Uplink transmission design for crowded correlated cell-free massive MIMO-OFDM systems

Junyuan GAO[1], Yongpeng WU[1*], Yongjian WANG[2*], Wenjun ZHANG[1] & Fan WEI[1]

[1]*The Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;*
[2]*The National Computer Network Emergency Response Technical Team/Coordination Center of China,*
Beijing 100029, China

Abstract In cell-free massive multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems, user equipments (UEs) are served by many distributed access points (APs), where channels are correlated owing to finite angle-delay spread in realistic outdoor wireless propagation environments. Meanwhile, the number of UEs is growing rapidly for future fully networked society. In this paper, we focus on the uplink transmission design in crowded correlated cell-free massive MIMO-OFDM systems with a limited number of orthogonal pilots. For the pilot transmission phase, we identify active UEs based on non-orthogonal pilot phase shift hopping patterns and non-orthogonal adjustable phase shift pilots (APSP). We derive a closed-form expression of mean square error of channel estimation (MSE-CE) and obtain an optimal condition for minimizing MSE-CE. According to this condition, the APSP set allocation scheme is proposed. Furthermore, for the data transmission, the max-min power control algorithm is devised to maximize the minimum spectral efficiency (SE) lower bound among active UEs. Simulation results indicate significant performance gains in terms of MSE-CE for the proposed APSP set allocation scheme. The proposed power control scheme can further improve the minimum SE among active UEs. Hence, they are crucial for crowded correlated cell-free massive MIMO-OFDM systems.

Keywords APSP set allocation, cell-free massive MIMO-OFDM, correlated channels, crowded scenarios, power control

1 Introduction

Massive multiple-input multiple-output (MIMO) has attracted great research interest in the past few decades and has been a fundamental structure in communication systems. Based on the antenna array deployment, massive MIMO architectures can be divided into co-located and distributed architectures. In co-located massive MIMO systems, the only base station (BS) serves all user equipments (UEs) in a cell [1]. It is recognized as a key component for the fifth-generation (5G) wireless communication networks, which can increase spectral efficiency (SE) and energy efficiency (EE) with simple signal processing [2]. However, these benefits are primarily enjoyed by cell-center UEs, but the performance of cell-edge UEs is usually limited by inter-cell interference. For distributed massive MIMO systems, access points (APs) equipped with single or multiple antennas are spread out over a large area [3]. Cell-free massive MIMO is a special implementation of distributed massive MIMO, where many APs cooperate to jointly serve UEs and UEs experience no inter-cell interference during data downlink transmission, and hence the performance of cell-edge UEs can be greatly improved [4]. Meanwhile, APs are connected to the central processing unit (CPU) via a fronthaul network. Thus, compared with co-located systems, cell-free massive MIMO needs more deployment costs and fronthaul overhead. However, cell-free massive MIMO systems have higher coverage probability since they can provide macro-diversity owing to the joint coherent transmission from APs in a distributed topology, which is the key motivation for the study of cell-free massive MIMO [5].

---

* Corresponding author (email: yongpeng.wu@sjtu.edu.cn, wyj@cert.org.cn)

In order to further improve the system performance, cell-free massive MIMO is expected to be combined with orthogonal frequency division multiplexing (OFDM), which is a multi-carrier modulation technology with high data rate and high robustness to channel frequency selectivity [6].

As we step into the fully networked society, many use cases are emerging, like the Internet of Things, social networking, and video streaming [7,8]. The massive machine-type communication (mMTC) has been one of the most representative services for future wireless communication systems. It aims to support massive connectivity of UEs which transmit packets in an intermittent pattern [8]. Since the cell-free massive MIMO-OFDM can offer spatial degrees of freedom and macro-diversity, it can be utilized in crowded scenarios to improve the connectivity reliability of UEs.

Channel state information (CSI) is crucial in wireless communication systems. Crowded scenarios pose new challenges in the pilot-based acquisition of CSI owing to two reasons. First, the number of orthogonal pilots is limited by power budget and coherence interval. It is difficult to assign a dedicated pilot to each UE for channel estimation [9,10]. Second, UEs are sporadically active [11]. In this case, pilot contamination becomes the performance bottleneck, and how to efficiently reduce channel estimation error and identify active UEs becomes an important topic.

1.1 Motivation and related work

Nowadays, many studies focus on the study of cell-free massive MIMO. For example, it was verified that cell-free massive MIMO can take advantage of the basic properties of co-located massive MIMO, including channel hardening and favorable propagation, when the number of antennas per AP was large and UEs were spatially well separated in independent and identically distributed (i.i.d.) fading channels [12]. In [5,13], the uplink and downlink SE of cell-free massive MIMO were derived with i.i.d. small-scale fading channels. In this case, cell-free massive MIMO can significantly outperform the small-cell operation in terms of the 95%-likely per-user throughput [13], and can outperform the conventional single-cell massive MIMO in terms of the SE of cell-edge UEs [4,14]. The uplink SE of different cell-free implementations is analyzed with spatially correlated fading channels [15]. Taking the fronthaul power consumption into account, the EE was optimized via power control in i.i.d. fading channels [16].

Some papers also study the pilot assignment in cell-free massive MIMO systems, but the vast majority of them consider i.i.d. fading channels and orthogonal pilot set. The most straightforward scheme is a random pilot assignment (RPA), where each UE is randomly assigned a pilot from the orthogonal pilot set. It can lead to serious pilot contamination owing to less consideration of the minimum distance among copilot UEs [17]. In [5], a greedy pilot assignment scheme was utilized, where pilots were randomly allocated to UEs and then iteratively updated to improve the lowest rate among UEs. A structured pilot assignment (SPA) scheme was proposed in [18] to maximize the minimum distance among copilot UEs. A tabu-based scheme was utilized to assign each UE a pilot from an orthogonal pilot set [19]. An efficient pilot assignment scheme based on graph coloring was proposed to mitigate the severe pilot contamination [20].

To mitigate severe pilot contamination, most existing studies consider how to assign a pilot for each UE from the orthogonal pilot set in i.i.d. fading channels. However, in crowded scenarios where UEs are active in an intermittent pattern, it is unnecessary and inefficient to allocate each UE with a pilot from the orthogonal pilot set. Besides, in realistic outdoor wireless propagation environments, the channel angle-delay spread is limited owing to the finite number of scattering clusters, i.e., practical channels between antennas of an AP and a UE over all subcarriers are spatially and frequently correlated [11]. Hence, how to design an uplink transmission scheme to identify active UEs, reduce channel estimation error, and efficiently transmit data in crowded cell-free massive MIMO-OFDM systems with spatially and frequently correlated channels becomes an important issue.

1.2 Contributions and organization

Motivated by the aforementioned discussion, we consider the uplink pilot and data transmission in crowded cell-free massive MIMO-OFDM systems with spatially and frequently correlated channels. The main contributions of this paper are as follows.

• We establish a comprehensive channel model with spatial and frequency correlation between antennas of an AP and a UE over all subcarriers resulting from finite angle-delay spread. We formulate the relationship between the space-frequency domain channel covariance matrix and the angle-delay domain channel power spectrum for cell-free massive MIMO-OFDM systems.
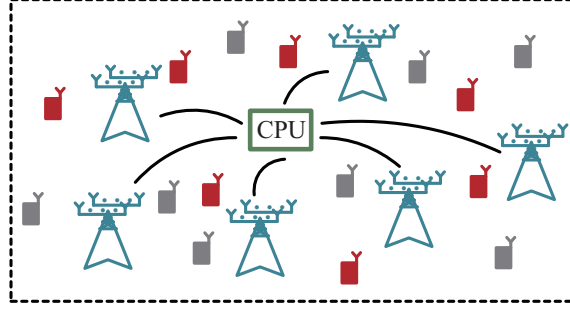
Figure 1 (Color online) Cell-free massive MIMO system. Active and inactive UEs are in red and gray, respectively.

• We achieve the identification of active UEs based on non-orthogonal pilot sequences and non-orthogonal pilot phase shift hopping patterns.

• We derive the expressions of mean square error (MSE) of channel estimation (MSE-CE) and obtain the condition of minimizing MSE-CE. Based on this condition, we propose the adjustable phase shift pilot (APSP) set allocation scheme. Simulation results indicate its performance gain over other schemes.

• We devise a max-min power control algorithm to maximize the minimum SE lower bound of active UEs. We obtain the globally optimal solutions by iteratively solving linear programs.

The rest of this paper is organized as follows. Section 2 describes the system model for cell-free massive MIMO-OFDM systems. Section 3 presents the channel estimation, the identification of active UEs, and the APSP set allocation scheme. The expressions of SE and a lower bound of SE are derived in Section 4. Here, the power control scheme is developed. Simulation results and discussion are given in Section 5. Section 6 concludes the work.

Notation. Let uppercase and lowercase boldface letters denote matrices and column vectors, respectively. Let $[\boldsymbol{A}]_{m,n}$, $[\boldsymbol{A}]_{m,:}$, and $[\boldsymbol{A}]_{:,n}$ denote the (m,n)-th element, the m-th row vector, and the n-th column vector of matrix $\boldsymbol{A}$, respectively. Let $\mathbf{I}_{N \times N}$ denote the $N \times N$ dimensional identity matrix, and $\mathbf{I}_{N \times G}$ denote the matrix comprising the first $G\,(\leqslant N)$ columns of $\mathbf{I}_{N \times N}$. Let $\mathbf{1}_{N \times 1}$ denote all-one vector, and $\mathbf{0}$ denote all-zero vector or matrix. $\|\cdot\|$ denotes the Euclidean norm. The conjugate, transpose, and conjugate transpose are denoted by $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$, respectively. $\mathrm{E}\{\cdot\}$ denotes the expectation operator. The notation vec, $\otimes$, and $\odot$ denote vectorization, Kronecker product, and Hadamard product, respectively. $\backslash$ denotes the set subtraction. $|\mathcal{A}|$ denotes the cardinal of set $\mathcal{A}$. Modulo-N is denoted by $\langle\cdot\rangle_N$. $\lfloor x \rfloor$ denotes the largest integer not greater than x. $z \sim \mathcal{CN}(0,\sigma^2)$ denotes a circularly symmetric complex Gaussian random variable (RV) with mean 0 and variance σ^2. Let

$$\delta(x) = \begin{cases} 1, \ x = 0, \\ 0, \ x \neq 0. \end{cases}$$

Let $\mathrm{diag}\,\{\boldsymbol{a}\}$ denote a diagonal matrix with vector $\boldsymbol{a}$ comprising its diagonal elements, $\mathrm{diag}\,\{\boldsymbol{A}, \boldsymbol{B}\}$ denote a block diagonal matrix with $\boldsymbol{A}$ and $\boldsymbol{B}$ in diagonal blocks.

2 System model

We consider an uplink cell-free massive MIMO-OFDM system where L APs serve a maximal number of K single-antenna UEs in the same time-frequency resource under the time-division duplex (TDD) mode. APs and UEs are randomly located over a large area. We assume each AP is equipped with two uniform linear arrays (ULAs). Each ULA comprises of N directional antennas and receives waves with the incidence angle in a range of 180 degrees to avoid ambiguities in the array manifold. Let $M = NL$. The sets of APs and UEs are denoted by $\mathcal{L} = \{0, 1, \ldots, L-1\}$ and $\mathcal{K} = \{0, 1, \ldots, K-1\}$, respectively. The set of active UEs is denoted as $\mathcal{K}_a$ with the size of $K_a \leqslant K$. APs are connected to a CPU via an error-free fronthaul network, as shown in Figure 1.

