Overloaded Pilot Assignment with Pilot Decontamination for Cell-Free Systems

Noboru Osawa*, Fabian Göttsch†, Issei Kanno*, Takeo Ohseki*,
Yoshiaki Amano*, Kosuke Yamazaki*, Giuseppe Caire†
*KDDI Research Inc., Saitama, Japan
†Technical University of Berlin, Germany
Emails: {nb-oosawa, is-kanno, ohseki, yo-amano, ko-yamazaki}@kddi.com, {fabian.goettsch, caire}@tu-berlin.de

Abstract—The pilot contamination in cell-free massive multiple-input-multiple-output (CF-mMIMO) must be addressed for accommodating a large number of users. In previous works, we have investigated a decontamination method called subspace projection (SP). The SP separates interference from co-pilot users by using the orthogonality of the principal components of the users' channel subspaces. Non-overloaded pilot assignment (PA), where each radio unit (RU) does not assign the same pilot to different users, limits the spectral efficiency (SE) of the system, since SP channel estimation is able to deal with co-pilot users that have nearly orthogonal subspaces. Motivated by this limitation, this paper introduces overloaded PA methods adjusted for the decontamination in order to improve the sum SE of CF systems. Numerical simulations show that the overloaded PA methods give higher SE than that of non-overloaded PA at a high user density scenario.

Index Terms—Cell-free massive MIMO, user-centric, pilot contamination, pilot assignment.

I. INTRODUCTION

A large number of works in wireless communication theory is dedicated to the joint processing of spatially distributed antennas. This idea can be traced back to the work of Wyner [1], and has been "re-marketed" several times under different names with slight nuances, such as coordinate multipoint (CoMP), cloud radio access network (CRAN), and currently, it is promoted as cell-free massive MIMO (CF-mMIMO). As the name implies, one of the purposes of CF-mMIMO is eliminating cell boundaries and preventing user equipment's (UE's) performance degradation depending on their geographical positions.

The channel state information (CSI) acquisition with pilot signals at the infrastructure antenna side is important for taking advantage of CF-mMIMO's spatial multiplexing gain. In this paper, we focus on uplink (UL) channel estimation in a time division duplex (TDD) system. Thanks to TDD operations and channel reciprocity, these estimates can also be used for the downlink precoding [2]. Generally, the number of orthogonal pilots is limited to reduce overhead and keep the training phase within coherence time. Therefore, pilot reuse is inevitable to increase the number of UEs to be accommodated, and it induces pilot contamination. In addition, CF-mMIMO system has no clear boundaries, so it is hard to apply a cell-based pilot reuse restriction.

As one of the pilot decontamination methods, in [3], [4], we have investigated a subspace projection (SP) based channel estimation, which uses receiver side discrete Fourier transform

(DFT) processing. The channel subspace is spanned by dominant components of channel covariance matrices. Assuming uniform linear arrays (ULAs) or uniform planar arrays (UPAs), the channel covariance matrix is Toeplitz (for ULA) or Block-Toeplitz (for UPA), which both are approximately diagonalized by discrete Fourier transforms (DFT) on the columns and rows [5]. This means that the dominant channel subspaces are approximately spanned by subsets of the DFT columns. If the antennas have correlation due to a limited scattering angular spread, each radio unit (RU) and UE link has an individual channel subspace. Therefore, DFT-based projection onto the target UE's subspace can reduce interference from co-pilot UEs if the subspaces of the co-pilot UEs are nearly orthogonal, which we call decontamination. We have shown that the effect of the pilot contamination can be reduced significantly with the SP-based decontamination, yielding a system performance close to the case of ideal CSI, where each RU has perfect knowledge of the channel vectors of its associated UEs [3].

The system model of CF in our previous works [3], [6] is similar to Björnson and Sanguinetti's scalable CF system [7], where each user-centric cluster is formed by a finite number of RUs. In [3], [6], we have not allowed the RUs to assign the same pilot to different users, which we call non-overloaded pilot assignment (PA). However, the SP enhances RUs to estimate channels of co-pilot UEs under overloaded PA, where RUs can assign the same pilot to multiple UEs. Adopting the overloaded PA leads to an increased number of associated RUs per UE, which can increase the spatial multiplexing gain per UE and improve spectral efficiency (SE).

Most of related works, which aimed to address contamination by strategic PA methods, implicitly adopted overloaded PA because pilot selection is significantly constrained by the cluster formation, when the non-overloaded PA is adopted. In [8], a greedy PA algorithm was introduced as an early study in the CF literature. In [9], a PA scheme aiming at user throughput maximization is investigated, where the maximization problem is solved by using an iterative scheme based on the Hungarian algorithm. In [10], user-group PA was proposed which assigns the same pilot to UEs who share the fewest number of serving RUs. Furthermore, graphic framework based pilot assignment schemes are investigated in [11], [12]. Since these works do not adopt the SP-based decontamination, combining these overloaded PA methods and our decontamination method has

potential to achieve better performance.

As mentioned above, to the best of our knowledge, combination of the SP decontamination and overloaded PA has not yet been investigated. Motivated by this, we propose two types of overloaded PA methods in this paper. We call these methods rough overloaded PA (R-OPA) and subspace information aided overloaded PA (SIA-OPA). The R-OPA forms clusters in advance without considering the contamination, and then assigns the pilots. We propose a specific implementation of R-OPA by adjusting a graphic framework based PA studied by Zeng et al. [12] to our SP based decontamination. The SIA-OPA has a restriction in the cluster formation phase that it adds an RU to a cluster of a UE only when the subspaces of co-pilot UEs are orthogonal.

