WILEY | Hindawi

Research Article

Deployment of Clustered-Based Small Cells in Interference-Limited Dense Scenarios: Analysis, Design, and Trade-Offs

David Alejandro Urquiza Villalonga [1], **Felip Riera-Palou** [2],
M. Julia Fernández-Getino García [1], **and Guillem Femenias** [2]

[1]*Department of Signal Theory and Communications, Carlos III University of Madrid, 28911, Leganés, Madrid, Spain*
[2]*Mobile Communications Group, Department of Mathematic and Informatics, University of the Balearic Islands, 07122 Majorca, Spain*

Correspondence should be addressed to David Alejandro Urquiza Villalonga; daurquiza@tsc.uc3m.es

Network densification is one of the most promising solutions to address the high data rate demands in 5G and beyond (B5G) wireless networks while ensuring an overall adequate quality of service. In this scenario, most users experience significant interference levels from neighbouring mobile stations (MSs) and access points (APs) making the use of advanced interference management techniques mandatory. Clustered interference alignment (IA) has been widely proposed to manage the interference in densely deployed scenarios with a large number of users. Nonetheless, the setups considered in previous works are still far from the densification levels envisaged for 5G/B5G networks that are considered in this paper. Moreover, prior designs of clustered-IA systems relied on oversimplified channel models and/or enforced single-stream transmission. In this paper, we explore an ultradense deployment of small cells (SCs) to provide coverage in 5G/B5G wireless networks. A novel cluster design based on a size-restricted k-means algorithm to divide the SCs into different clusters is proposed taking into account path loss and shadowing effects, thus providing a more realistic solution than those available in the current literature. Unlike previous works, this clustering method can also cater for spatial multiplexing scenarios. Also, several design parameters such as the number of transmit antennas, multiplexed data streams, and deployed APs are analyzed in order to identify trade-offs between performance and complexity. The relationship between density of network elements per area unit and performance is investigated, thus allowing to illustrate that there is an optimal coverage area value over which the network resources should be distributed. Moreover, it is shown that the spectral-efficiency degradation due to the intercluster interference in ultradense networks (UDNs) points to the need of designing an interference management algorithm that accounts for both intracluster and intercluster interferences. Simulation results provide key insights for the deployment of small cells in interference-limited dense scenarios.

1. Introduction

5G and beyond (B5G) wireless communication networks will strive for a capacity and coverage that are beyond the capabilities of current standards. Therefore, new radio access techniques and deployment strategies are being implemented by several standard development organizations and research groups to meet the demanding requirements of key performance indicators (KPIs) related to spectral efficiency (SE), delay, or outage probability, to name a few. Network densification, achieved by deploying a large number of small cells (SCs) throughout the coverage area, is one of the most promising solutions to boost the performance of wireless networks. SCs are based on the use of low-power and reduced-size wireless access points (APs) that provide short-range connectivity to nearby mobile stations (MSs). It has been shown that a dense deployment of SCs has the potential to increase capacity, extend network coverage, and improve spectrum utilization with reduced power consumption [1]. However, several technical challenges need to be addressed

to achieve the promised performance gains in an ultradense network (UDN) setting, including the efficient management of high levels of interference, the implementation of reliable and cost-effective backhaul links, and the design of efficient protocols for the management of initial access, mobility, and handover [1, 2]. In the following, we concentrate on the first of the challenges, which is interference management, notwithstanding the importance of the rest of mechanisms.

Advanced interference management techniques are required to combat the effects caused by the multiple concurrent transmissions taking place in a reduced area when considering a UDN scenario. Interference alignment (IA) has gained a lot of attention as one of the most promising interference management techniques for SCs given its capability to achieve optimal degrees of freedom (DoFs) [3, 4]. IA has been applied in several scenarios such as cognitive radio (CR), device-to-device (D2D), and cellular and heterogeneous networks (HetNets) [5, 6]. Although a closed-form expression to compute the precoder and decoder matrices according to the IA strategy is very difficult to be obtained in most scenarios, several linear iterative algorithms allow a practical implementation that exploits the availability of multiple antennas both at the transmitter and the receiver sides [7–9]. The main idea behind linear IA algorithms is to iteratively design the precoder and decoder matrices of all network users so that, for each user, the received interference signals are projected into certain subspaces orthogonal to the decoder matrix at the receiver side.

In a recent previous work [10], a performance comparison of different IA algorithms reported in scientific literature, such as minimum interference leakage (MIN-IL) [7], Gauss-Newton interference leakage (GN-IL) [9], maximum signal to interference and noise ratio (MAX-SINR) [7], and sum mean squared error (SUM-MSE) [8], is addressed. It is numerically shown that MAX-SINR achieves a good performance for low and medium signal to noise ratio (SNR) values, which is the scenario most often found in UDN settings.

The number of noninterfering data streams that can be simultaneously transmitted by using IA algorithms over an interference channel is directly related to the number of users in the network and the amount of transmit and receive antennas [11]. An IA algorithm is only feasible when all the interference signals are cancelled, and the desired data is properly recovered. Therefore, designing the precoder and decoder matrices of all the APs and MSs in the entire network by solving a single system-wide IA problem is not possible in practical densely deployed scenarios with a large number of SCs. In this context, the IA feasibility conditions that are difficult to be satisfied provided the large number of antennas required at each AP/MS to be able to align all the interference signals. In addition, a large amount of channel state information (CSI) overhead must be feedbacked to compute a global IA solution, which decreases the efficiency of the network. To overcome these problems, several authors have proposed the idea of dividing the network into several clusters with reduced size based on the fact that the strength of the received signals depends on propagation characteristics such as large-scale fading [12–19]. Towards this end, users are grouped in clusters such that the stronger interference links are captured in the same cluster, thus potentially generating a large amount of (intracluster) interference. These interfering transmissions can be efficiently suppressed by applying IA independently in each cluster. In contrast, the intercluster interference caused by geographically distant users is typically maintained weak so that it can be safely ignored. Therefore, a proper cluster design strategy is essential to increase the capacity of the network.

The idea of clustered IA is introduced in [12] to evaluate the applicability of IA in the emerging cellular networks with a high population of users. With this aim, the average throughput gains provided by clustered IA over the traditional noncooperating scheme with frequency reuse are evaluated considering a scenario made of 37 hexagonal cells each with 1 km of radius and 7 transmitter/receiver pairs per cell, a numerology still far from the densification levels to be found in 5G/B5G networks. In a later work [13], the same authors also evaluate the performance gains in terms of outage probability for large ad hoc networks. Nonetheless, the cluster design strategy was not clearly described in these seminal papers. In order to fill this gap, different cluster design algorithms have been developed in [14–19].

In [14], the authors propose a novel clusterization method based on graph partitioning for a multiuser interference network. An adaptation of the MAX-SINR algorithm based on diagonal loading (MAX-SINR-DL) is also described to consider both intracluster and intercluster interferences in the IA formulation, leading to improved rate performance. Finally, a soft-clustering design relaxing the IA feasibility constraint is proposed achieving higher performance for a high SNR regime. Numerical results are provided for a maximum of 15 transmitter/receiver pairs uniformly distributed over a 10×10 km square area. A different clustering strategy, based on the hedonic coalition formation game, is developed in [15] for large-scale cellular networks. Nevertheless, the cluster design in these research works assumes the use of single data stream transmission, thus limiting its applicability to scenarios without spatial multiplexing. Furthermore, recent papers have applied resource allocation with clustered IA to also reduce the intercluster interference and hence to boost the achievable SE in UDNs [16, 17]. However, the aforementioned research works propose cluster design algorithms that are based on oversimplified channel models that only take into account distance-dependent path loss propagation losses and do not consider the shadowing effect.

Clustered IA has also been exploited for D2D communications in [18, 19], where three grouping schemes based on fuzzy C-Means are proposed. In these papers, however, the precoder/decoder pairs are computed based on IA with symbol extensions, which is not practical design criteria because they require exponentially long symbol extensions proportional to the number of users and do not properly work under constant or slowly varying channels [3, 5, 14]. In this context, a partially connected IA is proposed in [6] for cellular and D2D communication networks to select the proper interference links to be aligned with IA while the residual weak interference, considered as noise, is managed by a power optimization method. However, a clustering algorithm is not considered, and the proposed method only applies IA

to the most powerful interference links that can be dealt with the available spatial dimension. Therefore, this algorithm is not viable for UDNs with few antennas at each AP since the level of interference that cannot be handled will severely impair the performance of the network.

1.1. Contribution. According to the works described above, the relevance of the study of IA in a UDN context with 5G/B5G numerology and the design of new efficient radio access strategies becomes evident. In this paper, the capacity of SCs to provide coverage in 5G/B5G wireless networks is explored under the umbrella of the IA paradigm. A fully distributed topology as in [20–22] is considered, where there are no macro base stations and a large number of APs act as small base stations to provide communication services to a much smaller number of MSs. The system is modeled as a downlink K-user interference channel network, where each MS receives data from a single AP [6, 7]. Moreover, each MS and its corresponding AP are considered an SC as in [20–22]. The transmission within an SC causes interference to the remaining users. We note that the network architecture is somewhat similar to that considered in the cell-free related literature (see [20–22]); however, the studied setup only requires for APs/MSs within a cluster to share their CSI whereas cell-free study requires the distribution of all users' data to all APs. Furthermore, we consider multiple antennas at the transmitter and receiver, thus enabling the incorporation of precoding/combining techniques at both sides to increase the achievable SE of the network.

