Aalto University

Ribeiro, Lucas; Leinonen, Markus; Al-Tous, Hanan; Tirkkonen, Olav; Juntti, Markku

Channel Charting Aided Pilot Reuse for Massive MIMO Systems with Spatially Correlated Channels

Channel Charting Aided Pilot Reuse for Massive MIMO Systems With Spatially Correlated Channels

LUCAS RIBEIRO[1] **(Graduate Student Member, IEEE), MARKUS LEINONEN**[1] **(Member, IEEE),**
HANAN AL-TOUS[2] **(Senior Member, IEEE), OLAV TIRKKONEN**[2] **(Fellow, IEEE),**
AND MARKKU JUNTTI[1] **(Fellow, IEEE)**

[1]Centre for Wireless Communications, University of Oulu, Oulu 90014, Finland

[2]Department of Communications and Networking, Aalto University, 00076 Espoo, Finland

CORRESPONDING AUTHOR: L. RIBEIRO (e-mail: lucas.ribeiro@oulu.fi)

ABSTRACT Massive multiple-input multiple-output (mMIMO) technology is a way to increase spectral efficiency and provide access to the Internet of Things (IoT) and machine-type communication (MTC) devices. To exploit the benefits of large antenna arrays, accurate channel estimation through pilot signals is needed. Massive IoT and MTC systems cannot avoid pilot reuse because of the enormous numbers of connected devices. We propose a pilot reuse algorithm based on channel charting (CC) to mitigate pilot contamination in a multi-sector single-cell mMIMO system having spatially correlated channels. We show that after creating an interference map via CC, a simple strategy to allocate the pilot sequences can be implemented. The simulation results show that the CC-based pilot reuse strategy improves channel estimation accuracy, which subsequently improves the symbol detection performance and increases the spectral efficiency compared to other existing schemes. Moreover, the performance of the CC pilot assignment method approaches that of exhaustive search pilot assignment for small network setups.

INDEX TERMS Channel charting, massive MIMO, pilot reuse, pilot contamination.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (mMIMO) technology is a key enabler for the Internet of Things (IoT) and massive machine-type communication (mMTC) systems [1], [2], [3], [4], [5]. It supports the access of huge numbers of connected devices through the spatial multiplexing of the user equipments (UEs). Furthermore, the effects of small-scale fading and uncorrelated noise have been shown to asymptotically vanish in mMIMO, significantly improving spectral and energy efficiencies for such systems [6], [7], [8]. However, to exploit spatial multiplexing and the benefits of channel hardening, accurate channel estimation through pilot signals is needed.

IoT and mMTC systems are characterized by sporadic UE activity, uplink-dominated transmissions, and small packets [3], [5], [9]. The design of access techniques to support the huge numbers of such devices is still an open challenge [10]. As opposed to conventional networks connecting mostly human users, the large number of UEs in mMIMO systems precludes the allocation of orthogonal pilot sequences to all of them [5], [11]. Different approaches such as the intelligent reuse of orthogonal pilot sequences and the use of non-orthogonal pilot sequences have been proposed to cope with the shortcoming of orthogonal resources. The first approach is the subject of this work. A related pilot sequence allocation problem appears also in non-orthogonal multiple-access (NOMA) systems [12] for which our approach can bring new insights as well.

Pilot reuse can be employed within the same cell to acquire the required channel state information (CSI) and

avoid excessive signaling, overcoming the lack of resources in mMIMO systems at the cost of *pilot contamination* known also as *pilot collision*. The contamination or collision of pilot sequences is the interference caused by reusing the same pilot sequence across different UEs to estimate the uplink channels, and it typically affects neighboring cells of a cellular network [13]. Particularly, when the number of antennas tends towards infinity, the ultimate limiting phenomenon affecting the performance of mMIMO systems is the pilot contamination [6]. To tackle the pilot contamination and improve both spectral and energy efficiency, several approaches exploiting the propagation environment through spatial multiplexing have been proposed [14], [15].

A. RELATED WORK

To address the shortage of resources and mitigate the pilot contamination problem in multi-user mMIMO systems, superimposed pilots have been studied in [16], [17]. Instead of time-division multiplexing of pilot symbols to estimate the channels and data, pilots are sent to the base station (BS) along with the data, not in dedicated time slots. In [17], Lago et al. proposed a mixed approach where first a set of orthogonal pilot sequences is sent, to later improve the accuracy of the channel estimation, and then superimposed pilot symbols are sent to estimate the uplink channels. The results showed that this approach can improve spectral efficiency as compared to purely superimposed or orthogonal pilot schemes.

The design of pilot sequences is considered in [18] to mitigate pilot contamination. Van Chien et al. proposed a framework where the pilots are linear combinations of orthogonal pilot sequences acting as basis vectors, and the resulting non-combinatorial optimization problem aims to find suitable weight coefficients for the basis vectors. Xu et al. [19] utilized deep learning to find such weight coefficients. They formulated a pilot power loading problem to minimize the sum mean square error (MSE) of channel estimation. They have shown that the proposed method improves the sum MSE as compared to other existing methods by employing a deep neural network to learn the MSE-minimizing optimal power allocation to each pilot symbol.

In addition to the aforementioned approaches, the pilot contamination can be alleviated by exploiting the spatial correlation of channels, as presented in [20]. Muppirisetty et al. [21] proposed a location-aided pilot allocation algorithm, which assumes the knowledge of UEs' physical locations and distribution of the scatterers to estimate the covariance matrices and, subsequently, to suppress pilot contamination. They consider an integer linear problem to assign the pilots. Their results show that the normalized channel estimation error approaches the interference-free scenario. A similar approach, yet using angle-of-arrival (AoA) information for allocating the pilot sequences was proposed in [22]. Therein, first the UEs are grouped within the cells to alleviate intra-cell pilot contamination. UEs

in the same group cannot use the same pilot sequence. Next, an algorithm matches the groups with strong potential interference in the neighboring cells into a group collection. Then, a pilot allocation algorithm assigns orthogonal pilot sequences to UEs with strong inter-cell interference. The potential mutual interference is measured on an undirected weighted graph, with vertices representing the UEs in each collection and the edges representing the interference. The main practical limitation of such location-aware pilot assignment methods is the requirement of knowledge about UEs' physical locations.

Another class of solutions, which do not require the explicit knowledge of UEs' positions, utilizing the spatial correlation to avoid pilot contamination includes [23], [24], [25], [26]. In [26], a pilot reuse strategy called statistic greedy pilot scheduling (SGPS), based on the second-order statistics of the channels, was proposed aiming at minimizing the MSE for both channel estimation and signal detection. The SGPS algorithm seeks to assign orthogonal pilot sequences to UEs with similar channel covariance matrices, according to a metric based on the covariance matrix distance (CMD) [27], resulting in significant improvement in terms of spectral efficiency as compared to the orthogonal pilot allocation without reuse.

Nguyen et al. [28] proposed to maximize the minimum weighted sum spectral efficiency by applying a pilot allocation algorithm and a transmission power control policy, aiming to diminish pilot contamination and to mitigate interference for the data transmission, respectively. In [29], Ma et al. utilize sparse Bayesian learning to estimate the channels' spatial signatures. Then, in the channel training phase, UEs with non-overlapping spatial signatures are grouped and assigned with the same pilot sequence, mitigating the overall pilot contamination.

Recently, we proposed a pilot reuse scheme in [30], [31] that utilizes *channel charting* (CC)[1] to exploit the spatial information existing in the CSI to allocate and reuse orthogonal pilot sequences so as to alleviate pilot contamination. CC is a framework proposed in [32] to estimate the relative positions of devices in an unsupervised manner. The CC maps the information obtained from the measured long-term CSI at the BS into a low-dimensional chart, in which the relative positions of UEs are preserved. The multi-cell extension of [30], [31] is considered in [33]. Our results [30], [31], [33] show that CC can be utilized to mitigate pilot contamination and improve the channel estimation accuracy.