The contributions of this work are summarized as follows.

- We propose R-OPA and SIA-OPA for CF systems. In addition, a graphic framework-based PA is designed as a specific implementation of R-OPA.
- We discuss the complexity of the proposed PA. Note that R-OPA with the graphic framework-based PA algorithm requires network-wide information exchange, whereas SIA-OPA does not.
- We compare the sum SE and outage performance of the non-overloaded PA, R-OPA and SIA-OPA to reveal a region of parameters where the performance is best.

II. SYSTEM MODEL

We consider a CF-mMIMO system with L RUs, each with M antennas, and K UEs. Both RUs and UEs are distributed on a squared region on the 2-dimensional plane. The set $\mathcal{C}_k \subseteq [L] = \{1, 2, \ldots, L\}$ denotes the cluster of RUs that serve UE k, and $\mathcal{U}_\ell \subseteq [K]$ denotes the set of UEs served by RU ℓ, where the set of integers from 1 to n is denoted by $[n]$. The RU-UE associations are described by a bipartite graph $\mathcal{G}$ whose vertices are the RUs and UEs, respectively. The set of edges accounting for associated RU-UE pairs is denoted by $\mathcal{E}$, i.e., $\mathcal{G} = \mathcal{G}([L], [K], \mathcal{E})$. We also define $\mathcal{U}(\mathcal{C}_k) = \bigcup_{\ell \in \mathcal{C}_k} \mathcal{U}_\ell$ as the set of UEs served by at least one RU in $\mathcal{C}_k$.

In UL transmissions, the UEs transmit with the same energy per symbol E_{s}, and we define the system parameter
$$\mathrm{SNR} = \frac{E_{\mathrm{s}}}{N_0}, \tag{1}$$
where N_0 denotes noise power spectral density. The large-scale-fading-coefficient (LSFC) between RU ℓ and UE k is denoted by $\beta_{\ell,k}$ and includes pathloss, blocking effects and shadowing, respectively. By assuming isotropic antennas, the maximum beamforming gain averaged over the small scale fading is M, therefore the maximum SNR at the receiver of RU ℓ from UE k is $\beta_{\ell,k} M \mathrm{SNR}$.

We consider orthogonal frequency-division multiplexing (OFDM) modulation and channels following the standard block-fading model [2], [13], [14], such that they are random but constant over coherence blocks of T signal dimensions in the time-frequency domain. The described methods are formulated for one resource block (RB), so the RB index is omitted for simplicity. We define channel matrices and its elements as

follows. $\mathbb{H} \in \mathbb{C}^{LM \times K}$ denotes the overall channel matrix between all LM RU antennas and all K UE antennas on a given RB. Next, $\mathbb{h}_k \in \mathbb{C}^{LM \times 1}$ denotes k-th column of $\mathbb{H} \in \mathbb{C}^{LM \times K}$. In addition, $\mathbf{h}_{\ell,k} \in \mathbb{C}^{M \times 1}$ denotes the channel vector between RU ℓ and UE k. Finally, $\mathbb{H}(\mathcal{C}_k) \in \mathbb{C}^{LM \times K}$ denotes the partial cluster-centric matrix for a cluster $\mathcal{C}_k$, whose $M \times 1$ blocks of RU-UE pairs $(\ell, k) \in \mathcal{E}$ are equal to $\mathbf{h}_{\ell,k}$, and equal to $\mathbf{0}$ (the identically zero vector) otherwise.

For ease of analysis, we assume that the individual channels between RUs and UEs follow the single ring local scattering model [5] where the covariance matrix is perfectly diagonalized by DFT. Then $\mathbf{h}_{\ell,k}$ is given by
$$\mathbf{h}_{\ell,k} = \sqrt{\frac{\beta_{\ell,k} M}{|\mathcal{S}_{\ell,k}|}} \mathbf{F}_{\ell,k} \boldsymbol{\nu}_{\ell,k}, \tag{2}$$
where $\mathcal{S}_{\ell,k} \subseteq \{0, \ldots, M-1\}$ and $\boldsymbol{\nu}_{\ell,k}$ are the angular support set according to [5], and an $|\mathcal{S}_{\ell,k}| \times 1$ i.i.d. Gaussian vector with components $\sim \mathcal{CN}(0,1)$, respectively. The set $\mathcal{S}_{\ell,k}$ is constructed by the angular support $\Theta_{\ell,k} = [\theta_{\ell,k} - \Delta/2, \theta_{\ell,k} + \Delta/2]$ centered at angle $\theta_{\ell,k}$ of the LOS between RU ℓ and UE k, with angular spread Δ. We let $\mathbf{F}$ denote the $M \times M$ unitary DFT matrix with (m,n)-elements $f_{m,n} = \frac{e^{-j\frac{2\pi}{M}mn}}{\sqrt{M}}$, and using a Matlab-like notation, $\mathbf{F}_{\ell,k} \triangleq \mathbf{F}(:, \mathcal{S}_{\ell,k})$ denotes the tall unitary matrix obtained by selecting the columns of $\mathbf{F}$ corresponding to the index set $\mathcal{S}_{\ell,k}$.