According to the deployment scenario considered in this paper, the proposed solution is divided into two general stages. In the first one, each MS is associated with the AP that exhibits the smallest large-scale channel loss. Then, in the second stage, a clustered-IA method is applied to cancel the interference signals. An IA iterative algorithm is implemented due to its optimality in terms of the number of DoFs. However, since in most practical scenarios the concurrent transmission links outnumber the amount of transmit and receive antennas, the fully connected IA feasibility conditions are not satisfied. Thus, a clusterization method is applied to divide the AP/MS pairs into different clusters taking into account similarities in propagation losses rather than proximity. The aim is to ensure that links causing the strongest interference levels belong to the same cluster and thus can be subsequently cancelled. Interference is only taken into account in the clustered-IA stage because considering the SINR to perform the AP/MS pair assignment and then to form the clusters would involve an iterative process with prohibitive complexity. Note that, in practical applications, the use of the SNR to decide the AP/MS pair is a common practice in the literature [20–22].

A novel cluster design is proposed based on the k-means algorithm since it is one of the most widely employed methods due to its speed, simplicity, and performance. K-means algorithm groups a set of objects into different clusters such that the distance function of each object with respect to the centroid of the cluster is minimized [23]. In this paper, the set of objects is defined by the large-scale propagation losses of the channel matrix including path loss and cor-

related shadowing which guarantees that the users potentially causing the highest levels of interference to each other are located in the same cluster. Therefore, the proposed cluster design provides, to the best of our knowledge, a more realistic solution than those available in the literature that only consider the distance-dependent path loss. In the conventional k-means algorithm, a restriction on the number of elements inside a cluster is not considered in the formulation. Consequently, the IA feasibility conditions may not be satisfied in some clusters due to the limited number of available transmit and receive antennas. To overcome this issue, we design a size-restricted k-means algorithm to constrain the maximum number of AP/MS pairs inside the same cluster. The proposed constrained k-means algorithm is less complex than those described in [24, 25].

A performance comparison is carried out between a network of SCs using singular value decomposition- (SVD-) based precoding/combining schemes and a network of SCs using IA algorithms based on either the traditional MAX-SINR approach or the MAX-SINR-DL strategy proposed in [14] to illustrate the advantage of using interference-targeting schemes that account for both intracluster and intercluster interferences. Note that SVD is the optimal linear beamforming strategy for a MIMO point-to-point transmission [26] and it is widely used in such circumstances (e.g., Wi-Fi standard IEEE 802.11 ac/ax). However, its performance is severely degraded as infrastructure and user deployment becomes denser. It is therefore interesting to evaluate the relative merits of IA-based processing and SVD beamforming under different levels of densification and taking into account performance and complexity. Furthermore and in contrast to [12–15], we consider multiple data stream transmission based on the results in [11] for a linear iterative IA algorithm. The impact of the cluster size on the average achievable rate of the networks is fully assessed as well as that of the number of antennas at each communication end alongside the network load. Also, the effect the AP and/or MS density might have on the system performance is investigated, thus allowing to illustrate the relationship between network elements per area unit and performance. Numerical results show that there is an optimal coverage area value over which the network resources should be distributed.

1.2. Paper Organization. This paper is organized as follows. The system model and problem formulation are presented in Section 2. In Section 3, the clustered IA solution to cancel intracluster interference is fully described. The proposed cluster design is detailed in Section 4. Numerical results are presented in Section 5 that serve to validate the proposed SC deployment. Finally, Section 6 concludes the paper and provides hints for further research.

1.2.1. Notation. Boldface lower-case letters are used for vectors, while boldface upper-case letters are used for matrices. $\mathbf{A}^T$, $\mathbf{A}^H$, and rank $(\mathbf{A})$ represent transpose, Hermitian transpose, and the rank of matrix $\mathbf{A}$, respectively. $\mathbf{A}_l$ denotes the lth column vector of matrix $\mathbf{A}$. $\mathbb{C}^{M \times N}$ is the space of complex $M \times N$ matrices, and $\mathscr{CN}(\mu, \sigma^2)$ is the complex Gaussian distribution with mean μ and variance σ^2. $\mathbb{E}(\cdot)$

stands for expectation. min (a, b) denotes the smallest element between a and b, and $\lceil a \rceil$ indicates the smallest integer larger than or equal to a. $\mathbf{I}_d$ is the $d \times d$ identity matrix.

2. System Model

Let us consider a network of densely deployed SCs similar to the one presented in [20–22] and which can be considered illustrative of the scenarios to be found in B5G networks. In particular, we assume a network comprised of M APs each equipped with N_t antennas that simultaneously provide service, on the same time-frequency resource, to K MSs equipped with N_r antennas. Similar to [20–22], a scenario with a larger number of APs than MSs is considered, as illustrated in Figure 1. In addition, each AP is assumed to be connected to a base station controller (BSC) by means of a fronthaul link. Although unlimited capacity and error-free fronthaul links are considered in this paper, the use of finite-capacity fronthaul links will be considered in future works similar to what was done in [27].

2.1. Channel Model. The classical channel propagation model considered in previous papers for clustered IA [14–19] oversimplifies the real radio environment. To deal with this issue, we assume the 3GPP Urban Microcell model described in [28] that takes into account the small-scale fading, large-scale propagation losses (i.e., path loss and shadowing), and the possibility of the channel to toggle between non-line-of-sight (NLOS) and line-of-sight (LOS) propagation conditions. The MIMO channel matrix describing the propagation between the ith AP and the kth MS is modeled as

$$\mathbf{H}^{[k,i]} = \sqrt{\beta^{[k,i]}} \mathbf{G}^{[k,i]}, \qquad (1)$$

where $\mathbf{H}^{[k,i]} \in \mathbb{C}^{N_r \times N_t}$, $i = \{1, 2, \cdots, M\}$ and $k = \{1, 2, \cdots, K\}$. The scalar $\beta^{[k,i]}$ represents the large-scale propagation losses, and $\mathbf{G}^{[k,i]} \in \mathbb{C}^{N_r \times N_t}$ denotes the matrix of small-scale fading coefficients. The large-scale fading is further decomposed as $\beta^{[k,i]} = \zeta^{[k,i]} \chi^{[k,i]}$ with $\chi^{[k,i]}$ corresponding to the shadowing component and $\zeta^{[k,i]}$ representing the distance-dependent path loss defined as [28]

$$\zeta^{[k,i]}[\text{dB}] = -\left(\zeta_0 + 10\alpha \log_{10}\left(r^{[k,i]}\right)\right), \qquad (2)$$

where ζ_0 is the path loss at a reference distance of 1 m, α is the path loss exponent, and $r^{[k,i]}$ is the three-dimensional (3D) distance between the ith AP and the kth MS where the height of each communication node, AP or MS, is also taken into account. Regarding the shadowing, $\chi^{[k,i]}$ follows the model described in ([20], (54)-(55)), and it is modelled as a spatially correlated log-normal random variable with variance σ_χ^2. Parameters ζ_0, α, and σ_χ^2 exhibit different behaviours depending on whether the link is in LOS or NLOS conditions. In particular, the channel between the ith AP and the kth MS will be in LOS according to the probability [29].

$$p_{\text{LOS}}\left(r^{[k,i]}\right) = \min\left(1, \frac{r_0}{r^{[k,i]}} + \left(1 - \frac{r_0}{r^{[k,i]}}\right)e^{-r^{[k,i]}/2r_0}\right), \qquad (3)$$

where r_0 is a reference distance. LOS propagation links are characterized by a Ricean K-factor $K^{[k,i]}$ and holding $10 \log_{10}(K^{[k,i]}) \sim \mathcal{N}(\mu_K, \sigma_K^2)$. For NLOS links, it holds that $K^{[k,i]} = 0$. Finally, the small-scale fading terms $\mathbf{G}^{[k,i]}$ consist of independent and identically distributed (i.i.d.) complex Gaussian random variables distributed as $\mathscr{CN}(0, 1)$. The channel coefficients $\mathbf{H}^{[k,i]}$ are assumed to be static throughout the coherence interval and then change independently (i.e., block fading). The channel model assumed in this paper is a more realistic representation than those proposed in previous works in the scientific literature since it considers both the path loss and correlated shadowing effects. We note at this point that the inclusion of spatial correlation effects of the antenna array is left for a future work.

2.2. AP Selection. Similar to the SC setup introduced in [20–22], we assume that each AP can serve only one MS, and hence, an AP selection procedure is required. Each AP is randomly selected, on a per-MS basis, such that each MS is served only by the available AP with the smallest channel loss $\beta^{[k,i]}$. Once an AP is selected, it becomes unavailable, and when all MSs have been paired with an AP, the idle APs are turned into a sleep mode to save energy. We consider that each MS and its corresponding AP form a reduced SC composed of a single transmitter/receiver pair. However, to simplify the notation and based on the terminology used in [20–22], each AP/MS pair will be referred to as an SC instance.