We adopt OFDM with N_c subcarriers to combat time dispersive channels. The sampling duration is T_s. Let $T_{\mathrm{sym}} = (N_c + N_{\mathrm{cp}})\,T_s$ and $T_c = N_c T_s$ denote the OFDM symbol duration with and without cyclic prefix (CP), where N_{cp} denotes the number of symbols in CP. The CP duration $T_{\mathrm{cp}} = N_{\mathrm{cp}} T_s$ is assumed to be longer than the maximum channel delay of UEs.

2.1 Channel model in space-frequency domain

In realistic outdoor wireless propagation environments, the channel angle spread is limited owing to finite scattering clusters. Hence, we assume two ULAs in an AP separately receive waves from UEs with the mean angle of arrival (AoA) in a range of 180 degrees, i.e., the waves from a UE are received by directional antennas of a ULA in each AP. Let $\boldsymbol{g}_{k,l,s}^{\beta}$ denote the uplink channel response vector between the l-th AP and the k-th UE over the s-th subcarrier. Let θ denote the incidence angle. $\boldsymbol{g}_{k,l,s}^{\beta}$ can be modeled as [21–23]

$$
\boldsymbol{g}_{k,l,s}^{\beta} \triangleq \sqrt{\beta_{k,l}}\,\boldsymbol{g}_{k,l,s} = \sum_{q=0}^{N_{\mathrm{cp}}-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \boldsymbol{v}_N(\theta) \exp\left\{-\jmath 2\pi \frac{s}{T_c}\tau\right\} a_{k,l}^{\beta}(\theta,\tau)\delta\left(\tau - qT_s\right) \mathrm{d}\theta \in \mathbb{C}^{N\times 1}, \tag{1}
$$

where $\beta_{k,l}$ denotes the large-scale fading between the l-th AP and the k-th UE, $\boldsymbol{g}_{k,l,s}$ denotes the small-scale fading, and $a_{k,l}^{\beta}(\theta,\tau) \triangleq \sqrt{\beta_{k,l}}a_{k,l}(\theta,\tau)$ represents the angle-delay domain channel gain function corresponding to the incidence angle θ and delay τ. If AP antennas are spaced with half of wavelength, the steering vector $\boldsymbol{v}_N(\theta) = [1, \mathrm{e}^{-\jmath \pi \sin\theta}, \ldots, \mathrm{e}^{-\jmath \pi(N-1)\sin\theta}]^{\mathrm{T}}$. The space-frequency domain channel response matrix between UE k and all APs over N_c subcarriers is denoted as

$$
\boldsymbol{G}_k^{\beta} \triangleq \begin{bmatrix} \boldsymbol{G}_{k,0}^{\beta} \\ \boldsymbol{G}_{k,1}^{\beta} \\ \vdots \\ \boldsymbol{G}_{k,L-1}^{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_{k,0,0}^{\beta} & \boldsymbol{g}_{k,0,1}^{\beta} & \cdots & \boldsymbol{g}_{k,0,N_c-1}^{\beta} \\ \boldsymbol{g}_{k,1,0}^{\beta} & \boldsymbol{g}_{k,1,1}^{\beta} & \cdots & \boldsymbol{g}_{k,1,N_c-1}^{\beta} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_{k,L-1,0}^{\beta} & \boldsymbol{g}_{k,L-1,1}^{\beta} & \cdots & \boldsymbol{g}_{k,L-1,N_c-1}^{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_{k,0}^{\beta} & \boldsymbol{g}_{k,1}^{\beta} & \cdots & \boldsymbol{g}_{k,N_c-1}^{\beta} \end{bmatrix} \in \mathbb{C}^{M\times N_c}, \tag{2}
$$

where $\boldsymbol{G}_{k,l}^{\beta} \in \mathbb{C}^{N\times N_c}$ and $\boldsymbol{g}_{k,s}^{\beta} \in \mathbb{C}^{M\times 1}$ are the space-frequency domain channel response matrix between the l-th AP and the k-th UE over N_c subcarriers, and the vector between all APs and the k-th UE over the s-th subcarrier, respectively. It is assumed that $\mathrm{vec}\{\boldsymbol{G}_k^{\beta}\} \sim \mathcal{CN}(0, \boldsymbol{R}_k^{\beta})$.

In this study, we consider the space-frequency domain correlation between the antennas of an AP and a UE over N_c subcarriers. We assume different UEs or APs are spatially separated by a few wavelengths, and the channel realizations between different UEs and an AP and the channel realizations between a UE and different APs are uncorrelated [24]. Besides, we assume the angle-delay domain channel gain function $a_{k,l}(\theta,\tau)$ is uncorrelated [12,22], i.e.,

$$
\mathrm{E}\left\{a_{k,l}(\theta,\tau)a_{k',l'}^{*}(\theta',\tau')\right\} = P_{k,l}^{A}(\theta)P_{k,l}^{D}(\tau)\delta\left(k-k'\right)\delta\left(l-l'\right)\delta\left(\theta-\theta'\right)\delta\left(\tau-\tau'\right), \tag{3}
$$

where $P_{k,l}^{A}(\theta)$ and $P_{k,l}^{D}(\tau)$ represent the channel power azimuth spectrum (PAS) and power delay spectrum (PDS), respectively. Let $\varsigma_{k,l}$, $\theta_{k,l}$, and $\zeta_{k,l}$ denote the angle spread, mean AoA, and delay spread between the k-th UE and the l-th AP, respectively. PAS and PDS are given by [25]

$$
P_{k,l}^{A}(\theta) \propto \exp\left\{-\sqrt{2}\,|\theta - \theta_{k,l}|/\varsigma_{k,l}\right\}, \text{ for } \theta, \theta_{k,l} \in [-\pi/2, \pi/2], \tag{4}
$$

$$
P_{k,l}^{D}(\tau) \propto \exp\left\{-\tau/\zeta_{k,l}\right\}, \text{ for } \tau \in [0, (N_{\mathrm{cp}}-1)T_s]. \tag{5}
$$

We assume the channel angle spreads between different UEs and APs are equal, i.e., $\varsigma_{k,l} = \varsigma$ for $k \in \mathcal{K}$ and $l \in \mathcal{L}$. Similarly, the delay spreads satisfy $\zeta_{k,l} = \zeta$ for $k \in \mathcal{K}$ and $l \in \mathcal{L}$.

2.2 Channel model in angle-delay domain

Let $\boldsymbol{H}_{k,l}^{\beta} \triangleq \sqrt{\beta_{k,l}}\boldsymbol{H}_{k,l} \in \mathbb{C}^{N\times N_{\mathrm{cp}}}$ denote the angle-delay domain channel response matrix between the k-th UE and the l-th AP. Let $\boldsymbol{H}_k^{\beta} \triangleq [(\boldsymbol{H}_{k,0}^{\beta})^{\mathrm{T}}(\boldsymbol{H}_{k,1}^{\beta})^{\mathrm{T}}\cdots(\boldsymbol{H}_{k,L-1}^{\beta})^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{H}_k \triangleq [\boldsymbol{H}_{k,0}^{\mathrm{T}}\ \boldsymbol{H}_{k,1}^{\mathrm{T}}\cdots\boldsymbol{H}_{k,L-1}^{\mathrm{T}}]^{\mathrm{T}}$ denote the angle-delay domain channel response matrices between UE k and all APs with and without large-scale fading, respectively. Different elements in $\boldsymbol{H}_k^{\beta}$ correspond to the channels for different incidence angles, delays or APs, which satisfy [12,22]

$$
\mathrm{E}\left\{\left[\boldsymbol{H}_k^{\beta}\right]_{m,q}\left[\boldsymbol{H}_k^{\beta}\right]_{m',q'}^{*}\right\} = \delta\left(m-m'\right)\delta\left(q-q'\right)\left[\boldsymbol{\Upsilon}_k^{\beta}\right]_{m,q}, \tag{6}
$$

where $\boldsymbol{\Upsilon}_k^\beta \triangleq [(\boldsymbol{\Upsilon}_{k,0}^\beta)^{\mathrm{T}} (\boldsymbol{\Upsilon}_{k,1}^\beta)^{\mathrm{T}} \cdots (\boldsymbol{\Upsilon}_{k,L-1}^\beta)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{M \times N_{\mathrm{cp}}}$ is the angle-delay domain channel power spectrum between UE k and all APs. We define $\boldsymbol{\Upsilon}_k \triangleq [\boldsymbol{\Upsilon}_{k,0}^{\mathrm{T}} \ \boldsymbol{\Upsilon}_{k,1}^{\mathrm{T}} \cdots \boldsymbol{\Upsilon}_{k,L-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{M \times N_{\mathrm{cp}}}$.

Wireless channels are sparse in many typical scenarios [26, 27]. In this study, we consider the angle-delay domain sparsity from a UE to all APs in a cell-free massive MIMO-OFDM system. Most elements in $\boldsymbol{\Upsilon}_k^\beta$ are approximately 0 owing to finite angle-delay spread.

2.3 Relationship between space-frequency domain and angle-delay domain channels

Since the channel power lies in a finite number of delays owing to limited scattering, channels are frequently correlated [22]. Considering most channel power is concentrated in a finite region of angles, channels between the antennas of an AP and a UE are spatially correlated [27]. Proposition 1 shows the relationship between the space-frequency domain channel covariance matrix and the angle-delay domain channel power spectrum in cell-free massive MIMO-OFDM systems.

Proposition 1. Define $\theta_n = \arcsin\left(\frac{2n}{N} - 1\right)$ for $n = 0, 1, \ldots, N$. Let $[\boldsymbol{V}_{N \times N}]_{i,j} \triangleq \frac{1}{\sqrt{N}} \exp\left\{-\overline{\jmath}\pi i \sin\theta_n\right\}$, $\boldsymbol{V}_{M \times M} \triangleq \boldsymbol{I}_{L \times L} \otimes \boldsymbol{V}_{N \times N}$, and $\left[\boldsymbol{F}_{N_c \times N_{\mathrm{cp}}}\right]_{s,q} \triangleq \frac{1}{\sqrt{N_c}} \exp\{-\overline{\jmath}2\pi\frac{s}{N_c}q\}$. We have $\boldsymbol{V}_{N \times N}^{\mathrm{H}} \boldsymbol{V}_{N \times N} \stackrel{N \to \infty}{=} \boldsymbol{I}_{N \times N}$ and $\boldsymbol{V}_{M \times M}^{\mathrm{H}} \boldsymbol{V}_{M \times M} \stackrel{N \to \infty}{=} \boldsymbol{I}_{M \times M}$. The channel power spectrum between the k-th UE and the l-th AP is denoted as $\boldsymbol{\Upsilon}_{k,l}^\beta \in \mathbb{R}^{N \times N_{\mathrm{cp}}}$ satisfying

$$\left[\boldsymbol{\Upsilon}_{k,l}^\beta\right]_{n,q} \triangleq \beta_{k,l} \left[\boldsymbol{\Upsilon}_{k,l}\right]_{n,q} = \beta_{k,l} N N_c \left(\theta_{n+1} - \theta_n\right) P_{k,l}^A(\theta_n) P_{k,l}^D(\tau_q), \tag{7}$$

where $\tau_q = qT_s$ for $q = 0, 1, \ldots, N_{\mathrm{cp}} - 1$. The relationship between the space-frequency domain channel covariance matrix and angle-delay domain channel power spectrum is given by

$$\boldsymbol{R}_k^\beta \stackrel{N \to \infty}{=} \left(\boldsymbol{F}_{N_c \times N_{\mathrm{cp}}} \otimes \boldsymbol{V}_{M \times M}\right) \mathrm{diag}\left\{\mathrm{vec}\left\{\boldsymbol{\Upsilon}_k^\beta\right\}\right\} \left(\boldsymbol{F}_{N_c \times N_{\mathrm{cp}}} \otimes \boldsymbol{V}_{M \times M}\right)^{\mathrm{H}} \in \mathbb{C}^{MN_c \times MN_c}. \tag{8}$$

The space-frequency domain channel response matrix $\boldsymbol{G}_k^\beta$ can be decomposed as

$$\boldsymbol{G}_k^\beta \stackrel{N \to \infty}{=} \boldsymbol{V}_{M \times M} \boldsymbol{H}_k^\beta \ \boldsymbol{F}_{N_c \times N_{\mathrm{cp}}}^{\mathrm{T}}. \tag{9}$$

Proof. See Appendix A.