A. UL data transmission

Let us note that the UEs transmit with same power in the UL. In general cellular network systems, UL transmit power is controlled per UE mainly for 1) countering the near-far problem for the cell edge UEs, or 2) saving power for UEs close to an RU. We believe that the UL power control for the "near-far problem" hardly makes sense for CF-mMIMO systems, since the CF system itself tries to solve the near-far problem by making UEs connect to multiple RUs. UL power control for energy saving or rate maximization still can be considered, however, it is very challenging to implement a centralized process for optimization in a distributed system or to find a decentralized way of UL power control, which will be future work.

According to the results in [3], we choose the local LMMSE with cluster-level combining for this work, since it showed high performance with scalability, compared to other methods such as cluster-level zero-forcing and local MRC.

The received $LM \times 1$ symbol vector at the LM RU antennas in UL is given by
$$\mathbb{y}^{\mathrm{ul}} = \sqrt{\mathsf{SNR}}\, \mathbb{H} \mathbb{s}^{\mathrm{ul}} + \mathbb{z}^{\mathrm{ul}}, \tag{3}$$
where $\mathbb{s}^{\mathrm{ul}} \in \mathbb{C}^{K \times 1}$ is the vector of information symbols transmitted by the UEs and $\mathbb{z}^{\mathrm{ul}}$ is an i.i.d. noise vector with components $\sim \mathcal{CN}(0,1)$. We define the received symbols vector at RU ℓ as $\mathbb{y}_\ell^{\mathrm{ul}} \in \mathbb{C}^{M \times 1}$, and the receiver unit norm vector for UE k as $\mathbb{v}_k \in \mathbb{C}^{LM \times 1}$ formed by $M \times 1$ blocks $\mathbf{v}_{\ell,k} : \ell = 1, \ldots, L$, such that $\mathbf{v}_{\ell,k} = \mathbf{0}$ if $(\ell, k) \notin \mathcal{E}$. Each RU

$\ell \in \mathcal{C}_k$ computes the local LMMSE receiving vector $\mathbf{v}_{\ell,k}$ for the UEs $k \in \mathcal{U}_\ell$; given by

$$\mathbf{v}_{\ell,k} = \left(\sigma_\ell^2 \mathbf{I} + \mathsf{SNR} \sum_{j \in \mathcal{U}_\ell} \mathbf{h}_{\ell,j} \mathbf{h}_{\ell,j}^{\mathsf{H}} \right)^{-1} \mathbf{h}_{\ell,k}, \qquad (4)$$

where σ_ℓ^2 denotes the approximated variance of noise and external interference [6], given by

$$\sigma_\ell^2 = 1 + \mathsf{SNR} \sum_{j \neq k} \beta_{\ell,j}. \qquad (5)$$

RU ℓ computes the local observation $r_{\ell,k}^{\mathrm{ul}} = \mathbf{v}_{\ell,k}^{\mathsf{H}} \mathbf{y}_\ell^{\mathrm{ul}}$ for each $k \in \mathcal{U}_\ell$ and sends the observations to a centralized unit for the cluster-level combining. Then the centralized unit computes the cluster-level combined observation as

$$r_k^{\mathrm{ul}} = \sum_{\ell \in \mathcal{C}_k} w_{\ell,k}^* r_{\ell,k}^{\mathrm{ul}}. \qquad (6)$$

The optimized combining weights vector $\mathbf{w}_k = \{w_{\ell,k} : \ell \in \mathcal{C}_k\}$, which maximizes the SINR [3], is given by

$$\mathbf{w}_k = \mathbf{\Gamma}_k^{-1} \mathbf{a}_k \in \mathbb{C}^{|\mathcal{C}_k| \times 1}, \qquad (7)$$

$$\mathbf{a}_k = \{g_{\ell,k,k} : \ell \in \mathcal{C}_k\}, \qquad (8)$$

$$g_{\ell,k,j} = \mathbf{v}_{\ell,k}^{\mathsf{H}} \mathbf{h}_{\ell,j}, \qquad (9)$$

$$\mathbf{\Gamma}_k = \mathbf{D}_k + \mathsf{SNR}\, \mathbf{G}_k \mathbf{G}_k^{\mathsf{H}} \in \mathbb{C}^{|\mathcal{C}_k| \times |\mathcal{C}_k|}, \qquad (10)$$

$$\mathbf{D}_k = \mathrm{diag}\left\{ \sigma_\ell^2 \|\mathbf{v}_{\ell,k}\|^2 : \ell \in \mathcal{C}_k \right\}, \qquad (11)$$

where the matrix $\mathbf{G}_k$ of dimension $|\mathcal{C}_k| \times (|\mathcal{U}(\mathcal{C}_k)| - 1)$ contains elements $g_{\ell,k,j}$ in position corresponding to RU ℓ and UE j (after a suitable index reordering) if $(\ell, j) \in \mathcal{E}$, and zero elsewhere. The overall receiving vector $\mathbb{v}_k$ is formed by stacking the vectors $w_{\ell,k} \mathbf{v}_{\ell,k}$, i.e., $\mathbb{v}_k = [w_{1,k} \mathbf{v}_{1,k}^{\mathsf{T}}, w_{2,k} \mathbf{v}_{2,k}^{\mathsf{T}}, ..., w_{L,k} \mathbf{v}_{L,k}^{\mathsf{T}}]^{\mathsf{T}}$.