B. CONTRIBUTIONS

In this paper, we address the problem of pilot assignment in a single-cell multi-sector mMIMO system with spatially correlated channels by employing pilot reuse of a set of orthogonal sequences. The interference in mMIMO systems is often related to the spatial configuration of UEs, especially

---

1. CC is interchangeably used for "channel charting" and "channel chart" throughout the paper.

to their angular position in relation to the BS. Therefore, a description of the whole azimuth plane is very important to characterize the interference among UEs. As directional antennas are usually employed in cellular systems, a multi-sector deployment is needed to provide coverage to the whole angular domain. Furthermore, the multi-sector arrangement allows us to get rid of the mirror angle phenomenon presented in uniform linear arrays (ULAs) [34, p. 240] or uniform planar arrays (UPAs).

We formulate a pilot allocation optimization problem that aims at minimizing pilot contamination by exploiting the spatial correlation of the UEs' channels through their second-order statistics. The formulated problem is, in general, hard to solve due to its combinatorial nature, which makes the complexity grow exponentially with the number of UEs and linearly with the number of available pilot sequences. To this end, we propose a practical solution based on CC to create a map that reveals the amount of interference between the UEs by mapping those with strong channel correlation closer than those with weak one. We show through simulation experiments that a simple pilot allocation strategy can be employed if an appropriate interference map is designed. Numerical results show that the intelligent CC-based pilot reuse method mitigates pilot contamination and greatly increases spectral efficiency. In a toy example, for a small network setup with few UEs deployed, the proposed method approaches the performance of exhaustive search.

The main **contributions** of this paper are summarized as follows:

- We formulate a pilot allocation problem to optimally reuse a pool of orthogonal pilots in single-cell multi-sector mMIMO systems with spatially correlated channels.
- We propose a CC method that translates the amount of interference between the UEs into a low-dimensional chart, creating an interference mapping. The higher the potential interference between two UEs, the closer they reside on the map.
- We propose to improve the interference map generated by CC by allowing a variable number of dimensions for CC. The extra degrees of freedom in CC allows us to construct more reliable interference mappings.
- We develop a CC-based greedy pilot allocation algorithm that exploits the spatial correlation of the UEs, embedded in the interference mapping. The pilot assignment algorithm aims at maximizing the distances between the UEs on the interference map that use the same pilot sequence to mitigate pilot contamination.
- Simulation results show that the proposed pilot allocation can cope with pilot contamination, i.e., it increases channel estimation accuracy and uplink achievable rate, or equivalently, reduces the symbol error rate.

The use of second-order statistical information to allocate the pilot sequences or to retrieve UEs' spatial information

is not a completely new idea. In [26], the covariance matrices were fed to a pilot allocation algorithm that computes a CMD-like feature and greedily assigns orthogonal pilot sequences, trying to minimize the interference with the set of UEs previously addressed. However, the novelty in our work lies in the design of an UE-interference map of the radio environment through CC. The UE-interference map reveals the big picture of the interference in the cell, which allows employing a low-complexity pilot allocation strategy to mitigate pilot contamination.

The present paper builds on our initial works [30], [31]. Herein, we take a more elaborate approach to the studied pilot contamination problem. We start by deriving an objective function to assess the overall amount of pilot contamination in the network, which works as a proxy function for the end performance metrics. Using this objective function, we approach the pilot contamination problem and the associated pilot reuse strategy by formulating a rigorous optimization problem that aims to minimize the amount of pilot contamination. This optimization problem is subsequently used as a basis to utilize meta-heuristics and propose the CC-based pilot reuse algorithm. Furthermore, to increase robustness of the proposed method to the changes in the UEs' density, here we propose the use of an adaptive number of dimensions for CC.

C. PAPER OUTLINE

The rest of the paper is organized as follows. Section II presents the system model for the multi-sector single-cell communication system. The channel estimation and the problem formulation are presented in Section III. The CC framework to retrieve a radio interference map is proposed in Section IV. In Section V, the proposed pilot allocation strategy utilizing the CC-based interference mapping is devised. The simulation results are presented and discussed in Section VI, and Section VII concludes the paper.

Notation: Boldface lowercase letters, $\mathbf{x}$, denote column vectors and boldface uppercase letters, $\mathbf{X}$, denote matrices. The superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, and $(\cdot)^{\mathrm{H}}$ denote transpose, conjugate, and conjugate transpose, respectively. The average value of $\mathbf{x}$ is $\bar{\mathbf{x}}$. The multivariate circularly symmetric complex Gaussian distribution with covariance matrix $\mathbf{R}$ and mean $\mathbf{x}$ is denoted as $\mathcal{CN}(\mathbf{x}, \mathbf{R})$. The $n \times n$ identity matrix is presented as $\mathbf{I}_n$. The vector of all entries 1 is denoted as $\mathbf{1}$. The element-wise product between $\mathbf{A}$ and $\mathbf{B}$ is expressed as $\mathbf{A} \odot \mathbf{B}$. The expected value of $\mathbf{x}$ is denoted as $\mathbb{E}[\mathbf{x}]$. The ℓ_2-norm and the Frobenius norm are denoted as $\| \cdot \|_2$ and $\| \cdot \|_F$, respectively. The trace of $\mathbf{X}$ is denoted as $\mathrm{tr}(\mathbf{X})$.

II. SYSTEM MODEL

We consider a mMIMO uplink communication scheme with a set $\mathcal{N} = \{1, \ldots, N\}$ of N single-antenna UEs uniformly distributed within a cell, from which only $K < N$ are randomly active at any given time. In order to focus on the development of pilot assignment strategies to improve the channel estimation accuracy, and, subsequently, enhance the
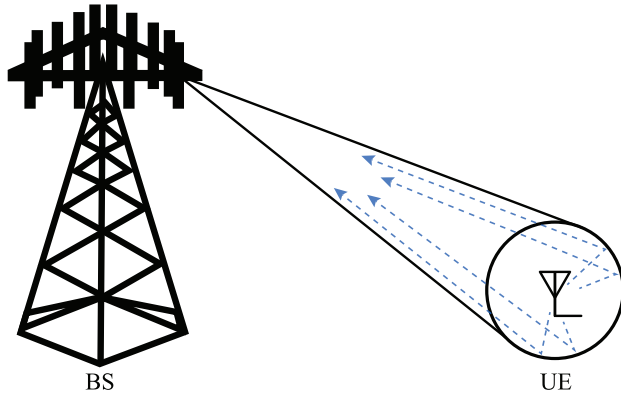
FIGURE 1. The uplink channel between a single-antenna UE and a BS equipped with $S = 3$ M-element ULAs. The multi-path components are concentrated in a ring around the UEs.

system performance, we assume that the BS knows the set of active UEs at each transmission instant. This assumption can be invoked by the fact that there exists a multitude of compressed sensing based approaches, e.g., [5], [35], [36], to detect the set of active UEs, and thus, we leave it outside of the scope of this paper.

The cell is divided into S sectors.[2] The BS receives the UEs' transmitted signals through S ULAs[3] each having M antenna elements. The array response vector for a ULA is given by

$$\mathbf{a}_r(\theta) = \left[1, e^{-j2\pi\Delta_r \cos(\theta)}, \ldots, e^{-j2\pi(M-1)\Delta_r \cos(\theta)}\right]^T, \quad (1)$$

where Δ_r is the normalized spacing between the antenna elements in units of wavelengths and θ is the AoA, i.e., the incident angle of the received signal on the antenna array [39, Sec. 7.2.1]. The ULA geometry causes resonance at certain angle pairs, a phenomenon known as mirror angles [34, p. 258]. Multi-sector processing allows the proposed CC-based method to handle the ULAs' mirror angles effect by combining the signal from different arrays, thus, granting coverage to UEs at the whole extent of the angular domain, $[0, 2\pi]$.

We adopt the one-ring channel model that assumes the multi-path components are concentrated around the UEs while the BS is in an elevated position, thus lacking scatterers in its proximity. Each multipath reaches the antenna array from a particular angle within a ring close to the UE, which makes the channels spatially correlated with independent gains and phase rotations [34, p. 236]. Fig. 1 depicts the channel model between a UE and the BS. Accordingly, the uplink channel vector for user $n \in \mathcal{N}$ at the ULA of sector $s \in \mathcal{S} = \{1, \ldots, S\}$ is modelled as a superposition of

<hr>

2. In practical systems, the outdoors coverage area is usually divided into three or six sectors of 120° or 60° [37]. We consider $S = 3$ in the examples.
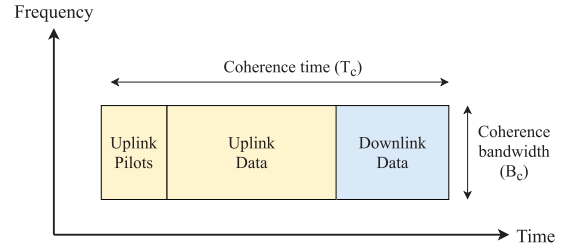
3. The extension to UPAs is straightforward [38].