When N is sufficiently large, the eigenvectors of space-frequency domain channel covariance matrices are independent of the locations of UEs and only depend on the AP antenna configurations. Eigenvalues can be approximated by the angle-delay domain channel power spectrum. Specifically, when ULAs are employed at each AP with antenna spacing of half-wavelength, $\boldsymbol{V}_{N \times N}$ can be set to discrete Fourier transform (DFT) matrices with some matrix elementary operations. Since the dimension of $\boldsymbol{\Upsilon}_k^\beta$ is much smaller than that of $\boldsymbol{R}_k^\beta$ and most elements in $\boldsymbol{\Upsilon}_k^\beta$ are approximately 0, we will estimate angle-delay domain channel parameters at first, and then obtain space-frequency domain channel parameters via (9). We assume angle-delay domain channel power spectrums are known by APs.

Proposition 1 is a generalization of existing results. It agrees with the results in [27] where channels are frequency-flat fading on a narrow-band subcarrier. It is consistent with the results in [23], where correlated channels in co-located systems are considered. The approximation is shown to be accurate enough when the number of antennas at the BS ranges from 64 to 512 in frequency-flat co-located massive MIMO channels [28].

3 Pilot transmission design with pilot assignment

UEs are sporadically active and far more than orthogonal pilots in crowded scenarios, which calls for proper design of uplink transmission scheme and pilot assignment scheme. We utilize APSP to increase the number of unique pilots as shown in Subsection 3.1. Besides, active UEs can be identified through pilot shift hopping patterns as introduced in Subsection 3.2. To better utilize channel characteristics and improve MSE-CE in correlated cell-free massive MIMO-OFDM systems, an APSP set allocation scheme is proposed in Subsection 3.3.

3.1 APSP and channel estimation

We assume channels are piece-wise constant over a coherence interval. The time coherence interval is assumed to be equal to the length of Z_a OFDM symbols. The frequency coherence interval is assumed to be equal to $\frac{1}{2\zeta}$. We assume Z OFDM symbols are used for pilot transmission and the remaining $Z_a - Z$ OFDM symbols are for uplink data transmission.

During the training phase, all active UEs simultaneously transmit pilots to APs. The space-frequency domain signal received at the l-th AP is given by

$$\boldsymbol{Y}_l = \sqrt{\rho_p Z} \sum_{k' \in \mathcal{K}_a} \boldsymbol{G}_{k',l}^{\beta} \boldsymbol{\Phi}_{k'} + \boldsymbol{W}_l \in \mathbb{C}^{N \times N_c Z}, \tag{10}$$

where ρ_p is the normalized signal-to-noise ratio (SNR) in the training phase, $\boldsymbol{W}_l$ is the additive white Gaussian noise (AWGN) matrix including i.i.d. $\mathcal{CN}(0,1)$ elements, $\boldsymbol{G}_{k',l}^{\beta} = \boldsymbol{V}_{N \times N} \boldsymbol{H}_{k',l}^{\beta} \boldsymbol{F}_{N_c \times N_{cp}}^{\mathrm{T}}$, and $\boldsymbol{\Phi}_{k'}$ is the pilot signal given in (11). We assume large-scale fading coefficients are known wherever required.

In cell-free massive MIMO-OFDM systems, the number of phase shift orthogonal pilots (PSOPs) is approximately $\lfloor ZN_c/N_{cp} \rfloor$. Their phase shift differences are no less than the maximum channel delay of UEs to promise orthogonality of different pilots [29]. However, pilot contamination can be serious since UEs are far more than PSOPs. Hence, APSP is adopted in this study [23]. The pilot signal for the k-th UE over N_c subcarriers and Z OFDM symbols is

$$\boldsymbol{\Phi}_k \triangleq [\boldsymbol{U}]_{\langle \phi_k \rangle_Z, :} \otimes \left(\boldsymbol{D}_{\lfloor \phi_k/Z \rfloor} \boldsymbol{\Phi}_b \right) \in \mathbb{C}^{N_c \times N_c Z}, \tag{11}$$

where $\boldsymbol{U}$ is an arbitrary $Z \times Z$ dimensional unitary matrix, $\phi_k \in \boldsymbol{\Psi} = \{0, 1, \ldots, N_c Z - 1\}$, $\boldsymbol{D}_i = \mathrm{diag}\{\boldsymbol{f}_{N_c, i}\}$, $\boldsymbol{f}_{N_c, i} = [1 \ \exp\{-\jmath 2\pi \frac{i}{N_c}\} \cdots \exp\{-\jmath 2\pi \frac{(N_c-1)i}{N_c}\}]^{\mathrm{T}}$, and $\boldsymbol{\Phi}_b$ is a diagonal matrix satisfying $\boldsymbol{\Phi}_b \boldsymbol{\Phi}_b^{\mathrm{H}} = \boldsymbol{I}_{N_c \times N_c}$, which is the basic pilot matrix shared by all UEs. Then we have, for $\forall k, k' \in \mathcal{K}$,

$$\boldsymbol{\Phi}_{k'} \boldsymbol{\Phi}_k^{\mathrm{H}} \overset{(a)}{=} \left([\boldsymbol{U}]_{\langle \phi_{k'} \rangle_Z, :} [\boldsymbol{U}]_{\langle \phi_k \rangle_Z, :}^{\mathrm{H}} \right) \otimes \left(\boldsymbol{D}_{\lfloor \phi_{k'}/Z \rfloor} \boldsymbol{D}_{\lfloor \phi_k/Z \rfloor}^{\mathrm{H}} \right) = \delta \left(\langle \phi_{k'} \rangle_Z - \langle \phi_k \rangle_Z \right) \boldsymbol{D}_{\lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor}, \tag{12}$$

where (a) follows from the properties of Kronecker product, i.e., $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A}\boldsymbol{C}) \otimes (\boldsymbol{B}\boldsymbol{D})$ and $(\boldsymbol{A} \otimes \boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}} \otimes \boldsymbol{B}^{\mathrm{H}}$. For the APSP, phase shifts are divided into Z groups. The group index is $\langle \phi_k \rangle_Z$. There is no pilot interference if UEs use phase shifts in different groups.

Based on (10)–(12), after decorrelation, we have

$$\check{\boldsymbol{Y}}_{k,l} = \frac{1}{\sqrt{\rho_p Z \beta_{k,l}}} \boldsymbol{V}_{N \times N}^{\mathrm{H}} \boldsymbol{Y}_l \boldsymbol{\Phi}_k^{\mathrm{H}} \boldsymbol{F}_{N_c \times N_{cp}}^*$$

$$= \boldsymbol{H}_{k,l} + \frac{1}{\sqrt{\beta_{k,l}}} \sum_{k' \in \mathcal{K}_a \backslash \{k\}} \delta \left(\langle \phi_{k'} \rangle_Z - \langle \phi_k \rangle_Z \right) \boldsymbol{H}_{k',l}^{\beta} \boldsymbol{F}_{N_c \times N_{cp}}^{\mathrm{T}} \boldsymbol{D}_{\lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} \boldsymbol{F}_{N_c \times N_{cp}}^* + \frac{1}{\sqrt{\rho_p Z \beta_{k,l}}} \check{\boldsymbol{W}}_{k,l}$$

$$= \boldsymbol{H}_{k,l} + \frac{1}{\sqrt{\beta_{k,l}}} \sum_{k' \in \mathcal{K}_a \backslash \{k\}} \delta \left(\langle \phi_{k'} \rangle_Z - \langle \phi_k \rangle_Z \right) \boldsymbol{H}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} + \frac{1}{\sqrt{\rho_p Z \beta_{k,l}}} \check{\boldsymbol{W}}_{k,l}, \tag{13}$$

where $\check{\boldsymbol{W}}_{k,l}$ includes i.i.d. $\mathcal{CN}(0,1)$ entries owing to unitary transformation. Define the permutation matrix

$$\boldsymbol{\Gamma}_{N_c}^i = \begin{bmatrix} \boldsymbol{0} & \mathbf{I}_{\left(N_c - \langle i \rangle_{N_c}\right) \times \left(N_c - \langle i \rangle_{N_c}\right)} \\ \mathbf{I}_{\langle i \rangle_{N_c} \times \langle i \rangle_{N_c}} & \boldsymbol{0} \end{bmatrix}.$$
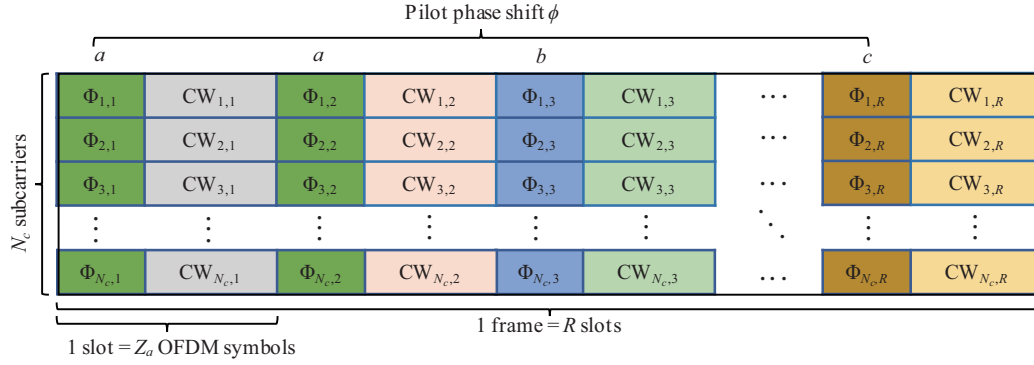
Then, $\boldsymbol{H}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor}$ can be expressed as

$$\boldsymbol{H}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} = \boldsymbol{H}_{k',l}^{\beta} \mathbf{I}_{N_c \times N_{cp}}^{\mathrm{T}} \boldsymbol{\Gamma}_{N_c}^{\lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} \mathbf{I}_{N_c \times N_{cp}}. \tag{14}$$

The power spectrum of (14) is given by

$$\boldsymbol{\Upsilon}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} = \mathrm{E} \left\{ \boldsymbol{H}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} \odot \left(\boldsymbol{H}_{k',l}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} \right)^* \right\}$$

$$= \boldsymbol{\Upsilon}_{k',l}^{\beta} \mathbf{I}_{N_c \times N_{cp}}^{\mathrm{T}} \boldsymbol{\Gamma}_{N_c}^{\lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} \mathbf{I}_{N_c \times N_{cp}}. \tag{15}$$

Pilot phase shift ϕ

Figure 2 (Color online) Illustration of the transmission frame. In this example, the hopping pattern for a given UE is $\{a, a, b, \dots, c\}$. Pilot signals among N_c subcarriers are formulated based on this pattern, and transmitted during the pilot transmission phase. Data codewords are transmitted afterwards.

Let $\boldsymbol{\Upsilon}_{k'}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor} \triangleq [(\boldsymbol{\Upsilon}_{k',0}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor})^{\mathrm{T}} \ (\boldsymbol{\Upsilon}_{k',1}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor})^{\mathrm{T}} \ \cdots \ (\boldsymbol{\Upsilon}_{k',L-1}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor})^{\mathrm{T}}]^{\mathrm{T}}$.