The resulting SINR for UE k's UL symbol is given by

$$\mathsf{SINR}_k^{\mathrm{ul}} = \frac{|\mathbb{v}_k^{\mathsf{H}} \mathbb{h}_k|^2}{\mathsf{SNR}^{-1} + \sum_{j \neq k} |\mathbb{v}_k^{\mathsf{H}} \mathbb{h}_j|^2}. \qquad (12)$$

We use the optimistic ergodic achievable rate R_k^{ul} for performance evaluation, which is given by

$$R_k^{\mathrm{ul}} = \mathbb{E}\left[\log(1 + \mathsf{SINR}_k^{\mathrm{ul}}) \right], \qquad (13)$$

where the expectation is with respect to the small scale fading. Then, the UL spectral efficiency (SE) is calculated as

$$\mathsf{SE}_k^{\mathrm{ul}} = (1 - \tau_p/T) R_k^{\mathrm{ul}}, \qquad (14)$$

where T is the dimension of an RB and τ_p is the pilot dimension.

B. UL channel estimation

As a practical remark, we note that in 5GNR two types of UL pilots are specified, the demodulation reference signals (DMRS) and the sounding reference signals (SRS). In this work we assume that the instantaneous channel coefficients are estimated from orthogonal DMRS pilot sequences, and the subspace information is estimated by utilizing SRS pilots. [1]

The DMRS pilot field received at RU ℓ is given by the $M \times \tau_p$ matrix as

$$\mathbf{Y}_\ell^{\mathrm{DMRS}} = \sum_{i=1}^{K} \mathbf{h}_{\ell,i} \boldsymbol{\phi}_{t_i}^{\mathsf{H}} + \mathbf{Z}_\ell^{\mathrm{DMRS}}, \qquad (15)$$

[1]An estimation method of subspace information by utilizing SRS pilots is discussed in [6].

where $\boldsymbol{\phi}_{t_i}$ denotes the DMRS pilot vector with index t_i of dimension τ_p used by UE i in the current RB, with total energy $\|\boldsymbol{\phi}_{t_i}\|^2 = \tau_p \mathsf{SNR}$. For each UE $k \in \mathcal{U}_\ell$, RU ℓ produces the *pilot matching* (PM) channel estimate

$$\begin{aligned} \widehat{\mathbf{h}}_{\ell,k}^{\mathrm{pm}} &= \frac{1}{\tau_p \mathsf{SNR}} \mathbf{Y}_\ell^{\mathrm{DMRS}} \boldsymbol{\phi}_{t_k} \\ &= \mathbf{h}_{\ell,k} + \sum_{\substack{i:t_i=t_k \\ i \neq k}} \mathbf{h}_{\ell,i} + \widetilde{\mathbf{z}}_{t_k,\ell}, \end{aligned} \qquad (16)$$

where $\widetilde{\mathbf{z}}_{t_k,\ell}$ has i.i.d. with components $\mathcal{CN}(0, \frac{1}{\tau_p \mathsf{SNR}})$. Notice that the presence of UEs $i \neq k$ using the same DMRS pilot t_k yields pilot contamination.

We consider the SP based decontamination scheme for which the projected channel estimate is given by the orthogonal projection of $\widehat{\mathbf{h}}_{\ell,k}^{\mathrm{pm}}$ onto the subspace spanned by the columns of $\mathbf{F}_{\ell,k}$, i.e.,

$$\begin{aligned} \widehat{\mathbf{h}}_{\ell,k}^{\mathrm{sp}} &= \mathbf{F}_{\ell,k} \mathbf{F}_{\ell,k}^{\mathsf{H}} \widehat{\mathbf{h}}_{\ell,k}^{\mathrm{pm}} \\ &= \mathbf{h}_{\ell,k} + \sum_{\substack{i:t_i=t_k \\ i \neq k}} \mathbf{F}_{\ell,k} \mathbf{F}_{\ell,k}^{\mathsf{H}} \mathbf{h}_{\ell,i} + \mathbf{F}_{\ell,k} \mathbf{F}_{\ell,k}^{\mathsf{H}} \widetilde{\mathbf{z}}_{t_k,\ell}. \end{aligned} \qquad (17)$$

The second term of the last equation corresponds to pilot contamination after the SP, which is a Gaussian vector with zero mean and its covariance matrix can be written as

$$\boldsymbol{\Sigma}_{\ell,k}^{\mathrm{co}} = \sum_{\substack{i:t_i=t_k \\ i \neq k}} \frac{\beta_{\ell,i} M}{|\mathcal{S}_{\ell,i}|} \mathbf{F}_{\ell,k} \mathbf{F}_{\ell,k}^{\mathsf{H}} \mathbf{F}_{\ell,i} \mathbf{F}_{\ell,i}^{\mathsf{H}} \mathbf{F}_{\ell,k} \mathbf{F}_{\ell,k}^{\mathsf{H}}. \qquad (18)$$

When $\mathbf{F}_{\ell,k}$ and $\mathbf{F}_{\ell,i}$ are nearly mutually orthogonal, i.e. $\mathbf{F}_{\ell,k}^{\mathsf{H}} \mathbf{F}_{\ell,i} \approx \mathbf{0}$, the subspace projection is able to reduce the pilot contamination effect.