Frequency

FIGURE 2. Illustration of transmission phases in the considered TDD system. At the beginning of each coherence block, τ symbols are sent to estimate the channel.

L propagation paths as

$$\mathbf{h}_{n,s} = \frac{1}{\sqrt{L}}\sum_{l=1}^{L}\sqrt{\beta_{n,s,l}}\alpha_{n,s,l}\mathbf{a}_r(\theta_{n,s,l}), \quad (2)$$

where $\alpha_{n,s,l}$ is the complex gain of the lth path, which is assumed to be an independent and identically distributed (i.i.d.) complex Gaussian random variable with zero mean and $\mathbb{E}[|\alpha_{n,s,l}|^2] = 1$. The large-scale propagation effects and the BS antenna gain in the channel are captured in $\beta_{n,s,l} \in \mathbb{R}$. Any path loss model can be used to model $\beta_{n,s,l}$, here, we consider the free-space path loss model as described in [40, eq. (2.7)], which is defined as

$$\beta_{n,s,l} \triangleq 10^{\frac{G_A(\theta_{n,s,l})}{10}}\left(\frac{\lambda}{4\pi d_n}\right)^2, \quad (3)$$

where λ is the wavelength and d_n is the distance between UE n and the BS. In (3), $G_A(\theta_{n,s,l}) \in \mathbb{R}$ represents the antenna gain through the lth path for user n at the ULA of sector s, and is given in dB by [37], [41, Table 7.1-1]

$$G_A(\theta_{n,s,l}) = G_{A_{max}} - \min\left[12\left(\frac{\theta_{n,s,l}}{\theta_{3dB}}\right)^2, A_{max}\right], \quad (4)$$

where $G_{A_{max}}$ is the maximum antenna gain, θ_{3dB} is the half power beamwidth, A_{max} is the maximum attenuation in dB at the ULAs. Above, $\theta_{n,s,l}$ is the AoA of the lth path, $l = 1, \ldots, L$. The AoA for the L paths between UE n and ULA s can be modelled as i.i.d. random variables with uniform distribution $U(\theta_{n,s}^{min}, \theta_{n,s}^{max})$, with $\theta_{n,s}^{min} = \bar{\theta}_{n,s} - \sqrt{3}\sigma_\theta$ and $\theta_{n,s}^{max} = \bar{\theta}_{n,s} + \sqrt{3}\sigma_\theta$. Here, $\bar{\theta}_{n,s} \in [0, 2\pi]$ is the incident angle between UE n and the ULA of sector s, σ_θ is the angular standard deviation, which specifies the AoA interval, $\mathcal{A}_{n,s} = [\theta_{n,s}^{min}, \theta_{n,s}^{max}]$, for the possible incoming multi-path components arriving from UE n at the ULA of sector s.

We consider a time-division duplex (TDD) system, as depicted in Fig. 2. We assume that during one coherence block, the channels are time-invariant and flat fading. For TDD systems, downlink and uplink channels are symmetric, i.e., the downlink channel can be inferred at the BS from the uplink one [39]. At each coherence block, the set of active UEs transmit τ known symbols to the BS for channel estimation. Right after transmitting the pilot symbols, the K active UEs transmit their data to the BS. Then, the remaining time within the coherence block is used for downlink communication.

III. CHANNEL ESTIMATION AND PROBLEM FORMULATION

The UEs are assigned with pilot sequences taken from a pool of τ *orthogonal* sequences. However, due to a vast number of UEs in mMIMO networks, we assume that $\tau \ll N$, so that global *pilot reuse* is employed in the cell area. Consequently, in the channel estimation phase, the same pilot is shared on the average by K/τ UEs. Let $\mathcal{K} = \{1, \ldots, K\} \subseteq \mathcal{N}$ represent the set of active UEs at a given transmission interval. Let $\mathcal{T} = \{1, \ldots, \tau\}$ be the set of indices of the available pilot sequences. UE $k \in \mathcal{K}$ transmits a pilot signal $\psi_k = \sqrt{p_\mathrm{u}}\phi_{\pi_k}$, where p_u is the power of each pilot symbol, $\pi_k \in \mathcal{T}$ is the index of the pilot sequence assigned to UE k, and $\phi_{\pi_k} \in \mathbb{C}^\tau$ is the corresponding pilot sequence, with power τ, taken from the orthogonal pilot book $\boldsymbol{\Phi} = [\phi_1, \ldots, \phi_\tau] \in \mathbb{C}^{\tau \times \tau}$. We define the set of interfering UEs to UE k as $\mathcal{I}_k$, i.e., $\mathcal{I}_k = \{j \mid j \in \mathcal{K} \setminus \{k\}, \pi_j = \pi_k\}$.

A. CHANNEL ESTIMATION

The compound channel $\mathbf{h}_k \in \mathbb{C}^{MS}$ between UE $k \in \mathcal{K}$ and the BS is given by

$$\mathbf{h}_k = \begin{bmatrix} \mathbf{h}_{k,1} \\ \vdots \\ \mathbf{h}_{k,S} \end{bmatrix}, \tag{5}$$

with covariance matrix $\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^\mathrm{H}] \in \mathbb{C}^{MS \times MS}$. The linear minimum mean square error (LMMSE) estimator is deployed at the BS to jointly estimate the channel vectors of the active UEs across the S sectors. Here, because the channels are Gaussian distributed, the MMSE estimator reduces to the LMMSE estimator [42, Sec. 12.3]. Importantly, Yin et al. [43] showed that the MMSE estimator can recover the channel estimates as in an interference-free scenario given that the covariances' signal subspaces are orthogonal. In our framework, this orthogonality raises from the restricted angles of the multipaths impinging on the receive antenna from each UE according to the one-ring channel model.

We assume that the channel covariance matrices of all UEs (active and inactive ones), i.e., $\mathbf{R}_n = \mathbb{E}[\mathbf{h}_n \mathbf{h}_n^\mathrm{H}]$, $n \in \mathcal{N}$, are known at the BS. This may be a challenging assumption for medium/high mobility UEs but is still feasible [20]. In practice, an initial training phase is required to obtain the first estimates of the covariance matrices. Several techniques to obtain the covariance matrices are explored in [20] and in the references therein. For instance, the BS can keep updating each $\mathbf{R}_n$ based on the estimated channel, i.e., the BS can approximate the covariance matrices by their sampled versions. Although a low-mobility scenario is readily applicable for the proposed method (e.g., electricity and water metering, environmental monitoring, building automation, city lightning, etc.), it is not stringently restricted to this scenario. Namely, the only requirement is that accurate channel statistics are available, meaning that for medium/high mobility, the statistics need to be updated more frequently.

The received signal for channel estimation at the BS, $\mathbf{Y} \in \mathbb{C}^{MS \times \tau}$, can be written as

$$\mathbf{Y} = \mathbf{H}\check{\boldsymbol{\Psi}} + \mathbf{N}, \tag{6}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{MS \times K}$ is the channel matrix for the active UEs, $\check{\boldsymbol{\Psi}} = [\psi_1, \ldots, \psi_K]^\mathrm{T} \in \mathbb{C}^{K \times \tau}$ is the pilot signal matrix, and $\mathbf{N} = [\mathbf{n}^1, \ldots, \mathbf{n}^\tau] \in \mathbb{C}^{MS \times \tau}$ is the noise matrix. We model the noise as an i.i.d. complex Gaussian random variable $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_\mathrm{n}^2 \mathbf{I}_{MS})$, where σ_n^2 is the noise power at each antenna element.