Each selected AP is intended to transmit d data streams to its corresponding MS, with the transmission of the kth AP/MS pair causing interference to the other $K - 1$ MSs. Since we assume a densely deployed network with a high number of SCs, a clustered IA strategy is implemented to cancel the interference signals. A limited amount of AP/MS pairs showing the highest levels of mutual interference are grouped into the same cluster in such a way that the intercluster interference is as weak as possible. This strategy is reminiscent of the one introduced in [30] where users clustered together are assigned orthogonal pilot sequences, thus leading to a minimization of the intracluster interference. In our IA-based setup, intracluster interference is completely eliminated using a linear iterative IA algorithm whereas the intercluster interference is expected to be relatively small since it comes from geographically distant users.

As it is typically postulated in most state-of-the-art wireless networks, downlink and uplink transmissions are organized in a time division duplex (TDD) operation assuming the reciprocity of the propagation channel $\overleftarrow{\mathbf{H}}^{[i,k]} = \mathbf{H}^{[k,i]H}$, where $\overleftarrow{\mathbf{H}}^{[i,k]}$ represents the reciprocal channel between the kth MS and the ith AP [7, 31]. Global information about large-scale fading between all AP-to-MS links (i.e., $\beta^{[k,i]}$, $\forall k, i$) is required for the clustering stage. Similar to [14], this large-scale information is obtained at each AP by measuring the averaged signal power of the pilot signals transmitted by each MS in the uplink training phase. Then, this information
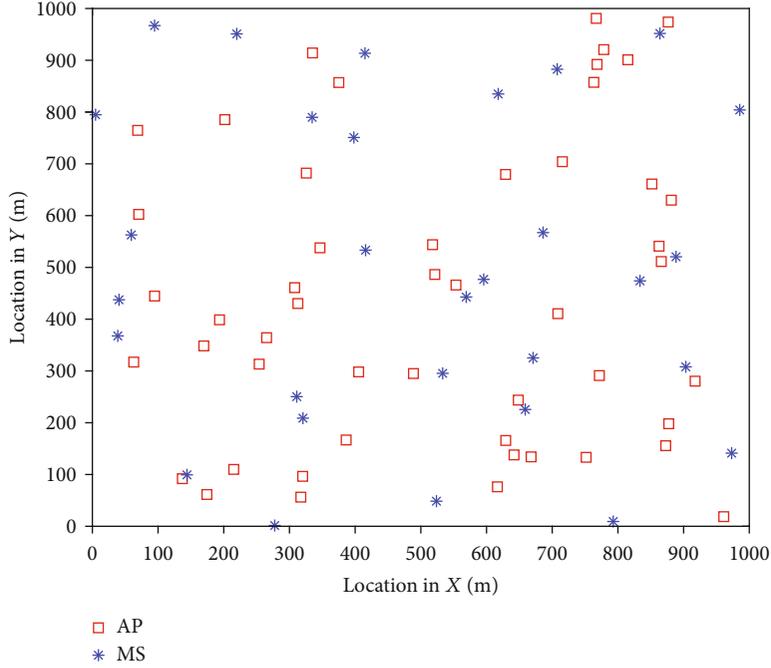
FIGURE 1: Illustration of a dense deployment scenario with 50 APs and 30 MSs randomly distributed within a square of $1 \, \text{km}^2$.

is sent to the BSC to perform the cluster design. As it is typically assumed in wireless communications, large-scale fading varies at a much slower pace than small-scale fading and, hence, remains virtually constant for many coherence periods [20, 32]. Therefore, the cluster design only needs to be updated when the large-scale fading conditions change, thus reducing signal overhead and complexity. Finally, the signal overhead for instantaneous CSI estimation is limited to AP/MS pairs inside the same cluster. Subsequently, these estimated channels are used by the IA algorithm to compute the precoder/decoder matrices. For brevity of exhibition, this paper focuses on the downlink only, notwithstanding the fact that most of the discussion also applies to the uplink segment. Furthermore, we assume that the estimation error is negligible, and hence, perfect CSI can be considered as in [12–15, 18, 19]. We note, however, that the design of robust IA algorithms to deal with channel estimation errors constitutes a promising line for further research.

3. Clustered Interference Alignment Algorithm

IA algorithms carefully design the precoder matrix $\mathbf{V}^{[k]} \in \mathbb{C}^{N_t \times d}$ at the AP side and the decoder matrix $\mathbf{U}^{[k]} \in \mathbb{C}^{N_r \times d}$ at the MS to avoid intracluster interference. $\mathbf{V}^{[k]}$ is designed to constrain all the interference signals into a subspace orthogonal to $\mathbf{U}^{[k]}$, while the desired signal is allocated into another subspace free from interference. IA algorithms have to satisfy the following feasibility conditions to suppress the interference and recover the desired signals [7]

$$\mathbf{U}^{[k]H}\mathbf{H}^{[k,j]}\mathbf{V}^{[j]} = 0, \quad \forall j \neq k,$$
$$\text{rank}\left(\boldsymbol{U}^{[k]H}\boldsymbol{H}^{[k,k]}\boldsymbol{V}^{[k]}\right) = d, \quad \forall k, \tag{4}$$

where $j = \{1, 2, \cdots, K\}$ represents the indexes of the selected APs.

The characterization of design parameters to ensure that IA algorithms meet feasibility conditions has been widely addressed in a scientific literature [11, 33–35]. The analysis provided in these papers shows that the ability of the IA algorithm to satisfy the previous constraints depends on the number of users K, the spatial dimension of the MIMO channel (number of antennas at transmitter (N_t) and receiver (N_r) side), and the number of data streams d to be multiplexed for each AP/MS pair. Results in [33–35] only address a partial description of the problem, applicable to specific values of the parameters K, d, N_t, and N_r. Based on these former results, the authors in [11] provide a full characterization of the feasibility of linear IA for a symmetric MIMO channel with constant coefficients and no symbol extension. Equation (26) in [11] describes the maximum achievable DoFs (d^*) as a function of the number of users and the number of antennas at the transmitter and receiver as $d^* = f(\lambda(K), \gamma(N_t, N_r))$, where $f(\lambda, \gamma)$ is a function that depends on $\lambda = 1/2(K - 1 - \sqrt{(K-1)^2 - 4})$ and $\gamma = N_r/N_t$, assuming that $N_r \leq N_t$ (see [11] for the details). In our clustered IA context, this equation offers us the framework to set the cluster size constraint.

Let $\mathbb{K} = \{1, 2, \cdots k, \cdots, K\}$ be the set of indices of the K AP/MS pairs. Moreover, let $\mathbb{A} = \{\mathscr{A}_1, \mathscr{A}_2, \cdots \mathscr{A}_c, \cdots, \mathscr{A}_{n_C}\}$ denote the set of clusters, where $\mathscr{A}_c$ contains the indices of the AP/MS pairs that are in the cth cluster, $c = \{1, 2, \cdots, n_C\}$, and n_C is the total number of clusters. For the given values of d, N_t, and N_r, the maximum number of SCs in each cluster, denoted as $S_{\max}$, is computed such that $d \leq d^* = f(\lambda (S_{\max}), \gamma(N_t, N_r))$. Then, denoting the size of each cluster as $\mathbb{S} = \{\mathscr{S}_1, \mathscr{S}_2, \cdots \mathscr{S}_c, \cdots, \mathscr{S}_{n_C}\}$, where each size represents

the cardinality of the cluster (i.e., $\mathscr{S}_c = |\mathscr{A}_c|$), the cluster design has to satisfy the following properties [14]:

$$\mathscr{A}_c \subseteq \mathbb{K}, \quad \forall c, \tag{5a}$$

$$\mathscr{A}_c \cap \mathscr{A}_m = \varnothing, \quad \forall c \neq m, \tag{5b}$$

$$\bigcup_{c=1}^{n_C} \mathscr{A}_c = \mathbb{K}, \tag{5c}$$

$$\mathscr{S}_c \leq S_{\max}, \quad \forall c. \tag{5d}$$

Let $\mathscr{A}_{c_k}$ denote the cth cluster where the kth AP/MS pair is located. Then, the received signal at the kth MS is given by

$$\mathbf{y}^{[k]} = \mathbf{U}^{[k]H} \mathbf{H}^{[k,k]} \mathbf{V}^{[k]} \mathbf{s}^{[k]} + \overbrace{\sum_{j \in \mathscr{A}_{c_k}, j \neq k}^{K} \mathbf{U}^{[k]H} \mathbf{H}^{[k,j]} \mathbf{V}^{[j]} \mathbf{s}^{[j]}}^{\text{intracluster interference}}$$

$$+ \underbrace{\sum_{j \notin \mathscr{A}_{c_k}}^{K} \mathbf{U}^{[k]H} \mathbf{H}^{[k,j]} \mathbf{V}^{[j]} \mathbf{s}^{[j]}}_{\text{intercluster interference}} + \mathbf{U}^{[k]H} \mathbf{n}^{[k]}, \quad k \in \{1, \cdots, K\}, \tag{6}$$

where $\mathbf{s}^{[k]} \in \mathbb{C}^{d \times 1}$ is the d-stream transmitted symbol vector to user k and $\mathbf{n}^{[k]} \in \mathbb{C}^{N_r \times 1}$ is the circularly symmetric additive white Gaussian noise (AWGN) vector at the kth MS ($\mathbf{n}^{[k]} \sim \mathscr{CN}(0, \sigma_n{}^2 \mathbf{I}_{N_r})$). The transmitted signal power by the jth AP is given by $P^{[j]} = \mathbb{E}[\|\mathbf{s}^{[j]}\|^2]$, and it is subject to a per-AP power constraint $P^{[j]} \leq P_T \forall j$.