III. PILOT ASSIGNMENT AND CLUSTER FORMATION

As mentioned in Sec. 1, the SP can enhance overloaded PA, and lead to improved SE compared to non-overloaded PA. This section introduces the non-overloaded PA, R-OPA and SIA-OPA schemes with their cluster formation rules.

A. Non-overloaded pilot assignment

1) Leader RU selection: When a UE k wishes to join the system, it selects its leading RU $\ell(k)$ as the one with the largest LSFC and free DMRS pilots. This is formulated as $\ell(k) = \arg\max_{\ell \in \mathcal{L}_{\mathrm{f}}} \beta_{\ell,k}$, where $\mathcal{L}_{\mathrm{f}}$ denotes the set of RUs with free DMRS pilots. The leader RU selection also needs to satisfy the condition

$$\beta_{\ell,k} \geq \frac{\eta}{M \mathsf{SNR}}, \qquad (19)$$

where $\eta > 0$ is a suitable threshold determining how much above the noise floor the useful signal in the presence of maximum possible beamforming gain (equal to M) should be. If such RU is not available, then the UE is declared in outage.

2) Pilot selection: The RU chooses the pilot with least interference for UE k, i.e., UE k's pilot t_k is given by

$$t_k = \arg\min_{i \in \mathcal{T}_\ell} \sum_{j \in \mathcal{P}_i} \beta_{\ell,j}, \qquad (20)$$

where $\mathcal{T}_\ell$ and $\mathcal{P}_i$ denote the set of not assigned pilots at RU ℓ and the set of UEs with pilot index i, respectively. The set $\mathcal{P}_i$ is updated as $\mathcal{P}_i = \mathcal{P}_i \cup k$ after assignment of t_k. The RU only knows the result of $\sum_{j \in \mathcal{P}_i} \beta_{\ell,j}$ as measured statistics.

3) Cluster formation: Suppose that UE k finds its leader RU $\ell(k)$ and it is allocated the DMRS pilot with index $t_k \in [\tau_p]$. Then, the cluster $\mathcal{C}_k$ is obtained sorting the RUs satisfying condition (19) and having pilot t_k still available in decreasing LSFC order, and adding them to the cluster until a maximum cluster size Q is reached, where Q is a design parameter imposed to limit the computational complexity. As a result, for all UEs k not in outage, $1 \leq |\mathcal{C}_k| \leq Q$ and for all the RUs $\ell \in \mathcal{C}_k$, the corresponding LSFC satisfies (19). Furthermore, for all RUs ℓ we have $|\mathcal{U}_\ell| \leq \tau_p$.

B. Subspace information aided overloaded PA

We introduce a SIA-OPA approach where each RU assigns the same pilot only to UEs with orthogonal subspaces. Let us assume that the RUs acquire subspace information of UEs within its coverage before the leader RU selection phase.

1) Leader RU selection: In the leader RU selection phase, UE k firstly try to choose the RU with the largest LSFC as $\ell(k)$. If the RU $\ell(k)$ has free pilots, same as non-overloaded PA, the RU selects a pilot based on eq. (20). On the other hand, when the RU has no free pilot, the RU tries to find a pilot, which is assigned only to UEs with an orthogonal subspace with respect to UE k's subspace.

2) Pilot selection: A contamination quantity per pilot index i on the subspace of the channel between RU $\ell(k)$ and UE k is formulated by

$$\zeta_{\ell,k,i} \triangleq \sum_{j \in (\mathcal{U}_\ell \cap \mathcal{P}_i \setminus k)} ||\mathbf{F}_{\ell,k}\mathbf{F}_{\ell,k}^{\mathsf{H}}\mathbf{F}_{\ell,j}||_F \beta_{\ell,j}. \quad (21)$$

At first, if $\zeta_{\ell,k,i} = 0$ due to that pilot index i is not assigned to any other UEs in the coverage of RU $\ell(k)$, the pilot can be assigned to UE k as with the non-overloaded PA. In addition, even if there are other UEs with pilot index i, the metric $\zeta_{\ell,k,i}$ also becomes 0 if the subspaces are orthogonal as $\mathbf{F}_{\ell,k}^{\mathsf{H}}\mathbf{F}_{\ell,j}$ results in a zero matrix. Therefore, with SIA-OPA, the pilot selection is given by

$$t_k = \arg\min_{i|\zeta_{\ell,k,i}=0} \sum_{j \in \mathcal{P}_i} \beta_{\ell,j}, \quad (22)$$

where t_k is the least interfered pilot. If the RU ℓ cannot satisfy the above conditions, the UE approaches the next candidate RU, or UEs are declared in outage if no RU is available.

3) Cluster formation: In the cluster formation of UE k, the cluster $\mathcal{C}_k$ is obtained sorting the RUs satisfying condition (19) and $\zeta_{\ell,k,t_k} = 0$ in decreasing LSFC order, and adding them to the cluster until a maximum cluster size Q is reached.

C. Rough overloaded pilot assignment

1) Cluster formation: Finally, we introduce R-OPA, where user-centric clusters are formed before pilot allocation. In this method, each UE k chooses up to Q RUs with condition $\beta_{\ell,k} > \frac{\eta}{M\mathsf{SNR}}$ in decreasing LSFC order.