The LMMSE estimate of the channel between UE k and the BS, $\mathbf{h}_k$ in (5), is given as [26]

$$\hat{\mathbf{h}}_k = \mathbf{R}_k \mathbf{Q}_k^{-1} \mathbf{y}_k^\mathrm{p}, \tag{7}$$

where $\mathbf{y}_k^\mathrm{p} \in \mathbb{C}^{MS}$ represents the processed received signal for UE k after correlating the received signal with the pilot sequence assigned to UE k, i.e.,

$$\begin{aligned} \mathbf{y}_k^\mathrm{p} &= \frac{1}{p_\mathrm{u}\tau} \mathbf{Y}\psi_k^* , \\ &= \mathbf{h}_k + \underbrace{\sum_{j \in \mathcal{I}_k} \mathbf{h}_j}_{\text{Pilot Interference}} + \frac{1}{p_\mathrm{u}\tau} \mathbf{N}\psi_k^* , \end{aligned} \tag{8}$$

and $\mathbf{Q}_k = \mathbb{E}[\mathbf{y}_k^\mathrm{p}(\mathbf{y}_k^\mathrm{p})^\mathrm{H}] \in \mathbb{C}^{MS \times MS}$ is the covariance matrix of the processed received signal, given as

$$\mathbf{Q}_k = \mathbf{R}_k + \sum_{j \in \mathcal{I}_k} \mathbf{R}_j + \frac{\sigma_\mathrm{n}^2}{p_\mathrm{u}\tau} \mathbf{I}_{MS}. \tag{9}$$

Due to the orthogonality principle of the MMSE estimator [42, Sec. 12.4], the channel estimation error is independent of $\hat{\mathbf{h}}_k$ and given by $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k \in \mathbb{C}^{MS}$. Thus, the error covariance matrix for user k is given by [26]

$$\mathbf{R}_{\tilde{\mathbf{h}}_k} = \mathbf{R}_k - \mathbf{R}_k \mathbf{Q}_k^{-1} \mathbf{R}_k, \tag{10}$$

and the corresponding MSE by

$$\begin{aligned} \mathrm{MSE}_k &= \mathrm{tr}\left(\mathbf{R}_{\tilde{\mathbf{h}}_k}\right) \\ &= \mathrm{tr}\left(\mathbf{R}_k - \mathbf{R}_k\left(\mathbf{R}_k + \sum_{j \in \mathcal{I}_k} \mathbf{R}_j + \frac{\sigma_\mathrm{n}^2}{p_\mathrm{u}\tau} \mathbf{I}_{MS}\right)^{-1} \mathbf{R}_k\right). \end{aligned} \tag{11}$$

To obtain an interference-free estimate of $\mathbf{h}_k$, a vector basis for the set of matrices $\{\mathbf{R}_j | j \in \mathcal{I}_k\}$ must be orthonormal to a basis vector for $\mathbf{R}_k$. As proven in [43], this condition is fulfilled, for an asymptotic regime ($M \to \infty$), if the multipath components of the interfering UEs lay outside the AoA interval for UE k, $\mathcal{A}_{k,s}$ in all sectors s. In that case, the interfering signals will fall in the null space of $\mathbf{R}_k$, and the MSE is going to be purely noise-limited. Therefore, the AoA intervals – or equivalently the eigenstructure of the channel covariance matrices – will determine the impact of the pilot contamination with large-antenna arrays [26], [43].

B. PROBLEM FORMULATION

The massive number of devices in mMIMO systems prevents the allocation of unique orthogonal pilot sequences to all UEs, which necessitates pilot reuse. Pilot contamination affects the capability to mitigate interference during the data transmission phase due to the reduced estimation quality and the statistical dependence of the channel estimates [34, p. 252]. Furthermore, the strongest interference is usually caused by UEs within the same cell [34, p. 246]. Therefore, the coordination of pilot assignment becomes essential for intra-cell pilot reuse systems.

The expression in (8) makes the effect of the pilot reuse on the channel estimation accuracy evident. However, we can avoid pilot contamination by assigning orthogonal pilot sequences to the UEs with non-orthogonal channels, i.e., $\mathrm{tr}\left(\mathbf{R}_n^{\mathrm{H}}\mathbf{R}_j\right) \neq 0$. In practice, it is not likely that the covariance matrices for distinct UEs are fully orthogonal, i.e., that their channels are spatially uncorrelated. Nevertheless, as shown in [43], it is still possible to mitigate pilot contamination between the UEs that have non-overlapping AoA intervals, i.e., $\mathcal{A}_{k,s} \cap \mathcal{A}_{j,s} = \{\emptyset\}$, $\forall k \neq j$, and $\forall s$. This behavior is also reported in [34, p. 341] as the asymptotic spatial orthogonality between covariance matrices, and it can be written as

$$\lim_{M \to \infty} \frac{1}{M}\mathrm{tr}\left(\mathbf{R}_n^{\mathrm{H}}\mathbf{R}_j\right) = 0, \quad \forall n \neq j, \tag{12}$$

where $\mathrm{tr}\left(\mathbf{R}_n^{\mathrm{H}}\mathbf{R}_j\right)$ is the standard inner product on $\mathbb{C}^{M \times M}$, measuring the orthogonality between $\mathbf{R}_n$ and $\mathbf{R}_j$.

Essentially, the pilot allocation algorithm must assign orthogonal pilot sequences to UEs with potentially strong interference. Thus, we can minimize the pilot contamination effect by wisely assigning the pilot sequences such that UEs with overlapping AoA intervals get orthogonal pilot sequences.

As motivated above, the angular separation between two UEs is closely related to their degree of orthogonality. An intuitive measure to assess the UEs' orthogonality level is presented in [26], based on the inner product of the covariance matrices. Similarly, we define the normalized channel spatial correlation between UEs n and j as

$$\delta\left(\mathbf{R}_n, \mathbf{R}_j\right) = \frac{\mathrm{tr}\left(\mathbf{R}_n^{\mathrm{H}}\mathbf{R}_j\right)}{\|\mathbf{R}_n\|_F \|\mathbf{R}_j\|_F}. \tag{13}$$

The spatial correlation between two UEs expresses how orthogonal the channels for the pair of UEs are, or how close these UEs are in the angular domain. The higher the correlation between the channel covariance matrices, the closer the UEs are in the angular domain, and the lower is the orthogonality between their channels.

To exploit the spatial characteristics of the channels for large antenna systems presented in (12), we formulate the pilot assignment problem to suppress the interference from UEs that share the same second-order statistics. More precisely, when UEs n and j share a pilot, we measure their contribution to the networks' pilot contamination by

$\delta(\mathbf{R}_n, \mathbf{R}_j)$; otherwise, the contribution is zero. Thus, we aim at minimizing the total amount of pilot contamination in the network through the optimization problem

$$\underset{\{\pi_n \in \mathcal{T}\}_{n=1}^N}{\text{minimize}} \quad \frac{1}{\tau N(N-1)/2} \sum_{n \in \mathcal{N}} \sum_{j > n} \delta\left(\mathbf{R}_n, \mathbf{R}_j\right) \phi_{\pi_n}^{\mathrm{T}} \phi_{\pi_j}, \tag{14}$$

where the optimization is over the UE-to-pilot index mapping, i.e., over the N integer variables π_n, each taking values on $\{1, \ldots, \tau\}$. The pre-summation factor normalizes the effect of changing the number of UEs and the length of the pilot sequences, but do not affect the minimization problem. Note that, due to orthogonality of the pilot codebook, the inner product $\phi_{\pi_n}^{\mathrm{T}}\phi_{\pi_j}$ acts as an indicator function taking the value τ if $\pi_n = \pi_j$ and 0 otherwise. Thus, the cost per UE pair (n, j) is always either $\delta(\mathbf{R}_n, \mathbf{R}_j)$ or 0. With this logic, to minimize (14), one should assign orthogonal pilots, distinct indices, to UEs with strong spatial correlation. It is also worth noting that the normalization of the inner product in (13) ensures fair pilot assignment for UEs with poor channels, as the interference level is normalized for each pair of UEs.

Problem (14) can be understood as a vertex coloring problem on an edge-weighted graph, and as such it is NP-hard [44]. For small-to-moderate-sized networks, the optimal pilot assignment can be found via exhaustive search, but for larger networks, sub-optimal algorithms have to be used. From [45], it is known that solving weighted graph problems with simplified local search algorithms can lead to far-from-optimal solutions. Motivated by this, we search for practical, yet efficient metaheuristics to allocate pilot sequences. We devise a two-stage method to find a practical solution to (14) by the following procedure. The first stage utilizes CC to retrieve an interference map of the UEs, determined by the coefficients $\delta(\mathbf{R}_n, \mathbf{R}_j)$ in (13). In particular, we tailor the underlying interference pattern into a lower dimensional map so that this map subsequently allows performing a simple greedy pilot assignment strategy conforming to the structure of the optimal solution of (14): assign orthogonal pilot sequences to UEs with strong pair-wise interference. In what follows, the proposed interference mapping is presented in Section IV, and the proposed pilot assignment method is presented in Section V.