Channels are estimated in a decentralized fashion in cell-free massive MIMO-OFDM systems, i.e., each AP autonomously estimates channels without interchanging information [5]. For an AP, the channel estimation is performed in an element-wise manner in the angle-delay domain. The minimum mean-square error (MMSE) estimate of $[\boldsymbol{H}_{k,l}]_{i,j}$ is

$$\left[\widehat{\boldsymbol{H}}_{k,l}\right]_{i,j} = \frac{[\boldsymbol{\Upsilon}_{k,l}^{\beta}]_{i,j}}{\sum_{k'\in\mathcal{K}_a}\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi_k\rangle_Z\right)[\boldsymbol{\Upsilon}_{k',l}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor}]_{i,j} + \frac{1}{\rho_p Z}}\left[\check{\boldsymbol{Y}}_{k,l}\right]_{i,j}. \tag{16}$$

Let $\widetilde{\boldsymbol{H}}_{k,l} = \boldsymbol{H}_{k,l} - \widehat{\boldsymbol{H}}_{k,l}$ denote the channel estimation error. Based on the property of MMSE estimation, $\widetilde{\boldsymbol{H}}_{k,l}$ is independent of $\widehat{\boldsymbol{H}}_{k,l}$, and the MSE-CE is shown as

$$[\boldsymbol{\Xi}_{k,l}]_{i,j} \triangleq \mathrm{E}\left\{\left[\widetilde{\boldsymbol{H}}_{k,l}\right]_{i,j}\left[\widetilde{\boldsymbol{H}}_{k,l}\right]_{i,j}^{*}\right\} = [\boldsymbol{\Upsilon}_{k,l}]_{i,j} - \frac{[\boldsymbol{\Upsilon}_{k,l}]_{i,j}[\boldsymbol{\Upsilon}_{k,l}^{\beta}]_{i,j}}{\sum_{k'\in\mathcal{K}_a}\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi_k\rangle_Z\right)[\boldsymbol{\Upsilon}_{k',l}^{\beta,\lfloor\phi_{k'}/Z\rfloor-\lfloor\phi_k/Z\rfloor}]_{i,j} + \frac{1}{\rho_p Z}}. \tag{17}$$

Let $\boldsymbol{\Xi}_k^{\beta} \triangleq [\beta_{k,0}\boldsymbol{\Xi}_{k,0}^{\mathrm{T}} \ \beta_{k,1}\boldsymbol{\Xi}_{k,1}^{\mathrm{T}} \ \cdots \ \beta_{k,L-1}\boldsymbol{\Xi}_{k,L-1}^{\mathrm{T}}]^{\mathrm{T}}$.

3.2 UE identification

Pilot contamination is unavoidable in crowded scenarios. In this case, transmitting UEs cannot be identified in a slot, and we have to resort to the information among some slots. In this work, the uplink transmission phase is divided into frames consisting of R slots, as shown in Figure 2. Each slot comprises Z_a OFDM symbols. The phase shifts of pilot signals among R slots, i.e., the hopping patterns, form the identification basis of active UEs. When PSOPs are utilized, the number of UEs that can be identified is $\lfloor ZN_c/N_{cp}\rfloor^R$. It can be further increased if APSPs are adopted owing to channel sparsity. Hence, we assume each UE has a unique and non-orthogonal pseudo-random hopping pattern, which is known by APs and CPU in advance.

Assuming the activation probability of UEs is p_a, the probability of having K_a active UEs among K UEs is given by

$$p(K_a|K) = \binom{K}{K_a}p_a^{K_a}(1-p_a)^{K-K_a}. \tag{18}$$

When a UE is active, it selects the pilot phase shift based on its hopping pattern. Next, it formulates the corresponding pilot over N_c subcarriers based on (11) and transmits this pilot during the pilot transmission phase. Data codewords are transmitted afterward. At the receiver, APs identify hopping patterns by running a correlation decoder across R slots to detect transmitting UEs. Normalized maximum ratio combining (MRC) is applied to the data at APs and the outputs are merged based on the identified hopping patterns. Since APs can know which UEs are active and what they transmit after R slots, this scheme is suitable for delay-tolerant mMTC.

3.3 APSP set allocation

In cell-free massive MIMO-OFDM systems, for a UE, the APs far from it cannot contribute significantly to the channel estimation performance and spatial diversity gains owing to heavy path loss. Besides, the transfer of data between APs and CPU causes much fronthaul energy consumption. Hence, in order to improve channel estimation performance and save fronthaul energy consumption, each UE should be served by a group of APs but not all APs within the network [30]. Hence, before allocating APSP set for each UE, AP selection is performed at CPU [16]. The set of APs selected to serve the k-th UE is denoted as $\mathcal{B}_k \subseteq \mathcal{L}$ satisfying $\sum_{b=0}^{|\mathcal{B}_k|-1} \overline{\beta}_{k,b} \geqslant \sum_{l=0}^{L-1} \beta_{k,l}\lambda$, where $\{\overline{\beta}_{k,0}, \overline{\beta}_{k,1}, \ldots, \overline{\beta}_{k,L-1}\}$ includes large-scale fading coefficients between UE k and APs in descending order, and λ is assumed to be equal for all UEs. The APs belonging to $\mathcal{B}_k$ need to estimate channels and decode signals of UE k. The set of UEs served by AP l is denoted as $\mathcal{D}_l$.

The sporadic and independent activation of UEs and the construction of pseudo-random hopping patterns can be modeled as the process that each active UE randomly selects a pilot phase shift from its allocated set in each slot [11]. We obtain the probability of having K_a active UEs in (18), but it is still uncertain which K_a UEs are active and which phase shifts they select. To quantify these uncertainties, we employ $\mathcal{U}_{\mathcal{K}_a} = \{\mathcal{U}_{\mathcal{K}_a}^0, \mathcal{U}_{\mathcal{K}_a}^1, \ldots, \mathcal{U}_{\mathcal{K}_a}^{N_{\mathcal{K}_a}-1}\}$ to denote all possible sets of K_a active UEs. The j-th element in $\mathcal{U}_{\mathcal{K}_a}^i$ is denoted as $\mathcal{K}_a^{i,j}$. In a slot, all possible choices of phase shifts for UEs in $\mathcal{U}_{\mathcal{K}_a}^i$ are denoted as $\mathcal{P}_i = \{\mathcal{P}_i^0, \mathcal{P}_i^1, \ldots, \mathcal{P}_i^{N_\phi-1}\}$. Assuming the p-th choice of shifts is considered, the phase shift of UE $\mathcal{K}_a^{i,j}$ is denoted as $\mathcal{P}_{i,j}^p$. We assume $\nu_{k,l}$ is equal to 1 when the l-th AP is selected to serve the k-th UE and is equal to 0 otherwise. Let $\boldsymbol{\Upsilon}_k^{\beta,0} \triangleq [\nu_{k,0}\beta_{k,0}\boldsymbol{\Upsilon}_{k,0}^{\mathrm{T}} \; \nu_{k,1}\beta_{k,1}\boldsymbol{\Upsilon}_{k,1}^{\mathrm{T}} \; \cdots \; \nu_{k,L-1}\beta_{k,L-1}\boldsymbol{\Upsilon}_{k,L-1}^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{\Upsilon}_k^0 \triangleq [\nu_{k,0}\boldsymbol{\Upsilon}_{k,0}^{\mathrm{T}} \; \nu_{k,1}\boldsymbol{\Upsilon}_{k,1}^{\mathrm{T}} \; \cdots \; \nu_{k,L-1}\boldsymbol{\Upsilon}_{k,L-1}^{\mathrm{T}}]^{\mathrm{T}}$. Based on the channel estimation analysis in Subsection 3.1, we obtain the MSE-CE of UE $\mathcal{K}_a^{i,j}$ averaged over its serving APs and N_c subcarriers:

$$\varepsilon_{\mathcal{K}_a^{i,j},p}^0 = \frac{1}{N_c|\mathcal{B}_{\mathcal{K}_a^{i,j}}|} \sum_{m=0}^{M-1} \sum_{q=0}^{N_{\mathrm{cp}}-1} \left\{ \left[\boldsymbol{\Upsilon}_{\mathcal{K}_a^{i,j}}^0\right]_{m,q} - \frac{[\boldsymbol{\Upsilon}_{\mathcal{K}_a^{i,j}}^0]_{m,q}[\boldsymbol{\Upsilon}_{\mathcal{K}_a^{i,j}}^{\beta,0}]_{m,q}}{\sum_{j'=0}^{K_a-1} \delta(\langle \mathcal{P}_{i,j'}^p \rangle_Z - \langle \mathcal{P}_{i,j}^p \rangle_Z)[\boldsymbol{\Upsilon}_{\mathcal{K}_a^{i,j'}}^{\beta,\lfloor \mathcal{P}_{i,j'}^p/Z \rfloor - \lfloor \mathcal{P}_{i,j}^p/Z \rfloor}]_{m,q} + \frac{1}{\rho_p Z}} \right\}. \tag{19}$$

We average (19) over all possible sets of active UEs, all UEs in the set, and all types of phase shift selection. We obtain

$$\mathrm{E}_{\mathcal{U},\mathcal{K}_a,\mathcal{P}}(\varepsilon^0) = \sum_{i=0}^{N_{\mathcal{K}_a}-1} \sum_{j=0}^{K_a-1} \sum_{p=0}^{N_\phi-1} \frac{\varepsilon_{\mathcal{K}_a^{i,j},p}^0}{N_{\mathcal{K}_a} K_a N_\phi}. \tag{20}$$

Then we calculate the expected value of (20) accounting for the number of active UEs and obtain

$$\overline{\varepsilon}^0 = \sum_{K_a=1}^{K} p(K_a|K)\mathrm{E}_{\mathcal{U},\mathcal{K}_a,\mathcal{P}}(\varepsilon^0). \tag{21}$$

The channel estimation performance suffers from pilot interference. In Proposition 2, we show that the effect of pilot interference can be eliminated by proper APSP set allocation for UEs. Phase shifts in the APSP set of each UE are used to form the unique pilot phase shift hopping pattern for it. In correlated channels, many elements in $\boldsymbol{\Upsilon}_k^\beta$ are approximately 0, which is the physical basis of the APSP set allocation scheme.

Proposition 2. The minimum value of $\mathrm{E}_{\mathcal{U},\mathcal{K}_a,\mathcal{P}}(\varepsilon^0)$ and $\overline{\varepsilon}^0$ are given by

$$\left[\mathrm{E}_{\mathcal{U},\mathcal{K}_a,\mathcal{P}}(\varepsilon^0)\right]_{\min} = \frac{1}{KN_c} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \sum_{q=0}^{N_{\mathrm{cp}}-1} \frac{1}{|\mathcal{B}_k|} \left([\boldsymbol{\Upsilon}_k^0]_{m,q} - \frac{[\boldsymbol{\Upsilon}_k^0]_{m,q}[\boldsymbol{\Upsilon}_k^{\beta,0}]_{m,q}}{[\boldsymbol{\Upsilon}_k^\beta]_{m,q} + \frac{1}{\rho_p Z}} \right), \tag{22}$$

$$\overline{\varepsilon}_{\min}^0 = \sum_{K_a=1}^{K} p(K_a|K) \left[\mathrm{E}_{\mathcal{U},\mathcal{K}_a,\mathcal{P}}(\varepsilon^0)\right]_{\min}, \tag{23}$$

and the minimum is achieved under the condition that for $\forall k, k' \in \mathcal{K}$ and $k \neq k'$,

$$\delta\left(\langle \phi_{k'} \rangle_Z - \langle \phi_k \rangle_Z\right) \boldsymbol{\Upsilon}_k^0 \odot \boldsymbol{\Upsilon}_k^0 \odot \boldsymbol{\Upsilon}_{k'}^{\beta,\lfloor \phi_{k'}/Z \rfloor - \lfloor \phi_k/Z \rfloor} = \mathbf{0}_{M \times N_{\mathrm{cp}}}. \tag{24}$$

Proof. See Appendix B.