2) Pilot selection: The pilots are assigned to each UE by random pilot assignment (RPA) or some strategic PA methods. Here we propose a strategic PA based on a weighted graphic framework (WGF) based PA proposed in [12]. At first, we introduce WGF-based heuristic PA from [12], where subspace orthogonality is not considered. The WGF scheme

consists of two main phases: the construction of a weighted pilot contamination graph and Max k-Cut PA. The aim of the Max k-Cut algorithm is finding the optimal τ_p co-pilot UEs sets $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_{\tau_p}\}$ so that the potential contamination is minimum. However, implementation of the pure Max k-Cut algorithm has high complexity, thus [12] introduces a heuristic approximation of the Max k-Cut. The heuristic approach consists of the following steps, where each variable is explained later.

1) Assign τ_p arbitrarily chosen UEs to τ_p subsets, one in each subset. Temporal subsets are given as $\mathcal{V}_1 = \{\mathrm{UE}_1\}, \ldots, \mathcal{V}_{\tau_p} = \{\mathrm{UE}_{\tau_p}\}$.
2) Select one remaining UE_i, calculate a weight between each subset and UE_i as $W_{i,q} = \sum_{j \in \mathcal{V}_q} \omega_{i,j}$, where $\omega_{i,j}$ denotes a weight between two UEs.
3) Assign the UE to the subset with the smallest increased weight as $q^* = \arg\min_q W_{i,q}$ and update subset as $\mathcal{V}_{q^*} = \mathcal{V}_{q^*} \cup \mathrm{UE}_i$.
4) Iteratively repeat step 2) and 3) until the remaining UEs are assigned.

Firstly, a potential pilot contamination $\omega_{k,k'}$ between the k-th and k'-th UE is defined as

$$\omega_{k,k'} = \left| \frac{\sum_{\ell \in \mathcal{C}_k} \beta_{\ell,k'}}{\sum_{\ell \in \mathcal{C}_k} \beta_{\ell,k}} \right|^2 + \left| \frac{\sum_{\ell \in \mathcal{C}_{k'}} \beta_{\ell,k}}{\sum_{\ell \in \mathcal{C}_{k'}} \beta_{\ell,k'}} \right|^2, \quad (23)$$

where the first term on the right-hand side corresponds to the amount of interference from the k'-th UE to the k-th cluster $\mathcal{C}_k$, and the second term corresponds to interference from the k-th UE to $\mathcal{C}_{k'}$. Secondly, a weight between subset $\mathcal{V}_p$ and $\mathcal{V}_q$ is defined as

$$W_{p,q} = \sum_{i \in \mathcal{V}_p, j \in \mathcal{V}_q} \omega_{i,j}. \quad (24)$$

This scheme tends to assign UEs with high potential interference into the different subsets, thus UEs with potentially high interference are given orthogonal pilots.

For further enhancing the robustness to the contamination, we modify the WGF PA so that the potential interference after the SP becomes small, by taking the subspace information into account in the WGF metric. Using SP with $\mathbf{F}_{\ell,k}\mathbf{F}_{\ell,k}^{\mathsf{H}}$, the new metric is given by

$$\omega_{k,k'}^{\mathrm{SP}} = \left| \sum_{\ell \in \mathcal{C}_k} \frac{||\mathbf{F}_{\ell,k}\mathbf{F}_{\ell,k}^{\mathsf{H}}\mathbf{F}_{\ell,k'}||_F^2}{|\mathcal{S}_{\ell,k}|} \beta_{\ell,k'} \bigg/ \sum_{\ell \in \mathcal{C}_k} \beta_{\ell,k} \right|^2$$
$$+ \left| \sum_{\ell \in \mathcal{C}_{k'}} \frac{||\mathbf{F}_{\ell,k'}\mathbf{F}_{\ell,k'}^{\mathsf{H}}\mathbf{F}_{\ell,k}||_F^2}{|\mathcal{S}_{\ell,k'}|} \beta_{\ell,k} \bigg/ \sum_{\ell \in \mathcal{C}_{k'}} \beta_{\ell,k'} \right|^2 \quad (25)$$

Note that $||\mathbf{F}_{\ell,k}\mathbf{F}_{\ell,k}^{\mathsf{H}}\mathbf{F}_{\ell,k'}||_F^2$ is given by $|\mathcal{S}_{\ell,k} \cap \mathcal{S}_{\ell,k'}|$. If the subspaces of UE k and k' are orthogonal, i.e. $|\mathcal{S}_{\ell,k} \cap \mathcal{S}_{\ell,k'}| = 0$, the potential interference after SP is 0 and $\omega_{k,k'}^{\mathrm{SP}}$ reflects this. One point to consider in using (25) is that these expressions make use of the LSFCs of not associated RU-UE pairs, i.e., $\beta_{\ell,k'}$ for pairs (ℓ, k') such that $\ell \notin \mathcal{C}_{k'}$. Since RU ℓ is not part of the cluster serving user k', such LSFCs may be difficult to be estimated and may not be available. On the other hand, if RU ℓ is not part of $\mathcal{C}_{k'}$ it is likely that $\beta_{\ell,k'}$ is very small (otherwise,