IV. CHANNEL CHARTING TO RECOVER AN INTERFERENCE MAP

In this section, we introduce a CC method to retrieve a UE-interference map from the information embedded in the coefficients $\delta(\mathbf{R}_n, \mathbf{R}_j)$ in (13). Recall from above that this represents the first (and crucial) stage of our proposed solution to problem (14): the generated CC-based interference map will be the main input to the second stage, i.e., for the actual pilot assignment algorithm presented in Section V.

The standard CC, as presented in [32], aims at reconstructing a low-dimensional embedding from the CSI features so as to be a good representation of the spatial geometry of UEs. Here, we want to create a map that preserves the angular neighborhood in a low-dimensional chart, building on the

fact that the angular separation between UEs represents their inter-UE interference, which is measured by (13). The target is to map spatially correlated UEs, with potentially strong interference, close in the interference map.

Clearly, the features to generate the interference mapping must contain spatial information about the UEs, particularly, their angular positions. To this end, the second-order statistics of CSI, i.e., the covariance matrices $\mathbf{R}_n$, $n \in \mathcal{N}$, available at the BS are good candidates to create CC map. Beyond carrying the required information, they possess desirable characteristics such as varying over larger time scales as compared to the instantaneous CSI. However, if one cannot properly estimate the channel covariance matrices, the UE-interference map may not reflect the true potential interference between UEs, which degrades the performance of the proposed method. Rapid and abrupt changes in the channel geometry may cause significant variations in the covariance matrices. Such cases are not in the scope of the current paper, but form an important item for further study.

The CMD metric [27] measures the normalized orthogonality between two covariance matrices, which makes it a good candidate as a feature for the optimization problem in (14). Accordingly, the (n, j)-th element of the feature matrix $\mathbf{F} \in \mathbb{R}^{N \times N}$ generated from the covariance matrices of UEs n and j is

$$[\mathbf{F}]_{n,j} = 1 - \delta(\mathbf{R}_n, \mathbf{R}_j), \tag{15}$$

where $\delta(\cdot, \cdot)$ is the normalized channel spatial correlation of (13). The more the signal spaces of $\mathbf{R}_n$ and $\mathbf{R}_j$ overlap, the larger the interference between UE n and j, and the smaller is the distance measured by (15) [27]. Because of its normalization factor, it can also be interpreted as an angular distance between the UEs. The n-th column of $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N] \in \mathbb{R}^{N \times N}$, $\mathbf{f}_n \in \mathbb{R}^N$, represents the feature or dissimilarity vector that contains the distance of UE n to all N UEs, including itself.

After acquiring the feature matrix $\mathbf{F}$, the last step to get CC is to apply a function $\mathcal{C}$ on $\mathbf{f}_n$, $n \in \mathcal{N}$, to map the set of UE features to a lower dimensional chart,

$$\mathcal{C} : \mathbf{f}_n \mapsto \mathbf{z}_n, \tag{16}$$

where $\mathbf{z}_n \in \mathbb{R}^C$ is the point in the C-dimensional CC corresponding to feature $\mathbf{f}_n$, where typically $C \ll N$. Apart from providing a lower dimensional representation of the mutual interference between UEs, which may help to visualize an interference pattern, the dimensionality reduction (DR) technique maps the features to the dimensions that have the most relevant attributes to characterize the interference. The DR technique is essential to alleviate the *distance concentration* effect. This phenomenon reduces the difference in distance between near and far neighbors as the features' dimensionality increases, causing problems for algorithms that rely on nearest neighbor search – the category in which our pilot allocation algorithm falls into [46], [47]. Therefore, DR techniques will be applied, as described next.

A. DIMENSIONALITY REDUCTION TECHNIQUES

Ultimately, we wish to design a function $\mathcal{C}$ in (16) that maps the feature vectors $\mathbf{f}_n$, $n \in \mathcal{N}$, into a lower dimensional chart such that similar features are mapped close while highly dissimilar features are mapped far away in the chart. Several DR techniques have been used to generate CCs [30], [31], [32], [48], [49], [50]. Here, two DR techniques are briefly presented: 1) principal component analysis (PCA) and 2) Isomap, which will be the DR technique used in our numerical experiments. PCA is a very robust and easy-to-implement algorithm, as it performs a linear mapping from the feature domain to CC, which, in turn, may be a limiting factor for non-linear embeddings. On the other hand, Isomap can handle more complex scenarios because it is not restricted to linear mappings.

1) PRINCIPAL COMPONENT ANALYSIS

PCA is a widely used algorithm, which performs a linear projection of high-dimensional data onto a subspace of lower dimension [51]. The main idea behind PCA is to find the components that maximize the variance of the projected data. Let $\mathbf{C} = \mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{N \times N}$ be the centering matrix. An interference map is then formed by applying singular value decomposition on the covariance of the centered feature matrix, $\mathbf{K} = \mathbf{F}^T\mathbf{C}\mathbf{F} \in \mathbb{R}^{N \times N}$, and retrieving the coordinates in the C-dimensional space with the largest variance values implying the maximum interference [52]. Thereby, we obtain the CC points, $\mathbf{Z}_C = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{C \times N}$, via PCA as

$$\mathbf{Z}_C^{\text{PCA}} = \left[\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_C}\mathbf{u}_C\right]^T, \tag{17}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_C$ are the C largest eigenvalues of $\mathbf{K}$ and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_C$ are the associated eigenvectors.

Although PCA has shown good performance for single-sector scenarios [30], its linearity reduces its capability to achieve good mapping in more challenging scenarios, like the 3-sector deployment considered herein. Thereby, more sophisticated DR techniques are required.

2) ISOMAP

Isomap is a DR technique that aims to preserve the geodesic distance or the curvilinear distance measured over the higher dimensional space between the features [53]. The geodesic distance is computed by constructing a neighborhood graph G, in which every UE is connected with its ν nearest neighbors, where the integer ν is a design parameter [54].