When APSPs are used, phase shifts are divided into Z groups. There is no pilot interference for UEs using shifts in different groups. Channel estimation of a UE is affected by interference from other UEs in the same group exhibiting corresponding cyclic shifts in the delay domain as shown in (13). $\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi_k\rangle_Z\right)\Upsilon_{k'}^{\beta,\lfloor\phi_{k'}/Z\rfloor - \lfloor\phi_k/Z\rfloor}$ is the angle-delay domain channel power spectrum interference from UE k' to UE k exhibiting cyclic shifts. By proper pilot allocation based on (24), there can be no overlapping between $\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi_k\rangle_Z\right)\Upsilon_{k'}^{\beta,\lfloor\phi_{k'}/Z\rfloor - \lfloor\phi_k/Z\rfloor}$ and Υ_k^0, i.e., the effect of pilot interference can be mitigated and $\overline{\varepsilon}^0$ can be minimized. In crowded correlated cell-free massive MIMO-OFDM systems, we can allocate an APSP set for each UE to reduce $\overline{\varepsilon}^0$. This problem can be formulated as

$$
\begin{aligned}
\min_{\mathcal{Y}_k, k\in\mathcal{K}} \quad & \overline{\varepsilon}^0 \\
\text{s.t.} \quad & 0 \leqslant \phi_{k,c} \leqslant N_c Z - 1 \ \text{ for } k\in\mathcal{K}, \ 0\leqslant c \leqslant |\mathcal{Y}| - 1, \\
& \phi_{k,c} \in \mathbb{Z} \ \text{ for } k\in\mathcal{K}, \ 0\leqslant c\leqslant |\mathcal{Y}| - 1,
\end{aligned}
\tag{25}
$$

where $\mathcal{Y}_k = \{\phi_{k,0}, \phi_{k,1}, \ldots, \phi_{k,|\mathcal{Y}|-1}\}$ is the APSP set allocated for the k-th UE.

This problem is combinatorial and can be solved through an exhaustive search with large complexity. Importantly, the problem can be simplified if we aim at satisfying the condition in (24) as much as possible. First, we define $\xi(\boldsymbol{A}, \boldsymbol{B})$ to measure the overlap degree between two non-negative matrices $\boldsymbol{A}, \boldsymbol{B}$. When $\boldsymbol{A} = \boldsymbol{0}$ or $\boldsymbol{B} = \boldsymbol{0}$, $\xi(\boldsymbol{A}, \boldsymbol{B}) \triangleq 0$. When $\boldsymbol{A}, \boldsymbol{B} \neq \boldsymbol{0}$, we have

$$
\xi(\boldsymbol{A}, \boldsymbol{B}) \triangleq \frac{\left|\sum_{i,j}[\boldsymbol{A}\odot\boldsymbol{B}]_{i,j}\right|}{\sqrt{\sum_{i,j}[\boldsymbol{A}]_{i,j}^2}\cdot\sqrt{\sum_{i,j}[\boldsymbol{B}]_{i,j}^2}} \in [0, 1].
\tag{26}
$$

In Algorithm 1, K-means clustering method is utilized to partition K UEs into J clusters. Next, APSP set is allocated for each UE to make the effect of pilot interference from other UEs in the same cluster as small as possible. Algorithm 1 should be performed before pilot transmission, i.e., without the knowledge of UEs' activity. Hence, it is computed at CPU assuming all UEs are active and their statistical CSI is known. Hence, this scheme is insensitive to the active pattern of UEs and does not need to be recomputed frequently.

Algorithm 1 APSP set allocation algorithm

Require: The angle-delay domain channel power spectrum $\{\Upsilon_k : k\in\mathcal{K}\}$; the threshold λ and γ; the phase shift set $\boldsymbol{\Psi}$; the large-scale fading $\boldsymbol{\beta}_k \triangleq [\beta_{k,0}\ \beta_{k,1}\cdots\beta_{k,L-1}]$ for $\forall k\in\mathcal{K}$.
Ensure: The APSP set allocated to each UE $\{\mathcal{Y}_k, k\in\mathcal{K}\}$.
1: Centroids are randomly chosen as $\tilde{\boldsymbol{\beta}}_j \in \mathbb{C}^{L\times 1}$ for $j = 0, 1, \ldots, J-1$;
2: For $\forall k\in\mathcal{K}$, UE k is assigned to the cluster with $\max_{j=0,1,\ldots,J-1}\xi(\boldsymbol{\beta}_k, \tilde{\boldsymbol{\beta}}_j)$;
3: Centroids are updated by averaging over UEs belonging to respective clusters and UEs are reassigned until the assignments no longer change. Let $\mathcal{C}_j$ denote the set of UEs belonging to the j-th cluster;
4: **for** $j = 0, 1, \ldots, J-1$ **do**
5:     $|\mathcal{Y}|$ shifts are randomly chosen to form the set $\mathcal{Y}_{\mathcal{C}_j(0)}$ allocated for UE $\mathcal{C}_j(0)$. Initialize the allocated UE set $\mathcal{K}_j^{\mathrm{al}} = \{\mathcal{C}_j(0)\}$ and the unallocated UE set $\mathcal{K}_j^{\mathrm{un}} = \mathcal{C}_j\backslash\{\mathcal{C}_j(0)\}$.
6:     **for** $k\in\mathcal{K}_j^{\mathrm{un}}$ **do**
7:         Search for $|\mathcal{Y}|$ pilot phase shifts $\phi\in\boldsymbol{\Psi}$ to form the APSP set $\mathcal{Y}_k$ allocated for the k-th UE, which should satisfy
$$\sum_{k'\in\mathcal{K}_j^{\mathrm{al}}}\frac{1}{|\mathcal{K}_j^{\mathrm{al}}|}\max_{\phi_{k'}\in\mathcal{Y}_{k'}}\left\{\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi\rangle_Z\right)\xi\left(\Upsilon_k^0\odot\Upsilon_k^0, \Upsilon_{k'}^{\beta,\lfloor\phi_{k'}/Z\rfloor - \lfloor\phi/Z\rfloor}\right)\right\} \leqslant \gamma;$$
8:         If $|\mathcal{Y}|$ pilot phase shifts are not found in step 7, then search for $|\mathcal{Y}|$ shifts from $\boldsymbol{\Psi}$ corresponding to the $|\mathcal{Y}|$ smallest
$$\sum_{k'\in\mathcal{K}_j^{\mathrm{al}}}\max_{\phi_{k'}\in\mathcal{Y}_{k'}}\left\{\delta\left(\langle\phi_{k'}\rangle_Z - \langle\phi\rangle_Z\right)\xi\left(\Upsilon_k^0\odot\Upsilon_k^0, \Upsilon_{k'}^{\beta,\lfloor\phi_{k'}/Z\rfloor - \lfloor\phi/Z\rfloor}\right)\right\};$$
9:         Update $\mathcal{K}_j^{\mathrm{un}} := \mathcal{K}_j^{\mathrm{un}}\backslash\{k\}, \mathcal{K}_j^{\mathrm{al}} := \mathcal{K}_j^{\mathrm{al}}\cup\{k\}$;
10:     **end for**
11: **end for**

We evaluate the complexity of the proposed scheme as follows. The complexity of K-means clustering is $\mathcal{O}(eKJL)$, where e is the number of iterations needed until convergence and is often small [31]. γ is the overlap degree threshold of two matrices, which can balance the complexity and performance of the APSP set allocation scheme. To search for $|\mathcal{Y}|$ phase shifts for each UE, no more than $N_c|\mathcal{Y}|\frac{K'(K'-1)}{2}$ calculations of (26) are needed assuming the number of UEs in a cluster is approximately $K' = \lfloor\frac{K}{J}\rfloor$, since there is no need to calculate (26) when $\delta(\langle\phi_{k'}\rangle_Z - \langle\phi\rangle_Z) = 0$. For each calculation, considering

N_{cp} is relatively small, the maximal scalar multiplication number for calculating $\mathbf{\Upsilon}_k^0 \odot \mathbf{\Upsilon}_k^0$ is $\mathcal{O}(M)$, the complexity for obtaining $\mathbf{\Upsilon}_{k'}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi/Z \rfloor}$ based on (15) is neglected because it only needs cyclic column shift and truncation, and the scalar multiplication number for calculating $\xi(\mathbf{\Upsilon}_k^0 \odot \mathbf{\Upsilon}_k^0, \mathbf{\Upsilon}_{k'}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi/Z \rfloor})$ is $\mathcal{O}(M)$ when $\mathbf{\Upsilon}_k^0 \odot \mathbf{\Upsilon}_k^0$ and $\mathbf{\Upsilon}_{k'}^{\beta, \lfloor \phi_{k'}/Z \rfloor - \lfloor \phi/Z \rfloor}$ are figured out. Hence, the computational complexity of allocating APSP sets for all UEs is no more than $\mathcal{O}\left(JM^2 K'^2 N_c\right)$. When λ is small enough, it can be reduced to $\mathcal{O}\left(JN^2 K'^2 N_c\right)$. When γ is large enough, it can be further reduced to $\mathcal{O}\left(JN^2 K'^2\right)$. Compared with the scheme allocating APSP sets for all UEs directly, the complexity of the proposed scheme is greatly reduced since UEs are partitioned via K-means method with liner complexity, which makes the proposed scheme can be utilized in crowded scenarios.

4 Spectral efficiency and power control

In this section, we analyse a lower bound of uplink SE in cell-free massive MIMO-OFDM systems. Then we propose a power control scheme to maximize the minimum SE lower bound among active UEs.