TABLE I
SIMULATION PARAMETERS

Area size (A)	2×2 km^2
Number of UEs (K)	100–1200
Number of RUs and RU antennas ($\{L, M\}$)	$\{10, 40\}, \{25, 16\},$ $\{50, 8\}, \{100, 4\}$
Pilot dimension (τ_p)	10, 20, 30
Maximum cluster size (Q)	10
Cluster formation threshold (η)	1
Pathloss model	3GPP urban microcell channel [15] with 3.7 GHz carrier frequency
Bandwidth per UE	10 MHz
Transmit power per UE	20 dBm
Noise power spectral density	-174 dBm/Hz
Dimension of a RB (T)	200
Angular spread (Δ)	$\pi/8$

TABLE II
THE LARGEST SUM SE WITH PARAMETERS

PA scheme	Sum SE [bit/s/Hz]	K	L	M	τ_p
Non-overloaded PA with PM	738	800	100	4	20
Non-overloaded PA with SP	770	800	100	4	20
SIA-OPA with SP	866	600	25	16	30
R-OPA Random with SP	740	400	50	8	30
R-OPA WGF with SP	841	600	25	16	30
R-OPA WGF with PM	534	200	50	8	30

the RU would likely be part of the cluster). Motivated by this, we introduce a partial LSFC-based WGF metric where only the LSFCs of associated UE-RU pairs are available, i.e.,

$$\overline{\omega}_{k,k'}^{\mathrm{SP}} = \left| \sum_{\ell \in \mathcal{C}_k \cap \mathcal{C}_{k'}} \frac{|\mathcal{S}_{\ell,k} \cap \mathcal{S}_{\ell,k'}|}{|\mathcal{S}_{\ell,k}|} \beta_{\ell,k'} \Big/ \sum_{\ell \in \mathcal{C}_k} \beta_{\ell,k} \right|^2$$
$$+ \left| \sum_{\ell \in \mathcal{C}_{k'} \cap \mathcal{C}_k} \frac{|\mathcal{S}_{\ell,k} \cap \mathcal{S}_{\ell,k'}|}{|\mathcal{S}_{\ell,k}|} \beta_{\ell,k} \Big/ \sum_{\ell \in \mathcal{C}_{k'}} \beta_{\ell,k'} \right|^2 . \quad (26)$$

We use this partial LSFC-based metric in (24) for numerical simulations. The complexity (number of weights calculations and set operations) of the heuristic WGF method is given by $\mathcal{O}(K^2/2 + K/2 + \tau_p)$ [12], where the actual operational complexity depends on the hardware implementation. On the other hand, non-overloaded PA and SIA-OPA are employed in a decentralized manner, where each RU mainly checks (20) or (22) for each pilot and each UE, which are much less computationally complex than the WGF methods and do not require network-wide information exchange.

IV. SIMULATION RESULTS

Table I shows the basic parameter specifications of the simulations. We consider a square coverage area of $A = 2 \times 2$ km^2 with a torus topology by using the wrapped around technique to avoid boundary effects. LSFCs are given according to the 3GPP urban microcell street canyon pathloss model from [15, Table 7.4.1-1], which differentiates between UEs in LOS and NLOS. The total number of receiving antennas is fixed as $LM = 400$. For each set of parameters we generate 40 independent layouts, and small scale fading coefficients are varied 30 times for each layout. RUs and UEs are randomly distributed in each setup, and the sum SE is given by the sum of SE$_k^{\mathrm{ul}}$ over all $k \in [K]$, and then averaging it with respect to the different layouts.

Table II shows the largest sum SE achieved by the different PA schemes and the corresponding system parameters, where R-OPA Random denotes rough overloaded PA using random pilot assignment and R-OPA WGF denotes rough overloaded PA using WGF-based pilot assignment. Firstly, we confirm that $\tau_p = 30$ outperforms $\tau_p = 20$ for the R-OPA and the SIA-OPA even taking into account the overhead in eq. (14), while

Fig. 1. Mean cluster size and outage probability vs. K.

$\tau_p = 20$ is the best for the non-overloaded PA. As a comparison with existing studies, the results of R-OPA with the WGF algorithm and PM channel estimation represent [12]. From Table II, the SIA-OPA with SP achieves the highest sum SE. In the following, we focus on the setup $\tau_p = 30, L = 25, M = 16$ to analyze the behavior around the largest sum SE.

Fig. 1 shows the mean cluster size and outage probability vs. K, and Fig. 2 illustrates the sum SE vs. K. Note that outage occurs in the region of more than 800 UEs for the SIA-OPA and the non-overloaded PA. From Fig. 1, the cluster size with rough overloaded PA (R-OPA) is almost equal to $Q = 10$ for all evaluated K. From this observation we deduce for the given simulation parameters that if there is no limitation on cluster formation imposed by the pilot assignment such as non-overloaded, Q RUs with the largest channel gain will form a user-centric cluster for most UEs. Note that the cluster size with R-OPA is independent of the pilot assignment scheme, since the clusters are formed before assigning pilots. The cluster size of the non-overloaded PA scheme is significantly reduced by the increase of K, while that of the SIA-OPA scheme decreases more gradual. For small channel estimation error, UEs with larger clusters can be given higher SE by obtaining more spatial diversity. Thus, the cluster formation of R-OPA and SIA-OPA have potential to outperform the non-overloaded PA in terms of SE.