The first step of the Isomap algorithm is to find the ν closest (neighboring) UEs for each UE in the feature domain, represented by $\mathbf{F} \in \mathbb{R}^{N \times N}$. To find the ν neighbors of UE n, we look for the $\nu + 1$ smallest elements in $\mathbf{f}_n$. Then, we construct a weighted graph G based on neighbor information. The weights in G connecting two vertices, i.e., UEs, that are deemed to be among the nearest neighbors correspond to the distances in $\mathbf{F}$ between these two UEs, whereas the weights of non-nearest neighbors are set to a large number. Note
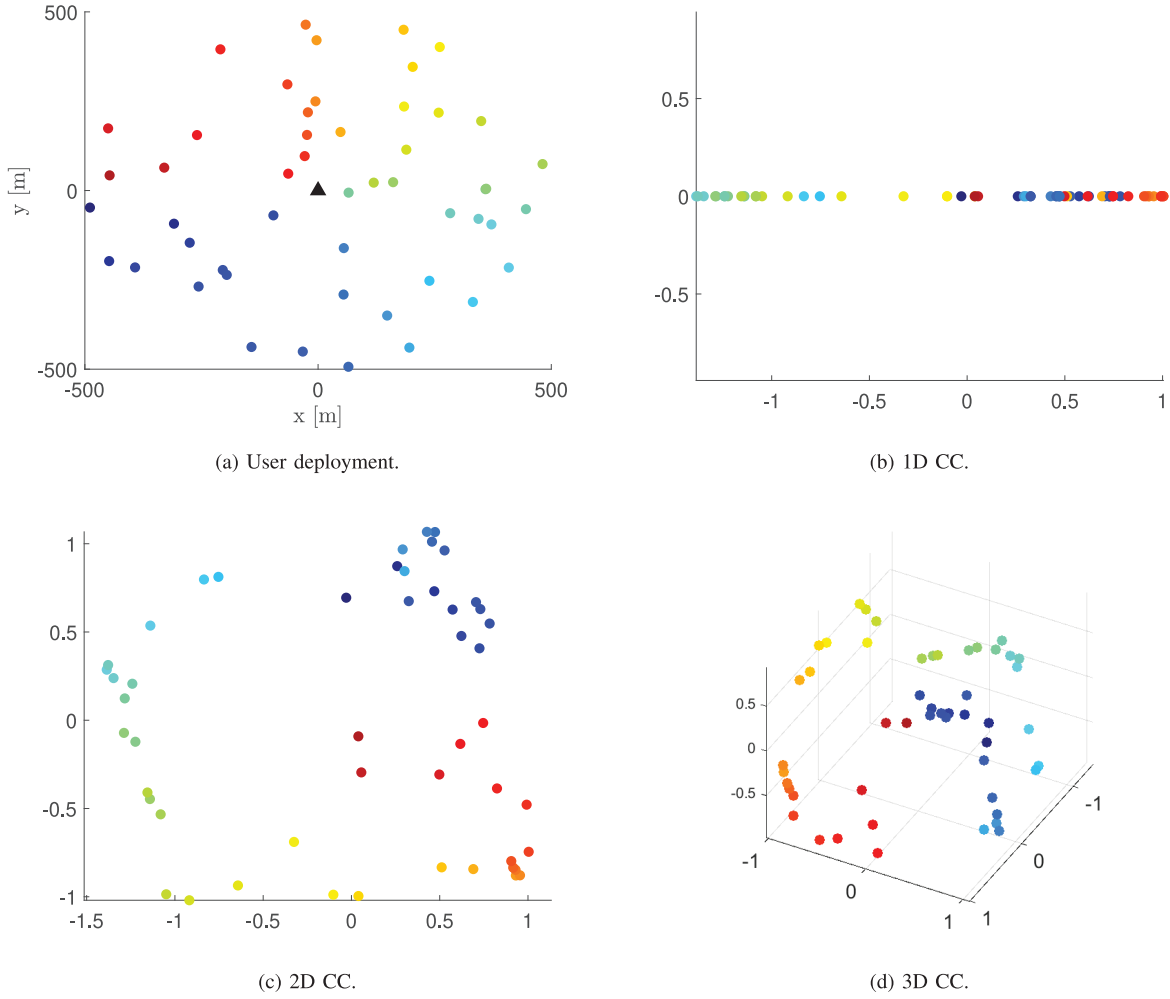
FIGURE 3. Channel charting mapping using Isomap. (a) shows the physical location of 40 UEs colored based on their AoA to the BS. The BS, represented by the black triangle, is equipped with $S = 3$ ULAs, each having $M = 64$ antenna elements. (b)–(d) represent the CC mapping of (a), for $C = 1$, $C = 2$ and $C = 3$ dimensions, respectively.

that G is a directed graph, i.e., UE j may not be in the neighborhood of UE n, even if UE n is in the neighborhood of j.

After acquiring the graph G, one needs to compute the geodesic distances across G using a shortest path algorithm, such as Dijkstra [55] or Floyd's [56] algorithm. The output of the shortest path algorithm is a dissimilarity matrix $\mathbf{F}' \in \mathbb{R}^{N \times N}$ which consists of geodesic distances computed along the nearest-neighbor graph.

The last step of Isomap consists of applying classical multidimensional scaling (MDS) to the dissimilarity matrix $\mathbf{F}'$ [53]. To start with this, we compute the doubly centered dissimilarity matrix $\mathbf{K}' = -\frac{1}{2}\mathbf{C}(\mathbf{F}' \odot \mathbf{F}')\mathbf{C}$. Next, we extract the CC coordinates by taking the eigenvalue decomposition of $\mathbf{K}'$. Let $\lambda_1', \lambda_2', \ldots, \lambda_C'$ be the C largest eigenvalues of $\mathbf{K}'$ and $\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_C'$ the associated eigenvectors. Finally, the CC mapping, when utilizing Isomap as the DR technique, is obtained as

$$\mathbf{Z}_C^{\text{ISO}} = \left[\sqrt{\lambda_1'}\mathbf{u}_1', \ldots, \sqrt{\lambda_C'}\mathbf{u}_C'\right]^{\text{T}}. \tag{18}$$

The dimensionality of the final embedding affects the accuracy of the representation of the features in the lower dimensional chart. This is illustrated in Fig. 3. In Fig. 3(a), $N = 40$ UEs are uniformly distributed in a 1 km^2 cell and colored with respect to their AoAs to the BS. Figs. 3(b), (c), and (d) depict the CC mapping obtained for $C = 1$, $C = 2$ and $C = 3$ dimensions, respectively. Here, one can clearly see that the accuracy of the angular domain representation increases as C increases.

In quantitative terms, the accuracy of the CC embedding is measured by the residual variance between the feature distances in the high-dimensional space, $\mathbf{F}$ in (15), and in the CC domain, $\mathbf{D}_{\mathbf{Z}_C} = [\mathbf{d}_1, \ldots, \mathbf{d}_N] \in \mathbb{R}^{N \times N}$. The (n, j)-th entry of the distance matrix $\mathbf{D}_{\mathbf{Z}_C}$ is computed as the ℓ_2-norm of the difference between the CC points $\mathbf{z}_n$ and $\mathbf{z}_j$, i.e., $[\mathbf{D}_{\mathbf{Z}_C}]_{n,j} = \|\mathbf{z}_n - \mathbf{z}_j\|_2$. Therefore, the residual variance of the CC mapping is given by

$$r(\mathbf{F}, \mathbf{D}_{\mathbf{Z}_C}) = 1 - \frac{\sum_{n=1}^{N}\left(\mathbf{f}_n - \bar{\mathbf{f}}_n\right)^{\text{T}}\left(\mathbf{d}_n - \bar{\mathbf{d}}_n\right)}{\sqrt{\text{Var}(\mathbf{F})\text{Var}(\mathbf{D}_{\mathbf{Z}_C})}}, \tag{19}$$

(a) UE deployment.

(b) 2D CC.

(c) Active UEs.

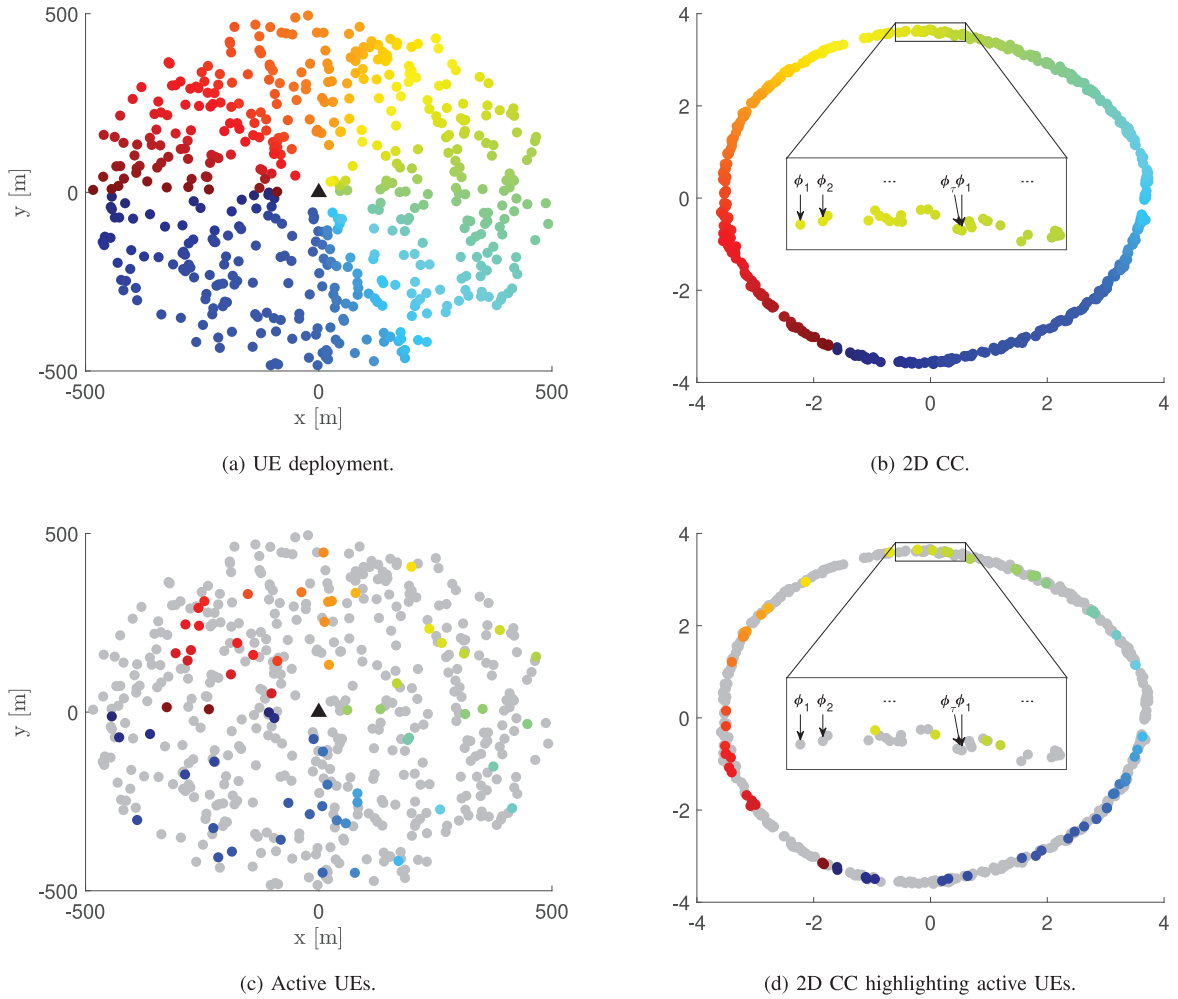(d) 2D CC highlighting active UEs.