The SE of the kth transmitter-receiver pair can be determined as [8, 36]

$$R^{[k]} = \sum_{l=1}^{d} \log_2 \left(1 + \text{SINR}^{[k,l]}\right), \tag{7}$$

where $\text{SINR}^{[k,l]}$ is the SINR at the kth receiver and lth data stream, $l = \{1, 2, \cdots, d\}$ and it can be expressed as

$$\text{SINR}^{[k,l]} = \frac{P^{[k]}}{d} \frac{\mathbf{U}_l^{[k]H} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]} \mathbf{V}_l^{[k]H} \mathbf{H}^{[k,k]H} \mathbf{U}_l^{[k]}}{\mathbf{U}_l^{[k]H} \left(\mathbf{B}_{\text{intra}}^{[k,l]} + \mathbf{B}_{\text{inter}}^{[k]}\right) \mathbf{U}_l^{[k]}}, \tag{8}$$

where

$$\mathbf{B}_{\text{intra}}^{[k,l]} = \sum_{j \in \mathscr{A}_{c_k}} \frac{P^{[j]}}{d} \mathbf{H}^{[k,j]} \mathbf{V}^{[j]} \mathbf{V}^{[j]H} \mathbf{H}^{[k,j]H}$$

$$- \frac{P^{[k]}}{d} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]} \mathbf{V}_l^{[k]H} \mathbf{H}^{[k,k]H} + \sigma_n{}^2 \mathbf{I}_{N_r} \tag{9}$$

is the intracluster interference plus noise covariance matrix at the kth MS and lth data stream preceding the decoding stage. The intercluster interference covariance matrix at the kth MS before the decoding stage is computed as

$$\mathbf{B}_{\text{inter}}^{[k]} = \sum_{j \notin \mathscr{A}_{c_k}} \frac{P^{[j]}}{d} \mathbf{H}^{[k,j]} \mathbf{V}^{[j]} \mathbf{V}^{[j]H} \mathbf{H}^{[k,j]H}. \tag{10}$$

One of our aims in this paper is to analyze how the average SE achieved for a given AP/MS pair is affected by several practical design parameters.

3.1. Max-SINR. Owing to its superior performance at low and intermediate SINR values, the MAX-SINR IA algorithm will be considered in this research work [7]. This is a linear iterative algorithm that considers the direct channel gain, the interference leakage, and the noise power to develop an alternating maximization of the SINR per transmitted data stream. Similar to [7], but considering the clustered IA scenario, the unitary decoder vector $\mathbf{U}_l^{[k]}$ that maximizes $\text{SINR}^{[k,l]}$ is given by

$$\mathbf{U}_l^{[k]} = \frac{\left(\mathbf{B}_{\text{intra}}^{[k,l]}\right)^{-1} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]}}{\left\|\left(\mathbf{B}_{\text{intra}}^{[k,l]}\right)^{-1} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]}\right\|}. \tag{11}$$

The transmission in the reciprocal network $\overleftarrow{\mathbf{H}}^{[j,k]}$ is considered to compute $\mathbf{V}^{[j]}$ in a similar way. In this case, the intracluster interference plus noise covariance matrix is given by

$$\overleftarrow{\mathbf{B}}_{\text{intra}}^{[j,l]} = \sum_{k \in \mathscr{A}_{c_j}} \frac{P^{[k]}}{d} \mathbf{H}^{[k,j]H} \mathbf{U}^{[k]} \mathbf{U}^{[k]H} \mathbf{H}^{[k,j]}$$

$$- \frac{P^{[j]}}{d} \mathbf{H}^{[j,j]H} \mathbf{U}_l^{[j]} \mathbf{U}_l^{[j]H} \mathbf{H}^{[j,j]} + \sigma_n{}^2 \mathbf{I}_{N_t}, \tag{12}$$

and the intercluster interference is computed as

$$\overleftarrow{\mathbf{B}}_{\text{inter}}^{[j]} = \sum_{k \notin \mathscr{A}_{c_j}} \frac{P^{[k]}}{d} \mathbf{H}^{[k,j]H} \mathbf{U}^{[k]} \mathbf{U}^{[k]H} \mathbf{H}^{[k,j]}. \tag{13}$$

Then, the unitary precoder vector is given by

$$\mathbf{V}_l^{[j]} = \frac{\left(\overleftarrow{\mathbf{B}}_{\text{intra}}^{[j,l]}\right)^{-1} \mathbf{H}^{[j,j]H} \mathbf{U}_l^{[j]}}{\left\|\left(\overleftarrow{\mathbf{B}}_{\text{intra}}^{[j,l]}\right)^{-1} \mathbf{H}^{[j,j]H} \mathbf{U}_l^{[j]}\right\|}. \tag{14}$$

This is an iterative process in which the precoder and decoder matrices are iteratively updated until the algorithm converges or the number of iterations reaches a given limit. Unfortunately, due to the nonconvex nature of the optimization problem, the convergence to a global minimum is not guaranteed.

Since MAX-SINR can only remove the intracluster interference, its performance can be severely degraded in highly dense scenarios. Consequently, a modification of the algorithm is proposed in [14] to also take into account the

intercluster interference in the design of the precoding/decoding matrices. This method is known as MAX-SINR-DL, and it is based on considering the aggregate intercluster interference as white noise which is then incorporated to the diagonal of the intracluster interference covariance matrix (i.e., Gaussian hypothesis). Therefore, equation (11) is modified as [14]

$$\mathbf{U}_l^{[k]} = \frac{\left(\mathbf{B}_{\mathrm{intra}}^{[k,l]} + \sum_{j \notin \mathcal{A}_{c_j}} P^{[j]} \beta^{[k,j]} \mathbf{I}_{N_r}\right)^{-1} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]}}{\left\| \left(\mathbf{B}_{\mathrm{intra}}^{[k,l]} + \sum_{j \notin \mathcal{A}_{c_j}} P^{[j]} \beta^{[k,j]} \mathbf{I}_{N_r}\right)^{-1} \mathbf{H}^{[k,k]} \mathbf{V}_l^{[k]} \right\|}. \tag{15}$$

We note from the above formulations that the MAX-SINR method only requires of a global CSI knowledge of all links between APs and MSs inside the same cluster. Therefore, the CSI signal overhead is only limited to the users inside the same cluster. A distributed implementation of MAX-SINR with only local knowledge of CSI is also proposed in [7]. Moreover, according to (15), the MAX-SINR-DL method also needs a global information regarding the propagation losses of the channel (e.g., path loss and correlated shadowing) of all links between APs and MSs (i.e., $\beta^{[k,j]} \forall k, j$). However, this information is as well required to implement the clustering stage. Hence, the practical complexity of the system is left unaltered.

Since intercluster interference is not perfectly eliminated with any of the described algorithms, the proper design of the cluster distribution will determine the performance of the system. The next section addresses the proposal to efficiently group the AP/MS pairs in different clusters to ensure that the residual intercluster interference has a minimal impact on the SE of the network.

4. Cluster Design: Size-Restricted k-Means

The cluster design objective is the partition of the K AP/MS pairs into n_C clusters such that

$$\min_{\mathbb{A}} \quad \sum_{k=1}^{K} \mathrm{Trace}\left[\mathbf{B}_{\mathrm{inter}}^{[k]}\right], \tag{16}$$
$$\mathrm{s.t.} \quad (6).$$

This optimization problem is known to be NP-hard, and therefore, the optimal solution can only be found via an exhaustive search that becomes computationally infeasible even for moderate values of K. Therefore, we propose a suboptimal solution based on the k-means algorithm that partitions a set of objects into a given number of clusters defined by its centroids to minimize a squared-error function [23]. We use the large-scale propagation losses including the path loss and correlated shadowing effect as the set of objects $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K\}$ corresponding to the k-means clustering algorithm such that the vector for the jth AP is given by $\mathbf{x}_j = [\beta^{[1,j]} \beta^{[K,j]}]$. This allows us to group inside the same cluster the APs with similar channel loss with respect to all MSs that in turn are precisely those that cause the

higher intracluster interference. We define $n_C = \lceil K/S_{\mathrm{max}} \rceil$ as the initial number of clusters, and their corresponding centroids $\boldsymbol{\mu}_c$ are computed following the method introduced in [37] to increase speed and accuracy. Then, each AP/MS is associated to the nearest cluster centroid. Formally, this is expressed as

$$\arg \min_{\mathbb{A}} \sum_{c=1}^{n_C} \sum_{j \in \mathcal{A}_c} g(\mathbf{x}_j, \boldsymbol{\mu}_c), \tag{17}$$

where $g(\mathbf{x}_j, \boldsymbol{\mu}_c)$ is a function used to measure the distance between the propagation loss vector x_j and the cluster centroid $\boldsymbol{\mu}_c$ that can also be viewed as a square-error function. In this paper, we use the cosine distance defined as $g(\mathbf{x}_j, \boldsymbol{\mu}_c) = 1 - (\mathbf{x}_j \boldsymbol{\mu}_c^T / \sqrt{(\mathbf{x}_j \mathbf{x}_j^T)(\boldsymbol{\mu}_c \boldsymbol{\mu}_c^T)})$. This distance has proved to result in roughly the same clusters as if it was conducted on the geographical positions of the APs/MSs using an Euclidean distance (i.e., clustering according to the AP/MS locations) [30]. The advantage of the clustering approach used in this paper is that it obviates the need to know the geographical locations of the communications nodes and instead relies on the already available large-scale propagation losses.