Then let us focus on Fig. 2 where Ideal CSI lines indicate
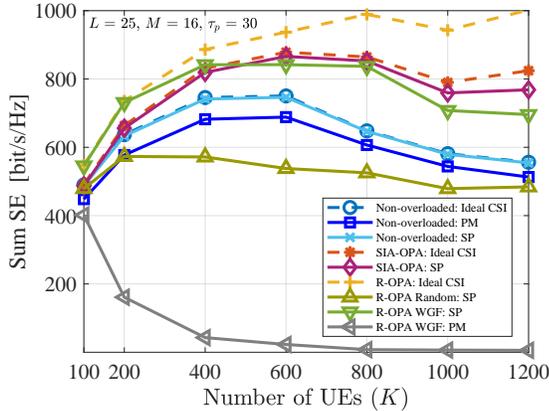
Fig. 2. Sum SE vs. K.

no channel estimation error. The SE degradation factor can be divided into two: channel estimation error due to pilot contamination, and deviation from the optimal multiplexing gain of massive MIMO. The fact that the sum SE saturates in the ideal CSI case indicates the latter factor.

The R-OPA WGF scheme with PM channel estimation is not tolerant of an increase in UE density. This is an interesting effect when compared to the non-overloaded PA with PM, where the sum SE first grows with the number of UEs, and then decreases again. The degradation of the R-OPA WGF scheme with PM is mainly due to an increased channel estimation error, while the non-overloaded PM is specifically designed to avoid assigning the same pilot to UEs that can potentially cause large mutual pilot contamination. The R-OPA WGF scheme allows co-pilot UEs with large potential interference if their subspaces are mutually orthogonal. Since the PM channel estimation only eliminates pilot contamination from UEs with a different UL pilot (and does not take into account the channel subspaces), the channel estimate is potentially contaminated to a large extent by co-pilot users with orthogonal subspaces, which leads to a strong degradation of the sum SE for growing K.

The non-overloaded PA with the SP achieves almost the same sum SE as the ideal case, which means that the SP effectively eliminates the contamination. For the R-OPA, the performance of random assignment significantly degrades compared to ideal CSI as K increases due to that randomly assigned pilots generate many co-pilot UEs within the same subspace. The WGF-based R-OPA also degrades compared to the ideal case, however, it achieves the highest sum SE at $K \leq 400$ among all PA methods. At last, SIA-OPA slightly outperforms the R-OPA at $K \geq 600$ and achieves the highest sum SE at $K = 600$. In addition, the gap between the SP and the ideal CSI stays very small, thus the SIA-OPA works as expected. In the region of $K \geq 800$, the sum SE saturates and slightly degrades, which can be considered the effect of increased outage as seen in Fig. 1.

The WGF R-OPA has the highest sum SE at a certain UE density ($K \leq 400$), where the per UE SE (sum SE/K) is the largest. On the other hand, the SIA-OPA achieves a competitive sum SE at $K = 400$ and is easier to implement in terms of scalability than WGF R-OPA.

V. CONCLUSIONS

This paper has investigated overloaded PA and cluster formation methods that outperform the non-overloaded PA in CF-mMIMO systems. The WGF R-OPA achieved the highest sum SE in the region of a certain UE density ($K \leq 400$), while it is challenging for practical implementation due to the unscalable metric calculation and information exchange. The SIA-OPA, which assigns pilots to UEs with orthogonal subspaces, achieves the highest sum SE in the region of $K \geq 600$. Furthermore, it also achieves a competitive sum SE at $K \geq 400$ with a more relaxed PA procedure. The appropriate method in a practical system can be chosen according to the trade-off between their performance and feasibility.

REFERENCES

[1] A. D. Wyner, "Shannon-theoretic approach to a gaussian cellular multiple-access channel," *IEEE Trans. on Inform. Theory*, vol. 40, no. 6, pp. 1713–1727, 1994.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Comm.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[3] F. Göttsch, N. Osawa, T. Ohseki, K. Yamazaki, and G. Caire, "The impact of subspace-based pilot decontamination in user-centric scalable cell-free wireless networks," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021, pp. 406–410.

[4] F. Göttsch, N. Osawa, T. Ohseki, K. Yamazaki, and G. Caire, "Uplink-downlink duality and precoding strategies with partial csi in cell-free wireless networks," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 614–619.

[5] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing—The Large-Scale Array Regime," *IEEE Trans. on Inform. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[6] F. Göttsch, N. Osawa, T. Ohseki, K. Yamazaki, and G. Caire, "Subspace-based pilot decontamination in user-centric scalable cell-free wireless networks," 2022. [Online]. Available: https://arxiv.org/abs/2203.00714

[7] E. Björnson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Trans. on Comm.*, vol. 68, no. 7, pp. 4247–4261, 2020.

[8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.

[9] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive mimo based on the hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, 2021.

[10] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, "Structured massive access for scalable cell-free massive mimo systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1086–1100, 2021.

[11] W. H. Hmida, V. Meghdadi, A. Bouallegue, and J.-P. Cances, "Graph coloring based pilot reuse among interfering users in cell-free massive mimo," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[12] W. Zeng, Y. He, B. Li, and S. Wang, "Pilot assignment for cell free massive mimo systems using a weighted graphic framework," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6190–6194, 2021.

[13] Ö. T. Demir, E. Björnson, L. Sanguinetti *et al.*, "Foundations of User-Centric Cell-Free Massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021.

[14] E. Björnson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Trans. on Comm.*, vol. 68, no. 7, pp. 4247–4261, 2020.

[15] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," 3GPP Technical Specification 38.901, 12 2019, Version 16.1.0.