FIGURE 4. Illustration of the proposed CC-based pilot allocation algorithm. (a) shows the physical locations of 512 UEs deployed in a cell with $S = 3$ ULAs and (b) depicts the pilot assignment for a fraction of the UEs, detailed in the zoomed region. (c) shows the same deployment as in (a), yet highlighting only the set of active UEs and (d) shows the corresponding pilot assignment.

where $\mathrm{Var}(\mathbf{F}) = \sum_{n=1}^{N}(\mathbf{f}_n - \bar{\mathbf{f}}_n)^{\mathrm{T}}(\mathbf{f}_n - \bar{\mathbf{f}}_n)$ and $\mathrm{Var}(\mathbf{D}_{\mathbf{Z}_C}) = \sum_{n=1}^{N}(\mathbf{d}_n - \bar{\mathbf{d}}_n)^{\mathrm{T}}(\mathbf{d}_n - \bar{\mathbf{d}}_n)$. Note that the second term in (19) is the Pearson's correlation coefficient for the vectorized distance matrices $\mathbf{F}$ and $\mathbf{D}_{\mathbf{Z}_C}$ [57, p. 157]. In Section V, the residual variance is utilized to find a good compromise between the number of dimensions of CC and the accuracy of the features' representation in the CC, which impacts the performance of the proposed pilot assignment algorithm.

V. PILOT REUSE ALGORITHM

To minimize the total amount of pilot contamination as formulated in (14), we propose a pilot reuse strategy that exploits the interference map generated through CC, as presented in Section IV, to assign orthogonal pilot sequences to UEs inducing potentially strong mutual interference. The pilot sequences are allocated to all UEs regardless of their activity pattern. Therefore, reassignment is needed only if the network setup or the UEs' relative positions vary considerably. Because the interference pattern depends on the UEs'

relative positions, if the latter changes, the pilot sequences must be reassigned to accommodate the new interference pattern. In practice, the relative positions of UEs are related to their physical positions, namely their angular positions. Therefore, if the UE mobility is high, the pilot allocation algorithm has to be re-executed more frequently. To handle cases where few devices join or disconnect from the network, out-of-sample CC [58] along with a pilot allocation margin could be implemented to avoid re-creating the whole UE-interference map and reassigning all the pilot sequences.

Furthermore, we propose an adaptive CC to recover the interference map with a flexible number of dimensions. This additional flexibility in the UE-interference map increases the robustness of the method to changes in the UEs' density by preserving the meaningful structures embedded in the features while discarding unnecessary dimensions. That improves the CC interference map representation, in particular, for low-density (sparse) deployments.

A. CC-AIDED PILOT ASSIGNMENT

The proposed CC-aided pilot reuse scheme for single-cell mMIMO deployments is divided into two main phases: 1) the acquisition of the UEs' interference map, and 2) the low-complexity nearest neighbor pilot assignment algorithm.

1) ADAPTIVE CC INTERFERENCE MAP

To generate the interference map with adaptive dimensions, we apply Isomap with the same principles as described in Section IV-A. However, instead of fixing C from the beginning, we propose to start with a small number of dimensions and increase it to meet the two mapping accuracy criteria described below. In practice, we start with $C = 1$ and increase the dimensionality of the resulting embedding only if the residual variance of CC given by (19) is higher than a threshold ϵ or if it varies less than ξ from increasing C, i.e., $r(\mathbf{F}, \mathbf{Z}_C) > \epsilon$ or $|r(\mathbf{F}, \mathbf{Z}_C) - r(\mathbf{F}, \mathbf{Z}_{C-1})| > \xi$. The thresholds ϵ and ξ are design parameters to tune the number of dimensions underlying the interference information embedded on $\mathbf{F}$. To shed some light, based on our empirical observations, $\epsilon = 0.0001$ and $\xi = 0.0001$ present a good compromise between the number of dimensions required and the residual variance to represent the features' embedding. Also, the numerical results will show that the number of dimensions required to have an accurate interference map depends on the density of UEs.

2) GREEDY PILOT ALLOCATION

The problem still remains weighted graph coloring, and thus NP-hard, but with distances computed in reduced dimensions. Intuitively, finding a solution has become easier, as the applied DR mechanism reduces the effect of the bulk of small interferences and enhances the large interferences, thus reducing the probability of falling to a bad local minimum. We apply greedy allocation of orthogonal pilots to the nearest UEs (neighbors) in the CC domain, as described in Algorithm 1 and depicted on the flow chart in Fig. 5. Having acquired an interference map where the strongly interfering UEs are mapped closer than the weakly ones, the main idea of the proposed pilot assignment algorithm is to allocate the pilot sequences in an ordered way, $\phi_1, \phi_2, \ldots, \phi_\tau, \phi_1, \phi_2, \ldots$, aiming at maximizing the distances between the same pilot sequences on the interference map. This is equivalent to maximizing the AoA distances between UEs that have the same pilot sequences. We delve into details of the proposed algorithm in the following.

To start the pilot allocation, the first orthogonal pilot sequence, ϕ_1, is assigned to a randomly chosen UE. In sequence, the next pilot sequence is assigned to the unassigned UE with the smallest Euclidean distance to the previously allocated user, i.e., ϕ_2 is allocated to the UE n that minimizes $\|\mathbf{z}_n - \mathbf{z}_{n'}\|_2^2$, where n' is the index of the previously allocated UE to ϕ_1. This process is repeated until all orthogonal pilot sequences have been allocated. From this point onward, the pilot sequences are to be reused by allocating the first sequence, ϕ_1, to the closest unassigned UE
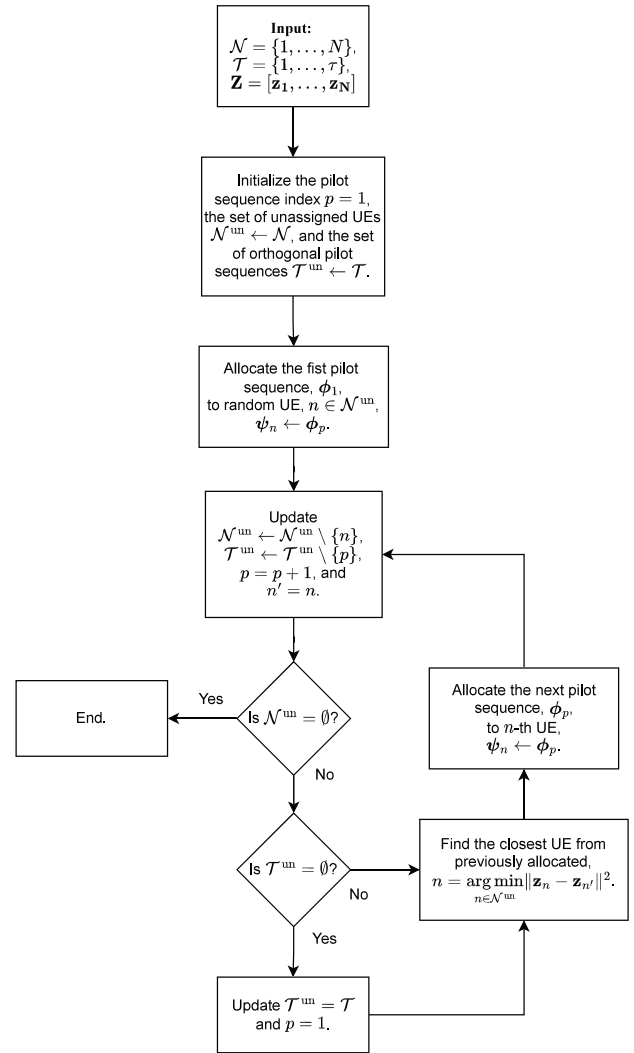