In each iteration of the k-means algorithm, the centroids are recalculated based on the current objects within each cluster and a new object assignment is developed based on (17). The design of the clusters is iteratively updated until no redistribution of the objects occurs. K-means is reported as the most commonly used clustering strategy, providing a local MMSE value. However, cluster size restriction is not considered in the basic formulation of the algorithm, and hence, there is no guarantee that this method satisfies constraint (5d). Several strategies to balance k-means have been studied to optimize both equal size clusters and local MMSE [24, 25]. In these papers, the assignment phase is different from the original k-means algorithm. Instead of selecting the users to the nearest centroids, a linear programming algorithm [24] or a Hungarian algorithm [25] is applied to ensure equal-size clusters and a local MMSE at the cost of increasing the complexity. However, since we only require to limit the maximum number of SCs in each cluster instead of obtaining equal-size clusters as in [24, 25], we propose a novel size-restricted k-means algorithm that is simpler than previous proposals.

The proposed algorithm is organized in two major steps. First, an initial cluster design is obtained by applying k-means algorithm with $n_{C_{ini}} = \lceil K/S_{\mathrm{max}} \rceil$ clusters. Then, if there is any cluster that does not satisfy constraint (5d), a size-restricted stage is implemented to ensure that each cluster has a maximum size S_{max}. This step is based on splitting the large clusters into smaller ones at the cost of increasing the number of clusters and hence the intercluster interference. A formal description of this method is given in Algorithm 1. The number of clusters is increased by one in each iteration of the size-restricted stage. In this way, the convergence of the method to a cluster distribution that satisfies constraint (5d) is guaranteed.

input : $n_{C_{ini}}$, β, $S_{\max}$, K.

output: $\mathbb{A}$, n_C

1 **// Step 1: Initial cluster design by applying k-means algorithm**

2 Compute an initial cluster distribution by applying k-means algorithm with input arguments: $\mathbb{X}$, $n_{C_{ini}} = \lceil K/S_{\max} \rceil$ and output argument $\mathbb{A}_{ini}$;

3 $\mathbb{A} = \mathbb{A}_{ini}$;

4 $n_C = n_{C_{ini}}$

5 **// Step 2: Size-restricted cluster distribution**

6 Compute the size of each cluster $\mathbb{S}$

7 Find the number of clusters n'_c that violate property (5d) being $\mathcal{S}_c > S_{\max}$

8 $a = 1$;

9 $b = n_C - n'_c$;

10 **while** $n'_c > 0$ **do**

11   Find the indexes set $\mathbb{A}'$ of clusters that violate property (5d) such that $\mathbb{A}' = \{\cdots, \mathscr{A}_{ini_c}, \cdots\}$ $\forall c : \mathcal{S}_c > S_{\max}$;

12   Update the indexes set $\mathbb{A}$ only with clusters that met property (5d) $\{\mathscr{A}_a, \cdots, \mathscr{A}_b\} = \mathbb{A}_{ini} \setminus \mathbb{A}'$;

13   Find the observation set $\mathbb{X}'$ of the APs/MSs pairs that are located in the n'_c clusters that violate property (5d) such that $\mathbb{X}' = \{\cdots, \boldsymbol{x}_j, \cdots\}$ $\forall j : j \in \mathbb{A}' c_j$;

14   Compute an updated cluster distribution by applying k-means algorithm with input arguments: $\mathbb{X}'$, $n_C' + 1$ and output argument $\mathbb{A}_{upd}$;

15   Compute the updated size $\mathbb{S}_{upd}$ of each cluster in the set $\mathbb{A}_{upd}$;

16   Find the new number of clusters n''_c that violate property (5d) being $\mathcal{S}_{upd_c} > S_{\max}$;

17     $\mathbb{A}_{ini} = \mathbb{A}_{upd}$;

18     $\mathbb{S}_{ini} = \mathbb{S}_{upd}$;

19     $a = b + 1$;

20     $b = n_C' + 1 - n_C''$;

21     $n_C = n_C + 1$;

22     $n_C' = n_C''$;

23 **end**

ALGORITHM 1: Size-restricted k-means algorithm.

At every iteration of the size-restricted stage, the k-means algorithm is computed with a lower number of data objects in the observation set $\mathbb{X}'$ and a lower number of clusters. This is due to the fact that once the initial cluster distribution has been obtained, only the objects that belong to the n_C' clusters that do not satisfy property (5d) are used to compute an updated version of the cluster distribution. Note that n_C' value is reduced after each iteration until the method converges. Then, according to [23, 38], the complexity of the initial cluster design in the first step of the Algorithm 1 is $O(n_{C_{ini}} K^2 t)$, where t is the number of iterations for the convergence of the k-means algorithm. It can be noted that the number of objects in the observation set and the dimensionality of the data are both equal to K. Therefore, the total complexity of the proposed size-restricted k-means algorithm is always lower than $O(T n_{C_{ini}} K^2 t)$ for $T > 1$, where T is the number of iterations required to guarantee the convergence of the Algorithm 1, and exactly $O(n_{C_{ini}} K^2 t)$ for $T = 1$. Since the final number of clusters n_C is expected to be much lower than the dimension of the set of objects (i.e., $n_C << K$), then our method has lower complexity than those reported in [24, 25].

The cluster design is performed at the BSC only using information related to the propagation losses of the channel instead of the complete channel matrix to reduce the signal overhead. Then, since the large-scale fading changes much more slowly than the coherence time, the cluster formation must only be updated a few times per second even in a high mobility scenario [20], thus leading to a low complexity design and a low signal overhead in the network. In fact, the most costly step of our proposal is the design of the precoder and decoder matrices that need to be updated at every coherence time. The computational complexity of the MAX-SINR algorithm is analyzed according to the number of complex multiplications as in [39]. For the direct channel, this complexity depends on the calculation of the intracluster interference plus noise covariance matrix in (9) and the computation of the decoder vector $\mathbf{U}_l^{[k]}$ according to (11). This last operation involves the matrix inversion of the covariance matrix. Then, similar operations have to be computed for the reciprocal channel. According to [39, 40], the multiplication of two complex matrices with sizes $N_r \times N_t$ and $N_t \times d$ requires $N_r N_t d$ complex multiplications. Therefore, the intracluster interference plus noise covariance matrix calculation at the kth AP/MS pair for all the data streams (i.e., $l = \{1, 2, .., d\}$) requires approximately a total of $\mathcal{S}_{c_k} \cdot (N_r N_t d + N_r^2 d + N_r^2)$ complex multiplications in the direct channel and $\mathcal{S}_{c_k} \cdot (N_r N_t d + N_t^2 d + N_t^2)$ in the reciprocal channel [39], where $\mathcal{S}_{c_k}$ is the number of SCs in the cluster containing

the kth AP/MS pair. The covariance matrix inversion per each data stream requires approximately N_r^3 complex multiplications for the direct channel and N_t^3 complex multiplications for the reciprocal channel. The exact complexity of the matrix inversion depends on the implemented algorithm [41]. Finally, computing the decoder vector $\mathbf{U}_l^{[k]}$ in the direct channel also requires the multiplication of two matrices, an Euclidean norm calculation of the resulting $N_r \times 1$ vector $\left(\mathbf{B}_{\mathbf{intra}}^{[k,l]}\right)^{-1}\mathbf{H}^{[k,k]}\mathbf{V}_l^{[k]}$ and the normalization by the Euclidean norm. These operations demand $N_rN_t + N_r^2$, N_r, and N_r complex multiplications, respectively. Consequently, the computational complexity to implement equations (11) in the direct channel and (14) in the reciprocal channel is approximately $N_r^3 + N_rN_t + N_r^2 + 2N_r$ and $N_t^3 + N_tN_r + N_t^2 + 2N_t$, respectively. These computations have to be repeated d times to account for each data stream. Therefore, the MAX-SINR algorithm demands in each iteration approximately $\mathcal{S}_{c_k} \cdot (2N_rN_td + N_r^2d + N_r^2 + N_t^2d + N_t^2) + d(N_r^3 + N_t^3 + N_r^2 + N_t^2 + 2N_rN_t + 2N_r + 2N_t)$ complex multiplications. Assuming the most complex scenario given by $N_r = N_t$ and the highest value of d computed as $d = (2N_t/\mathcal{S}_{c_k}) - 1$ [11], the complexity is approximately $8N_t^3 + 2N_t^2\mathcal{S}_{c_k} + (4N_t^4/\mathcal{S}_{c_k}) + (8N_t^3/\mathcal{S}_{c_k}) + (8N_t^2/\mathcal{S}_{c_k})$. Consequently, the highest complexity of the MAX-SINR algorithm in each iteration is given by $O(N_t^4)$. The complexity of the MAX-SINR-DL algorithm is very similar to that of MAX-SINR because the extra operations required to incorporate the effect of the intercluster interference as white noise demand a negligible complexity. These IA algorithms will be later compared in Section 5 with the SVD technique which is designed as in [26] with uniform power transmission per data stream. SVD is applied to every SC, and it only requires CSI corresponding to the selected AP/MS pair. Therefore, the signal overhead and the computational complexity are modest requiring only $14N_r^2N_t + 8N_r^3$ complex multiplications assuming that $N_r < N_t$ [39]. For the highest complexity scenario given by $N_r = N_t$, the computational complexity becomes $O(N_t^3)$ without requiring any iteration. However, this comes at the cost of totally neglecting interference and thus compromising the achieved performance.

