuser massive MIMO networks: User capacity analysis and pilot design," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5064–5077, Dec. 2016.

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

[18] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1644–1647, Sep. 2015.

[19] X. Zhu, L. Dai, and Z. Wang, "Graph coloring based pilot allocation to mitigate pilot contamination for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1842–1845, Oct. 2015.

[20] R. Mochaourab, E. Björnson, and M. Bengtsson, "Pilot clustering in asymmetric massive MIMO networks," in *Proc. of IEEE 16th Int. Workshop Signal Process. Advances Wireless Commun.*, Stockholm, Sweden, Jun. 2015, pp. 231–235.

[21] H. Ahmadi, A. Farhang, N. Marchetti, and A. MacKenzie, "A game theoretic approach for pilot contamination avoidance in massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 12–15, Feb. 2016.

[22] N. Akbar, S. Yan, N. Yang, and J. Yuan, "Location-aware pilot allocation in multicell multiuser massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7774–7778, Aug. 2018.

[23] A. Hu and H. Wang, "Single branch search based pilot allocation for multi-cell massive multiple-input multiple-output systems," *IET Commun.*, vol. 11, no. 5, pp. 726–732, Mar. 2017.

[24] T. V. Chien, T. N. Canh, E. Björnson, and E. G. Larsson, "Power control in cellular massive MIMO with varying user activity: A deep learning solution," May 2019. [Online]. Available:https://arxiv.org/abs/1901.03620.

[25] C. D'Andrea, A. Zappone, S. Buzzi, and M. Debbah, "Uplink power control in cell-free massive MIMO via deep learning," Aug. 2019. [Online]. Available:https://arxiv.org/abs/1908.11121.

[26] H. Q. Ngo, M. Matthaiou, and E. G. Larssson, "Performance analysis of large scale MU-MIMO with optimal linear receivers," in *Proc. of 2012 Swedish Commun. Tech. Workshop*, Lund, Sweden, Oct. 2012, pp. 59–64.

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

[28] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3–4, pp. 154–655.

[29] K. Mardia, J. T. Kent, and J. Bibby, *Multivariate Analysis*. San Diego, CA: Academic, 1979.

[30] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.

[31] G. H. Song, J. Brady, and A. M. Sayeed, "Beamspace MIMO Transceivers for low-complexity and near-optimal communication at mm-wave frequencies," in *Proc. of 2013 Int. Conf. Acoust., Speech, and Signal Process.*, Vancouver, Canada, May 2013, pp. 4394–4398.

[32] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[33] A. Hu, "Antenna tilt adaptation for multi-cell massive MIMO systems," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2436–2439, Nov. 2017.

[34] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz. 3rd Gener. Partnership Project, Tech. Rep. 38.901 V14.3.0 (2017).

Anzhong Hu (S'13-M'17) received the B.Eng. in communication engineering from Zhejiang University of Technology (ZJUT), Hangzhou, China, in 2009, and the Ph.D. in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. He joined Hangzhou Dianzi University (HDU), Hangzhou, China, in 2014, where he is currently an associate professor. From 2019 to 2020, he was a visiting researcher with Chalmers University of Technology (CTH), Gothenburg, Sweden. His research expertise include signal processing, system design, and parameter optimization in massive multiple-input multiple output (MIMO) systems and millimeter wave (mmWave) systems.