4.1 A lower bound of SE

In the uplink data transmission phase, K_a active UEs simultaneously transmit data to APs. The data of UE k over the s-th subcarrier in an OFDM symbol is denoted by $x_{k,s} \sim \mathcal{CN}(0,1)$, which is weighted by a power control coefficient $\sqrt{\eta_k}$ satisfying $0 \leqslant \eta_k \leqslant 1$. The received signal at AP l over the s-th subcarrier is given by

$$\boldsymbol{y}_{u,l,s} = \sqrt{\rho_u} \sum_{k' \in \mathcal{K}_a} \sqrt{\eta_{k'}} \boldsymbol{g}_{k',l,s}^{\beta} x_{k',s} + \boldsymbol{w}_{l,s} \in \mathbb{C}^{N \times 1}, \tag{27}$$

where ρ_u is the normalized uplink SNR and $\boldsymbol{w}_{l,s}$ is the AWGN vector at the l-th AP over the s-th subcarrier including i.i.d. $\mathcal{CN}(0,1)$ elements. The l-th AP needs to detect the symbols transmitted from active UEs belonging to the set $\mathcal{D}_l$. The normalized MRC is utilized, i.e., $\boldsymbol{c}_{k,l,s}^{\beta} = \hat{\boldsymbol{g}}_{k,l,s}^{\beta} / \|\hat{\boldsymbol{g}}_{k,l,s}^{\beta}\|^2$, where $\hat{\boldsymbol{g}}_{k,l,s}^{\beta}$ is the estimated channel response vector between the k-th UE and the l-th AP over the s-th subcarrier. The symbol of UE k over the s-th subcarrier detected by the l-th AP is given by

$$
\begin{aligned}
r_{k,l,s} &= \left(\boldsymbol{c}_{k,l,s}^{\beta}\right)^{\text{H}} \boldsymbol{y}_{u,l,s} \\
&= \sqrt{\rho_u}\sqrt{\eta_k} \left(\boldsymbol{c}_{k,l,s}^{\beta}\right)^{\text{H}} \boldsymbol{g}_{k,l,s}^{\beta} x_{k,s} + \left(\boldsymbol{c}_{k,l,s}^{\beta}\right)^{\text{H}} \left(\sqrt{\rho_u} \sum_{k' \in \mathcal{K}_a \setminus \{k\}} \sqrt{\eta_{k'}} \boldsymbol{g}_{k',l,s}^{\beta} x_{k',s}\right) + \left(\boldsymbol{c}_{k,l,s}^{\beta}\right)^{\text{H}} \boldsymbol{w}_{l,s}.
\end{aligned} \tag{28}
$$

Next, each AP sends the detected symbols to the CPU via a fronthaul network. For UE k, the CPU only sees the detected symbols from APs serving this UE, which can be represented as

$$r_{k,s} = \sum_{l=0}^{L-1} \nu_{k,l} r_{k,l,s} = \sqrt{\rho_u} \sum_{k' \in \mathcal{K}_a} \sqrt{\eta_{k'}} \left(\boldsymbol{c}_{k,s}^{\beta,0}\right)^{\text{H}} \boldsymbol{g}_{k',s}^{\beta} x_{k',s} + \left(\boldsymbol{c}_{k,s}^{\beta,0}\right)^{\text{H}} \boldsymbol{w}_s, \tag{29}$$

where $\boldsymbol{c}_{k,s}^{\beta,0} \triangleq \text{vec}\{\nu_{k,0} \boldsymbol{c}_{k,0,s}^{\beta}, \nu_{k,1} \boldsymbol{c}_{k,1,s}^{\beta}, \ldots, \nu_{k,L-1} \boldsymbol{c}_{k,L-1,s}^{\beta}\}$, and $\boldsymbol{w}_s \triangleq \text{vec}\{\boldsymbol{w}_{0,s}, \boldsymbol{w}_{1,s}, \ldots, \boldsymbol{w}_{L-1,s}\}$.

Let $\varpi \triangleq \frac{N_c}{N_c + N_{\text{cp}}} \frac{Z_a - Z}{Z_a}$. Based on the use-and-then-forget method [32], a lower bound of SE of UE k over the s-th subcarrier is given in (30) with $\text{SINR}_{k,s}^{\text{lb}}$ given in (31) (see Appendix C for derivations).

$$\text{SE}_{k,s}^{\text{lb}}(\{\eta_k\}) = \varpi \log_2 \left\{1 + \text{SINR}_{k,s}^{\text{lb}}(\{\eta_k\})\right\}, \tag{30}$$

$$\text{SINR}_{k,s}^{\text{lb}}(\{\eta_k\}) = \frac{\eta_k |\text{E}\{(\boldsymbol{c}_{k,s}^{\beta,0})^{\text{H}} \boldsymbol{g}_{k,s}^{\beta}\}|^2}{\sum\limits_{k' \in \mathcal{K}_a} \eta_{k'} \text{E}\{|(\boldsymbol{c}_{k,s}^{\beta,0})^{\text{H}} \boldsymbol{g}_{k',s}^{\beta}|^2\} - \eta_k |\text{E}\{(\boldsymbol{c}_{k,s}^{\beta,0})^{\text{H}} \boldsymbol{g}_{k,s}^{\beta}\}|^2 + \frac{1}{\rho_u} \text{E}\{\|\boldsymbol{c}_{k,s}^{\beta,0}\|^2\}}. \tag{31}$$

In correlated channels with less hardening, the lower bound is tighter if the normalized MRC is used, compared with the lower bound in [5, 12] using MRC as $\boldsymbol{c}_{k,l,s}^{\beta} = \hat{\boldsymbol{g}}_{k,l,s}^{\beta}$. The reason and comparison are shown in [33]. In this work, since the correlated channels in cell-free massive MIMO systems are

considered, we utilize the tighter normalized MRC. Considering the closed-form expression of SE lower bound is tricky to be derived for normalized MRC in correlated channels, we adopt Monte Carlo simulation to obtain each expectation over the random channel realizations in (31). In Subsection 4.2, we optimize the power coefficients based on (30). The reason why we adopt a lower bound of SE to optimize power coefficients is that the expression of SE is not analytically tractable, though it does not rely on channel hardening and can reflect the real SE even in correlated channels if the transmitted signals are $\mathcal{CN}(0,1)$ RVs [33]. For simulation comparison in Section 5, the expression of SE is also given here [33]

$$
\mathrm{SE}_{k,s}(\{\eta_k\}) = \varpi \mathrm{E}\left\{\log_2\left\{1 + \frac{\rho_u\eta_k|(\boldsymbol{c}_{k,s}^{\beta,0})^{\mathrm{H}}\hat{\boldsymbol{g}}_{k,s}^{\beta}|^2}{(\boldsymbol{c}_{k,s}^{\beta,0})^{\mathrm{H}}(\rho_u\sum_{k'\in\mathcal{K}_a\setminus\{k\}}\eta_{k'}\hat{\boldsymbol{g}}_{k',s}^{\beta}(\hat{\boldsymbol{g}}_{k',s}^{\beta})^{\mathrm{H}} + \rho_u\sum_{k'\in\mathcal{K}_a}\eta_{k'}\boldsymbol{R}_{\tilde{\boldsymbol{g}}_{k',s}^{\beta}} + \mathbf{I}_M)\boldsymbol{c}_{k,s}^{\beta,0}}\right\}\right\},
$$
(32)

where the expectation is with respect to channel estimates, and $\boldsymbol{R}_{\tilde{\boldsymbol{g}}_{k',s}^{\beta}}$ is represented as

$$
\boldsymbol{R}_{\tilde{\boldsymbol{g}}_{k',s}^{\beta}} = \frac{1}{N_c}\boldsymbol{V}_{M\times M}\,\mathrm{diag}\left\{\mathrm{sum}\left\{\left[\boldsymbol{\Xi}_{k'}^{\beta}\right]_{0,:}\right\}\,\mathrm{sum}\left\{\left[\boldsymbol{\Xi}_{k'}^{\beta}\right]_{1,:}\right\}\cdots\mathrm{sum}\left\{\left[\boldsymbol{\Xi}_{k'}^{\beta}\right]_{M-1,:}\right\}\right\}\boldsymbol{V}_{M\times M}^{\mathrm{H}}.
$$
(33)

4.2 Power control

A wireless system should provide good service to all UEs. The minimum data rate among UEs is an important indicator of system performance. In this subsection, we aim to maximize the minimum uplink SE lower bound among active UEs using max-min power control, and hence, improve the real minimum SE among active UEs in correlated cell-free massive MIMO-OFDM systems. The problem of maximizing the minimum uplink SE lower bound among active UEs can be formulated as

$$
\begin{aligned}
\max_{\{\eta_k\}}\quad&\min_{k\in\mathcal{K}_a}\quad\mathrm{SE}_{k,s}^{\mathrm{lb}}(\{\eta_k\})\\
\mathrm{s.t.}\quad&0\leqslant\eta_k\leqslant1,\ k\in\mathcal{K}_a.
\end{aligned}
$$
(34)

Problem (34) can be equivalently reformulated as

$$
\begin{aligned}
\max_{\{\eta_k\},t}\quad&t\\
\mathrm{s.t.}\quad&t\leqslant\mathrm{SINR}_{k,s}^{\mathrm{lb}}(\{\eta_k\}),\ k\in\mathcal{K}_a,\\
&0\leqslant\eta_k\leqslant1,\ k\in\mathcal{K}_a.
\end{aligned}
$$
(35)

Problem (35) is quasilinear because inequalities are linear for a given t. It can be solved using bisection as shown in [5, 13], where initial values of $t_{\min}$ and $t_{\max}$ are chosen and the range is bisected until it can be accepted. However, it is not easy to choose the proper initial value $t_{\max}$ in practice. This is because a larger value can increase the iteration times and a smaller value cannot satisfy the requirement. Hence, we utilize Dinkelbach's algorithm to efficiently solve (35) as shown in Algorithm 2, where only the initial value of $t_{\min}$ is needed. We can obtain the globally optimal solution of this problem by iteratively solving linear programs in (36) and updating $t_{\min}$.

$$
\max_{\{\eta_k\}}\ w
$$

$$
\mathrm{s.t.}\quad\eta_k(1+t_{\min})\left|\mathrm{E}\left\{\left(\boldsymbol{c}_{k,s}^{\beta,0}\right)^{\mathrm{H}}\boldsymbol{g}_{k,s}^{\beta}\right\}\right|^2 - t_{\min}\left(\sum_{k'\in\mathcal{K}_a}\eta_{k'}\mathrm{E}\left\{\left|\left(\boldsymbol{c}_{k,s}^{\beta,0}\right)^{\mathrm{H}}\boldsymbol{g}_{k',s}^{\beta}\right|^2\right\} + \frac{1}{\rho_u}\mathrm{E}\left\{\left\|\boldsymbol{c}_{k,s}^{\beta,0}\right\|^2\right\}\right)\geqslant w,\ k\in\mathcal{K}_a,
$$
(36)

$$
0\leqslant\eta_k\leqslant1,\ k\in\mathcal{K}_a.
$$

Algorithm 2 Dinkelbach's algorithm for solving (34)

1: Initialize $t_{\min} = 0$ satisfying the constraint in (35), and choose a tolerance ϵ.
2: Solve the linear problem in (36) and obtain the optimal solution $\{\eta_k^*\}$ and the optimal value w^*, where the expectations over the random channel realizations can be computed separately by means of Monte Carlo simulation.
3: Update $t_{\min} := \min_{k\in\mathcal{K}_a}\{\mathrm{SINR}_{k,s}^{\mathrm{lb}}(\{\eta_k^*\})\}$.
4: Stop if $w^* \leqslant \epsilon$. Otherwise, go to step 2.

The Dinkelbach's algorithm always converges superlinearly and often (locally) quadratically [34]. The number of iterations is assumed to be T_D. For each iteration, when the interior-point method is adopted, the linear problem in step 2 can be solved with polynomial complexity $\mathcal{O}\left(K_a^{3.5} L_b^2\right)$ [35]. In this case, the number of bits in the input satisfies

$$
\begin{aligned}
L_b = & \sum_{k \in \mathcal{K}_a} \left\lceil 1 + \log_2 \left(1 + (1 + t_{\min}) \left| \mathrm{E} \left\{ \left(\boldsymbol{c}_{k,s}^{\beta,0} \right)^{\mathrm{H}} \boldsymbol{g}_{k,s}^{\beta} \right\} \right|^2 \right) \right\rceil \\
& + \sum_{k,k' \in \mathcal{K}_a} \left\lceil 1 + \log_2 \left(1 + t_{\min} \mathrm{E} \left\{ \left| \left(\boldsymbol{c}_{k,s}^{\beta,0} \right)^{\mathrm{H}} \boldsymbol{g}_{k',s}^{\beta} \right|^2 \right\} \right) \right\rceil \\
& + \sum_{k \in \mathcal{K}_a} \left\lceil 1 + \log_2 \left(1 + \frac{t_{\min}}{\rho_u} \mathrm{E} \left\{ \left\| \boldsymbol{c}_{k,s}^{\beta,0} \right\|^2 \right\} \right) \right\rceil + 6 K_a.
\end{aligned}
$$

Step 3 is performed with complexity $\mathcal{O}\left(K_a^2\right)$. Hence, the complexity of Algorithm 2 is $\mathcal{O}\left(K_a^{3.5} L_b^2 T_D\right)$, i.e., a polynomial of the number of active UEs, which is usually far less than that of UEs in crowded scenarios. Algorithm 2 should be recomputed at CPU if statistical CSI or activity pattern is changed.