5. Numerical Results

We initially consider a scenario consisting of $M = 50$ APs and $K = 30$ MSs uniformly distributed randomly within a square of area of $11\,km^2$ (i.e., the length of the side of the square is $L = 1000\,m$) as shown in Figure 1. Similar to [20], cell-edge effects are avoided by implementing a wrap-around technique in which the nominal squared coverage area under evaluation is wrapped around with eight neighboring identical squared areas to imitate an infinite network. We assume a network where each AP is equipped with $N_t = 8$ antennas and transmits d data streams to its corresponding MS which is equipped with $N_r = 2$ antennas. In line with a literature on ultradense deployments, each AP transmits with a power of $P = 200\,mW$ distributing available power uniformly among the d streams to transmit (i.e., uniform power allocation). Numerical results are obtained averaging over 100 random

locations of APs/MSs within the coverage area, and for each location, 100 channel realizations are obtained by generating different fast fadings. The rest of the simulation parameters are detailed in Table 1 and chosen in accordance with the numerology typically used in UDNs as reported in [20, 28].

Figure 2 illustrates the cluster design procedure for the same scenario shown in Figure 1. AP/MS pairs with the same colors are those located in the same cluster. According to ([11], (26)), for the above simulation parameters and assuming $d = 1$, the maximum number of users in each cluster is $S_{max} = 9$, a constraint that is shown to be achieved with the proposed size-restricted cluster design. Therefore, the fulfillment of the IA feasibility conditions is guaranteed.

Figure 3 shows the average SE per user versus the number of MSs for the traditional MAX-SINR, the MAX-SINR-DL, and the SVD beamforming. Results are shown for three distinct coverage areas ($L = 1000, 500,$ and $250\,m$). We consider the maximum number of data streams allowed for the SVD, that is, $d = 2$, while SE results for the IA algorithm have been obtained for $d = 1$ and $d = 2$. According to ([11], (26)), the maximum number of AP/MS pairs in each cluster is $S_{max} = 9$ for $d = 1$ and $S_{max} = 4$ for $d = 2$. As expected, it can be observed that, irrespective of the coverage size, the network performance decreases as the number of MSs increases. This result is caused by the fact that a higher value of K demands for a higher number of clusters to group all SCs. Therefore, higher intercluster interference levels are generated, thus limiting the performance of the network. In Figure 3(a) ($L = 1000\,m$), both MAX-SINR and MAX-SINR-DL achieve almost the same performance for low values of K. However, when the level of densification increases, MAX-SINR-DL slightly outperforms the traditional MAX-SINR. This effect can be more clearly observed in Figures 3(b) and 3(c) when the same number of APs/MSs are deployed in a smaller area. The performance of the traditional MAX-SINR is severely degraded by the intercluster interference for values of $K > 20$ in Figure 3(b) and $K > 15$ in Figure 3(c).

The amount of data streams multiplexed by the IA algorithm represents a trade-off between the amount of desired data transmitted in every SC and the number of users within the same cluster. For given values of N_t, N_r, and K, the higher the d, the lower the S_{max}, and hence, the higher the intercluster interference. In the scenarios shown in Figures 3(a) and 3(b), a higher performance is achieved by transmitting with more data streams. However, for an extreme densification case like the one shown in Figure 3(c), transmitting with a single data stream achieves a higher average SE than with $d = 2$. The reason behind this fact is that in highly interference-limited systems, designing larger clusters with the aim of improving the interference cancellation is more efficient than multiplexing a higher number of data streams. Note that the shrinkage of the coverage area unavoidably results in lower spectral efficiencies, irrespective of the technique in use. It is worth pointing out that under extreme densification (50 APs in a $250 \times 250\,m$ square) and heavy network load ($K > 40$), a reasonable condition since the amount of densification responds to the expected user load and the SVD-based beamforming performs remarkably well indicating that a useful

TABLE 1: Simulation parameters.

Parameters	Value
Carrier frequency	2 GHz
Bandwidth	20 MHz
Noise figure	9 dB
Antennas height	AP antenna height (h_{AP} = 15 m), MS antenna height (h_{MS} = 1.65 m)
Path loss parameters	$LOS\begin{cases} \zeta_0 = 34, \alpha = 2.2 & 10 < r^{[k,i]} < 150m \\ \zeta_0 = -5.17, \alpha = 4 & r^{[k,i]} > 150m \end{cases}$ $NLOS\{\zeta_0 = 30, \alpha = 3.67$
LOS/NLOS reference distance	r_0 = 20 m
Ricean K-factor	$\mu_K = 9, \sigma_K^2 = 5$
Shadowing parameters	$\sigma_{\chi_{LOS}}$ = 3dB, $\sigma_{\chi_{NLOS}}$ = 4dB, decorrelation distance (r_c = 9 m), correlation parameter (δ = 0.5)
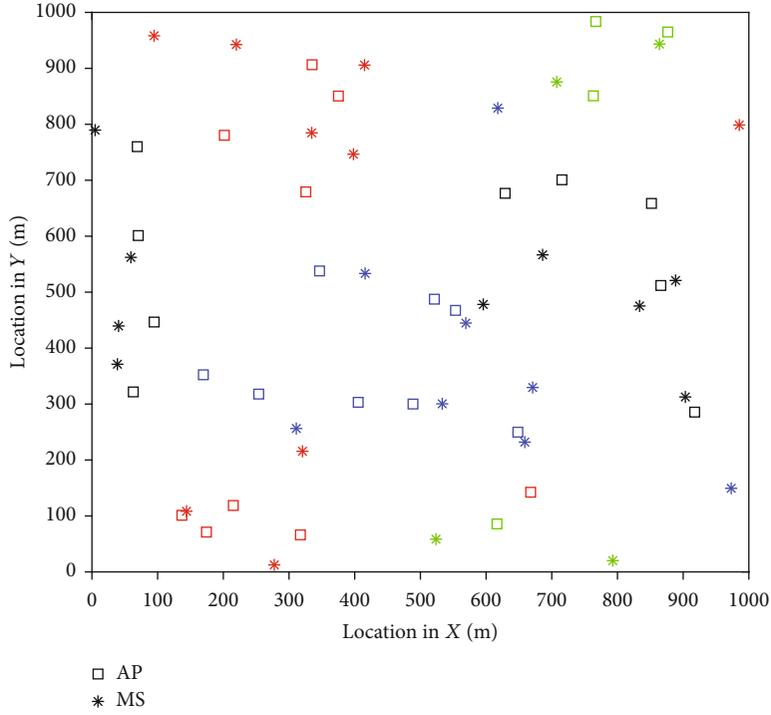
FIGURE 2: Illustration of the cluster design procedure for the dense deployment scenario composed of 50 APs and 30 MSs shown in Figure 1.

strategy in this case is to concentrate the radiated energy towards the direction of maximum gain of the user-specific channel matrix. Since SVD does not require any exchange of CSI information among APs, it seems it would be a reasonable processing choice under these specific conditions.

Overall, Figure 3 shows that MAX-SINR-DL is the IA-based scheme achieving the best performance in interference-limited systems since it deals with both intracluster and intercluster interference. The performance improvement is achieved by considering the information regarding the propagation losses from all users in the network. Since MAX-SINR-DL reports higher values of SE than traditional MAX-SINR in a clustered deployment of SCs, the remaining figures only illustrate the performance provided by this strategy.

Figure 4 illustrates the average SE per user versus the number of APs when considering a fixed number of MSs (K = 30). It can be observed that the performance increases as the number of APs grow while keeping the network load fixed. Note that the probability of having APs closer to each MS becomes higher for a larger amount of APs, leading to an increase of the probability of LOS propagation within an SC, a reduction of the corresponding path loss, and a more efficient exploitation of the diversity against shadow fading. Therefore, we can note that a network deployment with a high number of distributed APs increases the coverage probability at the cost of higher complexity and a more expensive design.

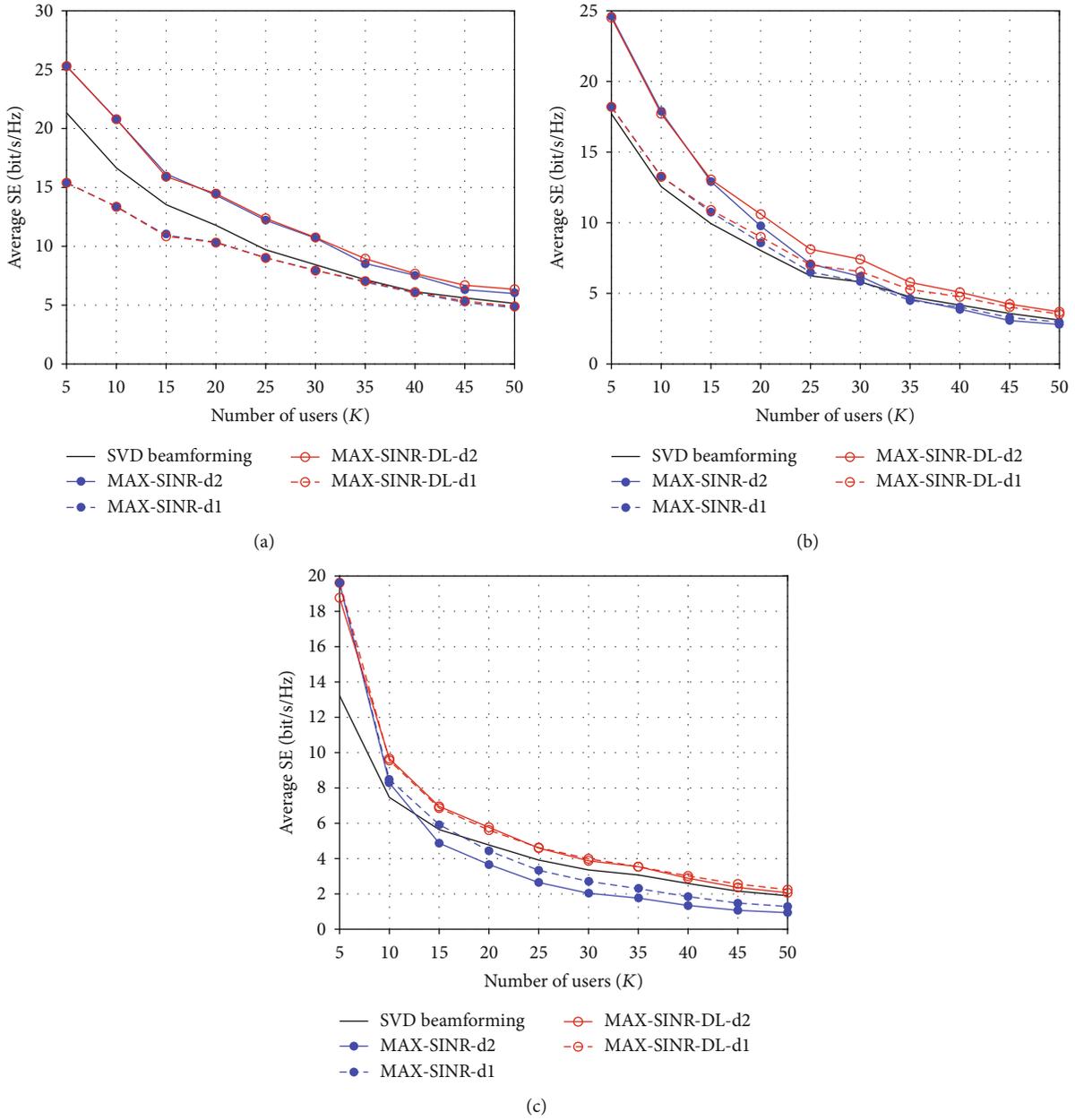A similar result is obtained in Figure 5, which shows a performance improvement as the APs are equipped with a

(a)

(b)

(c)

FIGURE 3: Average SE per user vs. K for traditional MAX-SINR, MAX-SINR-DL, and SVD by assuming $M = 50, N_t = 8, N_r = 2, d = 1$, and $d = 2$ for IA; $d = 2$ for SVD; and (a) $L = 1000$ m; (b) $L = 500$ m; (c) $L = 250$ m.

larger antenna array. Higher values of N_t imply the availability of a higher spatial dimension to align the interfering signals. Therefore, according to the analysis in Section 3, a larger number of AP/MS pairs could be grouped in the same cluster, thus reducing the intercluster interference. The performance curves obtained for the IA algorithm can be divided into two sections. For a single data stream transmission when $N_t < 29$, there are several clusters, and hence, the growth of SE is produced by the reduction of intercluster interference as the number of clusters decreases. For $N_t \geq 29$, there is only one cluster which means that all interfering terms are dealt with in the interference cancellation process and the SE is significantly increased. In fact, as N_t increases beyond 29

antennas, the SE keeps increasing but more slowly, because in this interval the growth is given by how well the interference is cancelled, and the desired signal is located in a subspace completely orthogonal to the interference subspace. In the case of $d = 2$, all SCs are grouped inside one single cluster for $N_t \geq 60$ which explains the abrupt change of the SE behaviour at $N_t = 60$.

Finally, Figure 6 shows the relation between the average SE per user and the length of the square side (L) when fixing the number of APs and MSs in the network ($M = 100, K = 30$). We notice that as a consequence of the realistic 3D channel model assumed in the paper, there is an interval where SE grows with L while in another one it decreases. This is because
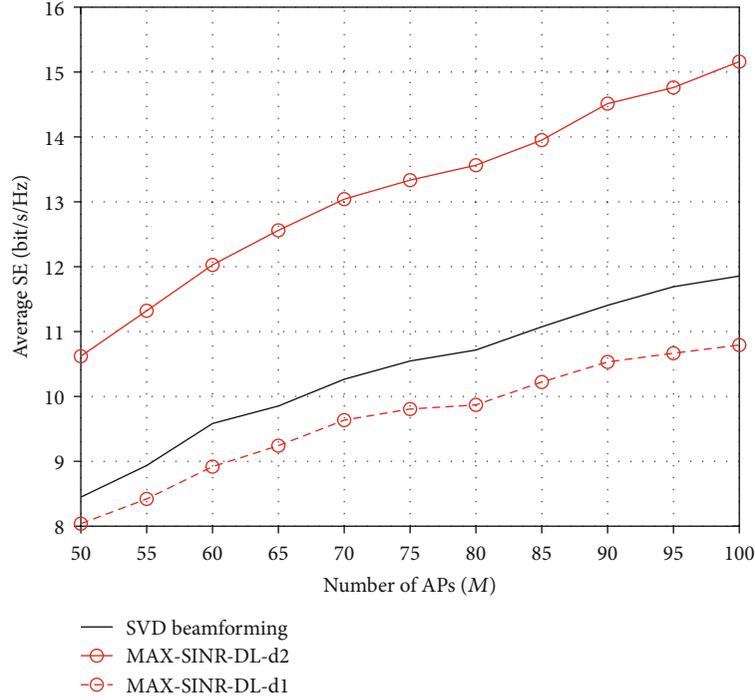
FIGURE 4: Average SE per user vs. M for $K = 30$, $N_t = 4$, $N_r = 2$, and $L = 1000$ m; $d = 1$ and $d = 2$ for IA; and $d = 2$ for SVD.
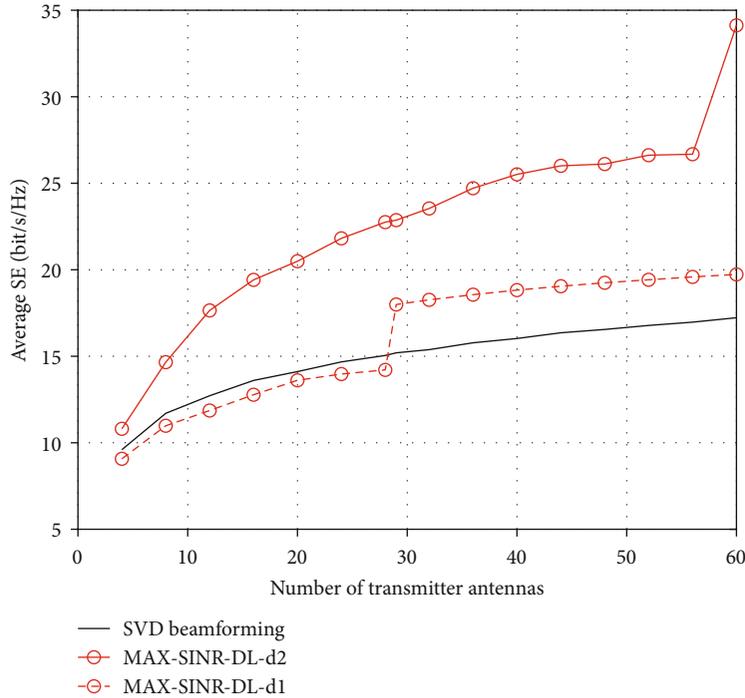

FIGURE 5: Average SE per user vs. N_t for $M = 100$, $K = 30$, $N_r = 2$, and $L = 1000$ m; $d = 1$ and $d = 2$ for IA; and $d = 2$ for SVD.