In practice, it should be performed after UE detection in two ways depending on specific demands. If data needs to be decoded in a short time, active UEs should transmit pilots and data with full power in the first uplink frame shown in Figure 2. Once APs receive these pilots, active UEs can be detected based on hopping patterns and power coefficients can be optimized and sent to UEs during downlink transmission. Next, pilots can be transmitted with full power and data can be transmitted with power control. If the minimum SE among active UEs needs to be improved, the first ZR OFDM symbols can be used to transmit pilots for UE detection. Then power coefficients is optimized at CPU and transmitted to UEs. Finally, pilots are transmitted with full power and data is transmitted with power control. These pilots are for UE detection and channel estimation.

5 Numerical results and discussion

5.1 Large-scale fading model and system parameters

We assume that APs and UEs are independently and uniformly distributed within a square of size $1 \text{ km} \times 1 \text{ km}$. Each AP is equipped with two 100-antenna ULAs. The large-scale fading coefficient $\beta_{k,l}$ between the k-th UE and the l-th AP is modeled as $10 \log_{10}\left(\beta_{k,l}\right) = \mathrm{PL}_{k,l} \mathrm{SH}_{k,l}$, where the shadow fading $\mathrm{SH}_{k,l} \sim \mathcal{N}\left(0, \sigma_{\mathrm{sh}}^2\right)$ with $\sigma_{\mathrm{sh}} = 8$ dB. Let $d_{k,l}$ denote the distance between UE k and AP l. If $d_{k,l} \leqslant d_1$, there is no shadowing. We model the three-slope path loss $\mathrm{PL}_{k,l}$ as [36]

$$
\mathrm{PL}_{k,l} = \begin{cases} -\chi - 35 \log_{10}\left(d_{k,l} / 1 \, \mathrm{m}\right), & \text{if } d_{k,l} > d_1, \\ -\chi - 15 \log_{10}\left(d_1 / 1 \, \mathrm{m}\right) - 20 \log_{10}\left(d_{k,l} / 1 \, \mathrm{m}\right), & \text{if } d_0 < d_{k,l} \leqslant d_1, \\ -\chi - 15 \log_{10}\left(d_1 / 1 \, \mathrm{m}\right) - 20 \log_{10}\left(d_0 / 1 \, \mathrm{m}\right), & \text{if } d_{k,l} \leqslant d_0, \end{cases} \tag{37}
$$

where $\chi \triangleq 46.3 + 33.9 \log_{10}(f_c / 1 \, \mathrm{MHz}) - 13.82 \log_{10}\left(h_{\mathrm{AP}} / 1 \, \mathrm{m}\right) - \left(1.1 \log_{10}(f_c / 1 \, \mathrm{MHz}) - 0.7\right) h_{\mathrm{u}} + \left(1.56 \log_{10}(f_c / 1 \, \mathrm{MHz}) - 0.8\right)$, f_c is the carrier frequency (in MHz), and h_{AP} and h_{u} are the antenna height (in m) of AP and UE, respectively. The values of major parameters are given in Table 1.

The noise power is $\sigma_w^2 = B \times k_{\mathrm{B}} \times T_0 \times \mathrm{NF}$ (W), where $T_0 = 290$ (Kelvin) is the noise temperature, $k_{\mathrm{B}} = 1.381 \times 10^{-23}$ (Joule per Kelvin) is the Boltzmann constant, and $\mathrm{NF} = 9$ dB is the noise figure [13]. The corresponding normalized transmitting SNRs in the training phase and data transmission phase are assumed to be equal, which is computed by dividing P_x by σ_w^2.

We consider channels with 30 taps in the delay domain, unless stated otherwise. The channel power from a UE to an AP is normalized as $\sum_{i,j} \left[\boldsymbol{\Upsilon}_{k,l}\right]_{i,j} = N N_{\mathrm{cp}}$.

5.2 Results and discussion

5.2.1 *The MSE-CE performance with different numbers of active UEs and pilot symbols*

In Figure 3, we compare the MSE-CE performance of pilot assignment schemes and theoretical lower bound versus different numbers of active UEs. In the legend, we rewrite "lower bound" and "allocation"

Table 1 System parameters for the simulation

Parameter	Value	Unit
Bandwidth B	20	MHz
Carrier frequency f_c	2	GHz
Sampling duration T_s	48.8	ns
Subcarrier number N_c, guard interval N_{cp}	1024, 144	–
The number of UEs K, the number of clusters J	200, 2	–
The number of OFDM symbols in a slot Z_a	7	–
h_{AP}, h_{u}, d_1, d_0	15, 1.65, 50, 10	m
Maximum transmitting power P_x	0.5	W
Pilot phase shift number per UE $|\mathcal{Y}|$	4	–
γ	$10^{-8}(Z=1), 10^{-13}(Z=2)$	–
ι_k	0.4	–
$P_{\mathrm{ip},k}$, $P_{\mathrm{ip},l}$, $P_{0,l}$	0.1, 0.1, 0.825	W
$P_{\mathrm{bt},l}$	0.25	W/(Gbits/s)
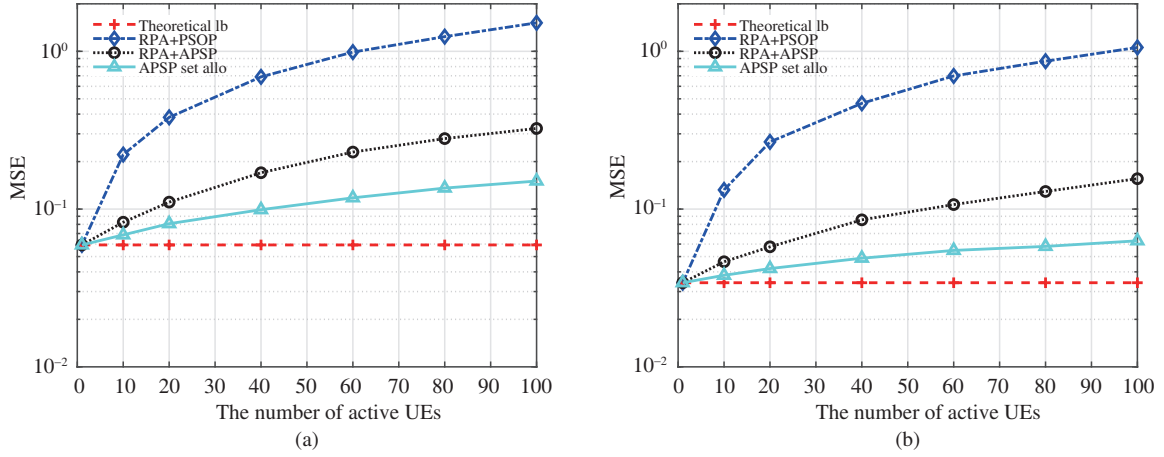
Figure 3 (Color online) Comparison of MSE-CE of pilot assignment schemes and theoretical lower bound with AP selection versus different numbers of active UEs assuming $\zeta = 0.2$ μs, $\varsigma = 2°$, $L = 10$, and $\lambda = 0.7$. (a) Results are shown with $Z = 1$; (b) results are shown with $Z = 2$.

to "lb" and "allo" for short, respectively. Figure 3(a) and (b) show the performance with $Z = 1$ and $Z = 2$ OFDM symbols used for pilot transmission in a slot, respectively. It is shown that the performance of each scheme is improved when more OFDM symbols are used for pilot transmission. For the PSOP-based RPA scheme, UEs are randomly assigned pilots from PSOP sets. The number of PSOPs is $\lfloor ZN_c/N_{cp} \rfloor$. It can cause serious pilot interference, since the number of PSOPs is far less than that of UEs. For the APSP-based RPA scheme, UEs are randomly assigned phase shifts from $\boldsymbol{\Psi}$. It can utilize channel sparsity and outperform the PSOP-based scheme. Although the proposed APSP set allocation scheme is suboptimal in general compared with exhaustive search, substantial performance gains in terms of MSE-CE can still be achieved for different numbers of active UEs and pilot symbols. This is because it can reduce the overlap of angle-delay domain channel power distributions between the desired UE and interference with corresponding cyclic shifts in the delay domain. As observed from the trend of each curve, the performance gain of the proposed scheme can still be obvious when the number of active UEs is more than 100. Moreover, the theoretical lower bound is plotted, where each UE is assumed to be assigned a dedicated orthogonal pilot. As the number of active UEs increases, the MSE-CE of each pilot assignment scheme and the gap between it and the corresponding lower bound increase.

5.2.2 The MSE-CE performance with different channel sparsity

In Figure 4, we show the MSE-CE performance of pilot assignment schemes and lower bound versus different channel sparsity. In Figure 4(a), performance comparisons are presented versus different angle spreads. Different delay spreads are considered in Figure 4(b).

The APSP-based RPA scheme outperforms the scheme based on PSOP. If the APSP set allocation
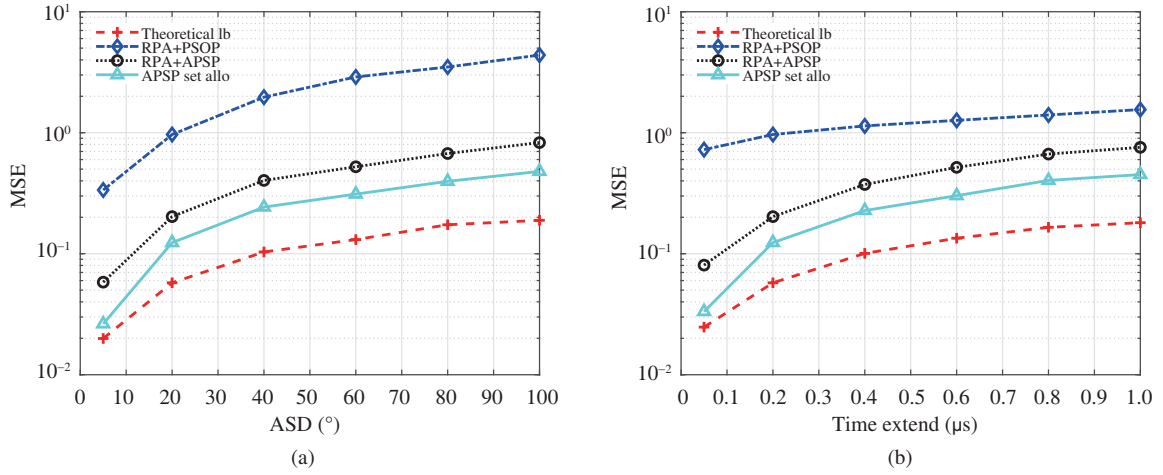
Figure 4 (Color online) Comparison of MSE-CE of pilot assignment schemes and theoretical lower bound with AP selection versus channel sparsity with $L = 10$, $K_a = 50$, $\lambda = 0.7$, and $Z = 1$. (a) Results are given versus angle spreads with $\zeta = 0.2$ μs; (b) results are given versus delay spreads with $\varsigma = 2°$.
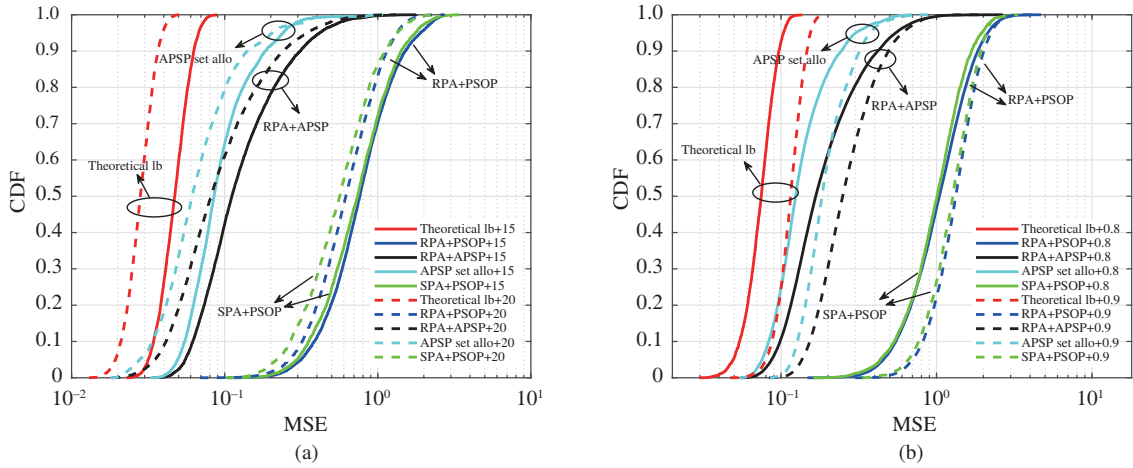
Figure 5 (Color online) CDFs of MSE-CE for pilot assignment schemes and theoretical lower bounds with AP selection assuming $K_a = 50$, $\zeta = 0.2$ μs, $\varsigma = 2°$, and $Z = 1$. (a) Results are shown with $\lambda = 0.7$ and $L = 15$ and 20, respectively; (b) results are shown with $L = 10$ and $\lambda = 0.8$ and 0.9, respectively.

scheme is adopted, the performance can be further improved and close to its theoretical lower bound especially when the angle spread and delay spread are small. Besides, the performance of each scheme improves as angle spread or delay spread decreases, i.e., as channels become more sparse.

5.2.3 *The MSE-CE performance with different numbers of APs and different AP selection coefficients*

Figure 5 compares the MSE-CE cumulative distributions (CDFs) for pilot assignment schemes and lower bounds. In Figure 5(a), MSE-CE CDFs are presented for different numbers of APs.

The performance of each scheme improves as the number of APs increases. This is because the macro-diversity provided by many APs reduces the risk that a UE has large distances to all APs. Figure 5(b) shows the MSE-CE CDFs of pilot assignment schemes with different AP selection coefficients. The performance of each scheme improves as λ decreases. This is because APs with better channel estimation performance for a UE are possible to have larger large-scale fading coefficients. When λ is increased, more APs far from a given UE are taken into account, and the MSE-CE averaged over serving APs of a UE can be worse. Besides, compared with the PSOP-based RPA scheme, the SPA scheme [18] has limited median gain in crowded scenarios. Owing to the reduced pilot interference effect, the performance of the proposed APSP set allocation scheme is close to the lower bound and significantly outperforms other schemes in both median and 95%-likely performance.
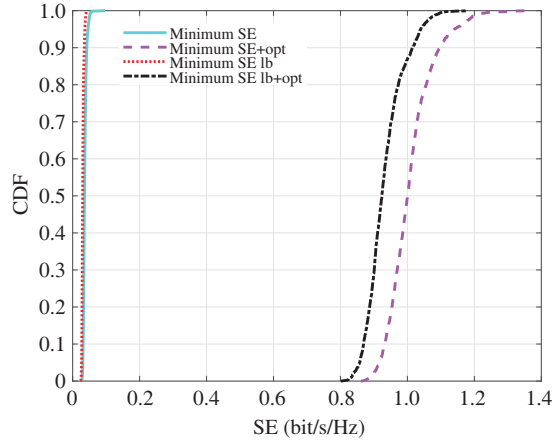
Figure 6 (Color online) CDFs of the minimum SE lower bound among active UEs and corresponding SE with and without power control under the proposed APSP set allocation scheme with $K_a = 50$, $L = 10$, $\lambda = 0.7$, $\zeta = 0.8$ μs, $\varsigma = 8^\circ$, $Z = 1$, and 50 taps.

5.2.4   *The SE performance*

In Figure 6, we present the CDFs of the minimum SE and minimum SE lower bound among active UEs with and without power control under the APSP set allocation scheme. Based on Algorithm 2 with $\epsilon = 0.02$, we maximize the minimum SE lower bound among K_a active UEs, which significantly outperforms the case without power control in both median and 95%-likely performance. Applying the optimized power coefficients of problem (35) to the real minimum SE among active UEs, it is shown that power control can significantly benefit the minimum SE.

6   Conclusion

In this paper, we considered uplink pilot and data transmission in crowded cell-free massive MIMO-OFDM systems with spatial and frequency correlation. For the pilot transmission, we utilized pilot phase shift hopping patterns to identify active UEs. Meanwhile, we derived a closed-form expression of MSE-CE with APSP, and provided an optimal condition of minimizing MSE-CE. According to this condition, we further developed an APSP set allocation scheme to reduce channel estimation error. This scheme is insensitive to the active pattern of UEs and does not need to be recomputed frequently. Besides, the expressions of SE and a lower bound of SE were derived. For the data transmission, we devised a max-min power control algorithm to maximize the minimum SE lower bound among active UEs. We can obtain the globally optimal solution to this problem by iteratively solving linear programs. Significant performance gains in terms of MSE-CE were observed for the proposed APSP set allocation scheme because it can fully utilize the channel sparsity. Compared with the equal power control, our proposed power allocation scheme can improve the minimum SE among active UEs in both median and 95%-likely performance. Hence, the proposed APSP set allocation scheme and the power control scheme are crucial and suitable for crowded cell-free massive MIMO-OFDM systems with frequently and spatially correlated channels.

References

1 Larsson E G, Edfors O, Tufvesson F, et al. Massive MIMO for next generation wireless systems. IEEE Commun Mag, 2014, 52: 186–195

2 Lu L, Li G Y, Swindlehurst A L, et al. An overview of massive MIMO: benefits and challenges. IEEE J Sel Top Signal Process, 2014, 8: 742–758

3 Ngo H Q, Ashikhmin A, Yang H, et al. Cell-free massive MIMO versus small cells. IEEE Trans Wireless Commun, 2017, 16: 1834–1850

4 Zhang J Y, Bjornson E, Matthaiou M, et al. Prospective multiple antenna technologies for beyond 5G. IEEE J Sel Areas Commun, 2020, 38: 1637–1660

5 Zhang J, Chen S, Lin Y, et al. Cell-free massive MIMO: a new next-generation paradigm. IEEE Access, 2019, 7: 99878–99888

6 Cimini L J. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. IEEE Trans Commun, 1985, 33: 665–675

7 Zhao J, Ni S, Yang L, et al. Multiband cooperation for 5G HetNets: a promising network paradigm. IEEE Veh Technol Mag, 2019, 14: 85–93

8 Bockelmann C, Pratas N, Nikopour H, et al. Massive machine-type communications in 5G: physical and MAC-layer solutions. IEEE Commun Mag, 2016, 54: 59–65

9 de Carvalho E, Bjornson E, Sorensen J H, et al. Random access protocols for massive MIMO. IEEE Commun Mag, 2017, 55: 216–222

10 Wu Y P, Gao X Q, Zhou S D, et al. Massive access for future wireless communication systems. IEEE Wireless Commun, 2020, 27: 148–156

11 de Carvalho E, Bjornson E, Sorensen J H, et al. Random pilot and data access in massive MIMO for machine-type communications. IEEE Trans Wireless Commun, 2017, 16: 7703–7717

12 Chen Z, Bjornson E. Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry. IEEE Trans Commun, 2018, 66: 5205–5219

13 Nayebi E, Ashikhmin A, Marzetta T L, et al. Precoding and power optimization in cell-free massive MIMO systems. IEEE Trans Wireless Commun, 2017, 16: 4445–4459

14 Yang H, Marzetta T L. Energy efficiency of massive MIMO: cell-free vs. cellular. In: Proceedings of IEEE 87th Vehicular Technology Conference, 2018. 1–5

15 Bjornson E, Sanguinetti L. Making cell-free massive MIMO competitive with MMSE processing and centralized implementation. IEEE Trans Wireless Commun, 2020, 19: 77–90

16 Ngo H Q, Tran L N, Duong T Q, et al. On the total energy efficiency of cell-free massive MIMO. IEEE Trans Green Commun Netw, 2018, 2: 25–39

17 Interdonato G, Bjornson E, Ngo H Q, et al. Ubiquitous cell-free massive MIMO communications. J Wireless Com Network, 2019, 1: 197

18 Attarifar M, Abbasfar A, Lozano A. Random vs structured pilot assignment in cell-free massive MIMO wireless networks. In: Proceedings of IEEE International Conference on Communications Workshops, Kansas City, 2018. 1–6

19 Liu H, Zhang J, Zhang X, et al. Tabu-search-based pilot assignment for cell-free massive MIMO systems. IEEE Trans Veh Technol, 2020, 69: 2286–2290

20 Liu H, Zhang J, Jin S, et al. Graph coloring based pilot assignment for cell-free massive MIMO systems. IEEE Trans Veh Technol, 2020, 69: 9180–9184

21 Liu K, Raghavan V, Sayeed A M. Capacity scaling and spectral efficiency in wide-band correlated mimo channels. IEEE Trans Inform Theor, 2003, 49: 2504–2526

22 Fleury B H. First- and second-order characterization of direction dispersion and space selectivity in the radio channel. IEEE Trans Inform Theor, 2000, 46: 2027–2044

23 You L, Gao X, Swindlehurst A L, et al. Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots. IEEE Trans Signal Process, 2016, 64: 1461–1476

24 Adhikary A, Nam J, Ahn J-Y, et al. Joint spatial division and multiplexing-the large-scale array regime. IEEE Trans Inform Theor, 2013, 59: 6441–6463

25 Pedersen K I, Mogensen P E, Fleury B H. A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments. IEEE Trans Veh Technol, 2000, 49: 437–447

26 Tse D, Viswanath P. Fundamentals of Wireless Communication. Cambridge: Cambridge University Press, 2005

27 You L, Gao X Q, Xia X G, et al. Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels. IEEE Trans Wireless Commun, 2015, 14: 3352–3366

28 Wen C K, Jin S, Wong K K, et al. Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning. IEEE Trans Wireless Commun, 2015, 14: 1356–1368

29 Li Y G. Simplified channel estimation for OFDM systems with multiple transmit antennas. IEEE Trans Wireless Commun, 2002, 1: 67–75

30 Bjornson E, Jalden N, Bengtsson M, et al. Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission. IEEE Trans Signal Process, 2011, 59: 6086–6101

31 Hartigan J A, Wong M A. Algorithm AS 136: a k-means clustering algorithm. J R Stat Soc C-Appl, 1979, 28: 100–108

32 Marzetta T L, Larsson E G, Yang H, et al. Fundamentals of Massive MIMO. Cambridge: Cambridge University Press, 2016

33 Bjornson E, Hoydis J, Sanguinetti L. Massive MIMO networks: spectral, energy, and hardware efficiency. FNT Signal Process, 2017, 11: 154–655

34 Schaible S. Fractional programming. II, on Dinkelbach's algorithm. Manage Sci, 1976, 22: 868–873

35 Karmarkar N K. A new polynomial-time algorithm for linear programming. Combinatorica, 1984, 4: 373–395

36 Tang A, Sun J, Gong K. Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area. In: Proceedings of IEEE VTS 53rd Vehicular Technology Conference, 2001