in dense scenarios limited by interference, as L increases the 3D distances between a given MS and the interfering APs increase faster than those between the MS and the selected AP. Then, since the large-scale fading depends on the distance, the power of the desired signal decreases more slowly than the interference power, and hence, the SE increases. As L keeps increasing and the spatial density of APs and MSs decreases,

however, the aforementioned 3D distances start behaving as 2D distances and the SE decreases with L as the noise term becomes more and more significant with respect to the interference term and the desired signal power is drastically reduced. Indeed this is also what causes the SVD-beamforming solution to perform remarkably close to the best IA solution. This figure proves that in the given values of K, M,
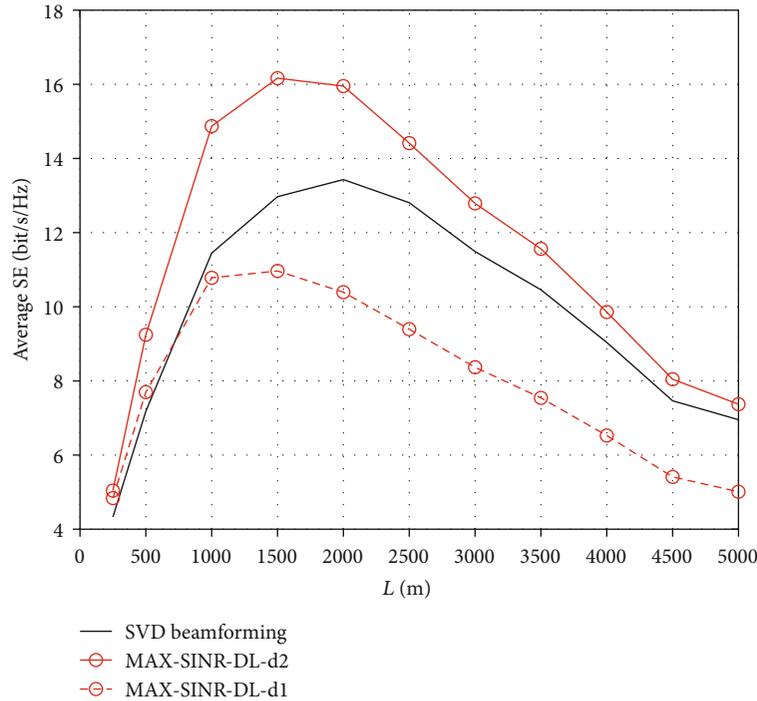
FIGURE 6: Average SE per user vs. L for $M = 100$, $K = 30$, $N_t = 4$, and $N_r = 2$; $d = 1$ and $d = 2$ for IA; and $d = 2$ for SVD.

N_t, N_r, h_{AP}, and h_{MS}, there is an optimal value of L for which the SE is maximum. Then, although it may seem surprising in an interference-limited system, distributing a certain amount of communication resources in a very small coverage may imply a loss in system performance.

6. Conclusion

This paper has considered the application of various flavours of IA in the context of UDN within a 5G/B5G framework. In particular, an algorithm based on a clusterized version of MAX-SINR IA has been shown to be an effective technique to achieve remarkably large spectral efficiencies in deployments of SCs with a large number of APs and MSs in the coverage area. Results reveal that a larger cluster size, effectively implemented by using more antennas at the APs, can drastically improve the network performance, but this comes at the cost of requiring the exchange of CSI among more network elements (all the APs/MSs forming the cluster). Also, the impact modifying the densities of APs and/or MSs might have on the system performance has been investigated and has served to illustrate the relationship between the number of network elements per area unit and performance. According to the considered propagation model, given a certain number of APs and MSs equipped with a fixed amount of antennas, there is an optimal coverage area value over which these resources should be distributed. Numerical results have provided us with valuable insight to make practical designs in UDNs showing a proper balance between performance and network complexity. A future work will seek to incorporate the effects the availability of imperfect CSI and/or the use of spatially correlated antenna arrays might have on the net-

work performance as well as the implementation of more sophisticated power allocation techniques.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Ahmadi, *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*, Academic Press, 2019.

[2] X. Wang, L. Kong, F. Kong et al., "Millimeter wave communication: a comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1616–1653, 2018.

[3] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[4] S. A. Jafar, "Interference alignment — a new look at signal dimensions in a communication network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, pp. 1–134, 2010.

[5] N. Zhao, F. R. Yu, M. Jin, Q. Yan, and V. C. M. Leung, "Interference alignment and its applications: a survey, research issues, and challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1779–1803, 2016.

[6] X.-Y. Li, C. He, H.-S. Shan, J. Li, and Z. Jane Wang, "Feasibility-aware partial interference alignment for hybrid D2D and cellular communication networks," *IEEE Access*, vol. 6, pp. 71069–71083, 2018.

[7] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3309–3322, 2011.

[8] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3480–3489, 2010.

[9] O. González, C. Lameiro, and I. Santamaría, "A quadratically convergent method for interference alignment in MIMO interference channels," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1423–1427, 2014.

[10] D. A. U. Villalonga and M. J. Fernández-Getino García, "Performance comparison of interference alignment algorithms in an energy harvesting scenario," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 1–6, Porto, Portugal, July 2020.

[11] O. González, C. Beltrán, and I. Santamaría, "A feasibility test for linear interference alignment in MIMO channels with constant coefficients," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1840–1856, 2014.

[12] R. Tresch and M. Guillaud, "Clustered interference alignment in large cellular networks," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1024–1028, Tokyo, Japan, September 2009.

[13] R. Tresch and M. Guillaud, "Performance of interference alignment in clustered wireless ad hoc networks," in *2010 IEEE International Symposium on Information Theory*, pp. 1703–1707, Austin, TX, USA, June 2010.

[14] S. Chen and R. S. Cheng, "Clustering for interference alignment in multiuser interference network," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2613–2624, 2014.

[15] F. Zhao, P. Zhu, J. Li, and X. You, "Hedonic coalition formation game for clustered interference alignment," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–5, Nanjing, China, October 2017.

[16] H. Zhang, K. Yang, S. Zhang, and X. Wang, "Efficient subchannel allocation based on partial clustered interference alignment and clusters merging in dense small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16191–16196, 2020.

[17] H. Zhang, K. Yang, S. Zhang, and O. A. Dobre, "Subchannel allocation based on clustered interference alignment for differentiated data streams in dense small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14049–14054, 2020.

[18] H. E. Elkotby, K. M. F. Elsayed, and M. H. Ismail, "Exploiting interference alignment for sum rate enhancement in D2D-enabled cellular networks," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1624–1629, Paris, France, April 2012.

[19] H. E. Elkotby, K. M. F. Elsayed, and M. H. Ismail, "Shrinking the reuse distance: spectrally-efficient radio resource management in D2D-enabled cellular networks with interference alignment," in *2012 IFIP Wireless Days*, pp. 1–6, Dublin, Ireland, November 2012.

[20] 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.

[21] S. Elhoushy and W. Hamouda, "Performance of distributed massive MIMO and small-cell systems under hardware and channel impairments," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8627–8642, 2020.

[22] S. Elhoushy and W. Hamouda, "Towards high data rates in dynamic environments using hybrid cell-free massive MIMO/small-cell system," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 201–205, 2021.

[23] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, 3rd ed*, Morgan Kaufmann Publishers, 2011.

[24] K. P. Bennett, P. S. Bradley, and A. Demiriz, "Constrained K-means clustering," 2000, https://www.microsoft.com/en-us/research/publication/constrained-k-means-clustering/.

[25] M. I. Malinen and P. Fränti, "Balanced K-means for clustering," in *Structural, Syntactic, and Statistical Pattern Recognition, ser. Lecture Notes in Computer Science*, P. Fränti, G. Brown, M. Loog, F. Escolano, and M. Pelillo, Eds., pp. 32–41, Springer, Berlin, Heidelberg, 2014.

[26] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[27] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44596–44612, 2019.

[28] 3GPP, "Study on 3D channel model for LTE (release 12)," 2017, 3GPP TR 36.873 (version 12.7.0).

[29] A. A. Polegre, F. Riera-Palou, G. Femenias, and A. G. Armada, "Channel hardening in cell-free and user-centric massive MIMO networks with spatially correlated Ricean fading," *IEEE Access*, vol. 8, pp. 139827–139845, 2020.

[30] F. Riera-Palou, G. Femenias, A. G. Armada, and A. Pérez-Neira, "Clustered cell-free massive MIMO," in *2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Abu Dhabi, United Arab Emirates, December 2018.

[31] T. L. Marzetta, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[32] B. Sklar, "Rayleigh fading channels in mobile digital communication systems .I. Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.

[33] C. M. Yetis, Tiangao Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4771–4782, 2010.

[34] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 511–515, San Francisco, CA, USA, June 2011.

[35] Chenwei Wang, Tiangao Gou, and S. A. Jafar, "Subspace alignment chains and the degrees of freedom of the three-user MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2432–2479, 2014.

[36] Z. Zong, H. Feng, F. R. Yu, N. Zhao, T. Yang, and B. Hu, "Optimal transceiver design for SWIPT in K-user MIMO interference channels," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 430–445, 2016.

[37] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, ser. SODA '07*, pp. 1027–1035, USA: Society for Industrial and Applied Mathematics, January 2007.

[38] O. Bachem, M. Lucic, S. H. Hassani, and A. Krause, "Approximate K-means++ in sublinear time," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, p. 1, 2016, https://ojs.aaai.org/index.php/AAAI/article/view/10259.

[39] C. Sun, Y. Yang, and Y. Yuan, "Low complexity interference alignment algorithms for desired signal power maximization problem of MIMO channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, 2012.

[40] R. Hunger, *Floating Point Operations in Matrix-Vector Calculus*, Munich University of Technology, Inst. for Circuit Theory and Signal Processing, 2005.

[41] D. Yu, Z. Lu, S. He, B. Wu, Y. Huang, and L. Yang, "Low complexity complex matrix inversion method for MIMO communication systems," in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, pp. 1–5, Nanjing, China, October 2015.