FIGURE 5. Flow chart of the proposed nearest neighbor pilot allocation algorithm. The algorithm greedily allocates the pilot sequences to maximize the distance between the same pilot sequences in the CC domain.

from the precedent allocated user. This process is repeated until all UEs have been assigned a pilot sequence.

Fig. 4 exemplifies the pilot allocation algorithm procedure for a cell with $N = 512$ UEs. Fig. 4(a) shows the UEs' real positions, which are colored based on their AoA with respect to the BS, and Fig. 4(b) shows the respective two-dimensional CC, when Isomap is used as the DR technique. The proposed pilot allocation procedure, as described in Algorithm 1, is exemplified at the zoomed region in Fig. 4(b). Figs. 4(c) and (d) illustrate the effect of UEs' activity on the pilot assignment for a random set of active UEs. Comparing Fig. 4(b) and Fig. 4(d), one can see that the UEs' activity does not affect the pilot assignment. Moreover, comparing Fig. 4(b) with Fig. 3(b), we see that a two-dimensional CC may be a good representation of the angular relationship between UEs when their density in the cell is high. This evidences that the density of UEs in the cell affects the capability of Isomap to accurately map

---

Algorithm 1: Nearest Neighbor Pilot Allocation

Input: 1) The set of UEs $\mathcal{N} = \{1, \ldots, N\}$, 2) the index set of orthogonal pilots $\mathcal{T} = \{1, \ldots, \tau\}$, and 3) the suitable features $\mathbf{z}_n$, $n \in \mathcal{N}$.

Output: A pilot assignment $\mathbf{\Psi} = [\psi_1, \ldots, \psi_N]^{\mathrm{T}}$.

1   Initialize the set of unassigned UEs $\mathcal{N}^{\mathrm{un}} = \mathcal{N}$, and the set of unassigned pilots $\mathcal{T}^{\mathrm{un}} = \mathcal{T}$.

2   Select a random UE n and initialize the auxiliary variable n' with it, i.e., $n' = n$.

3   Assign ϕ_1 to user n and update the set of unassigned UEs and pilots, i.e., $\mathcal{N}^{\mathrm{un}} \leftarrow \mathcal{N}^{\mathrm{un}} \setminus \{n\}$ and $\mathcal{T}^{\mathrm{un}} \leftarrow \mathcal{T}^{\mathrm{un}} \setminus \{1\}$.

4   Initialize the auxiliary variable: $p = 2$.

5   **while** $\mathcal{N}^{\mathrm{un}} \neq \emptyset$ **do**

6   **if** $\mathcal{T}^{\mathrm{un}} = \emptyset$ **then**

7   Reinitialize: $\mathcal{T}^{\mathrm{un}} = \mathcal{T}$ and $p = 1$.

8   **end**

9   Assign pilot ϕ_p to user n, i.e., $\psi_n = \phi_p$, that satisfies $n = \arg\min_{n \in \mathcal{N}^{\mathrm{un}}} \|\mathbf{z}_n - \mathbf{z}_{n'}\|^2$.

10   Update the set of unassigned UEs, $\mathcal{N}^{\mathrm{un}} \leftarrow \mathcal{N}^{\mathrm{un}} \setminus \{n\}$, and the set of unassigned pilots, $\mathcal{T}^{\mathrm{un}} \leftarrow \mathcal{T}^{\mathrm{un}} \setminus \{p\}$.

11   Update $n' = n$ and $p = p + 1$.

12 **end**

---

the underlying interference pattern among the UEs into a low-dimensional chart.

B. COMPLEXITY ANALYSIS

We evaluate the total complexity order of the proposed algorithm through $\mathcal{O}(\cdot)$ notation as: 1) the complexity of generating the adaptive CC interference map through Isomap and 2) the complexity of allocating the pilot sequences using the nearest neighbor pilot algorithm, presented in Algorithm 1.

The complexity of generating the CC through Isomap is divided into computing the feature matrix $\mathbf{F}$ in (15), finding the ν nearest neighbors, computing the shortest paths across the neighborhood graph G, and performing classical MDS. The cost of computing the CMD distances between all the UE pairs in the MS-dimensional space is $\mathcal{O}(M^3 S^3 N^2)$. The complexity to find the ν nearest neighbors is $\mathcal{O}(\nu N^2)$. To compute the shortest paths across the graph G using Floyd's algorithm (used in the simulations) requires $\mathcal{O}(N^3)$. Note that the complexity to compute the shortest paths can be reduced to $\mathcal{O}(\nu N^2 \log(N))$ using Dijkstra's algorithm [59]. Lastly, the complexity of computing the doubly centered matrix $\mathbf{K}'$ and its eigenvalue decomposition is $\mathcal{O}(N^3)$ [60].

The complexity of the pilot allocation algorithm in Fig. 5 is given as the cost to compute the pair-wise distances between the N CC points in the C-dimensional embedding, which has the complexity $\mathcal{O}(CN^2)$, and then finding the closest unassigned UE which has complexity $\mathcal{O}(N^2)$.

TABLE 1. Algorithm complexity analysis.

Scheme	Complexity Order
AoA UE grouping [22]	$\mathcal{O}(N^2)$
SGPS [26]	$\mathcal{O}(M^3 S^3 N^3)$
CMD [31]	$\mathcal{O}((M^3 S^3 + 1)N^2 + N^3)$
CC-aided[4]	$\mathcal{O}((M^3 S^3 + \nu + C + 1)N^2 + 2N^3)$

In summary, the complexity order of running the proposed algorithm is $\mathcal{O}((M^3 S^3 + \nu + C + 1)N^2 + 2N^3)$. As comparison, the complexity of SGPS [26], AoA UE grouping [22], and CMD-based [31] pilot allocation algorithms are also summarized in Table 1. The proposed method has the same order of complexity as the CMD method, up to a scalar factor, namely $\nu + C$. AoA UE grouping has a much lower complexity than the other baseline methods, only $\mathcal{O}(N^2)$. For scenarios where the total number N of UEs is much larger than the number of antenna elements M times the number of sectors S, the complexity of the proposed method is lower than that of SGPS. Also, assuming that the CC interference map is utilized for other network management purposes, thereby already available, the cost of assigning the pilots is only that of applying the greedy pilot assignment algorithm, which has order of complexity $\mathcal{O}((C + 1)N^2)$.

VI. SIMULATION RESULTS

We consider a mMIMO MTC system with $N = 512$ UEs uniformly distributed within a 1 km^2 cell area, where $K = 64$ UEs are simultaneously active at any given transmission interval. The BS is equipped with $S = 3$ ULAs each having $M = 64$ critically spaced antenna elements ($\Delta_{\mathrm{r}} = 0.5$). The propagation channel between each UE and the BS is modelled as the superposition of $L = 200$ paths with a fixed angular standard deviation $\sigma_\theta = 10°$, unless otherwise specified. The maximum antenna gain in (4) is $G_{A_{\max}} = 0$ dB, the maximum attenuation is $A_{\max} = 30$ dB, and the half-power beamwidth is $\theta_{\mathrm{3dB}} = 65°$ [41]. Isomap DR is used. The residual variance thresholds are $\epsilon = 0.0001$ and $\xi = 0.0001$. Binary phase shift keying (BPSK) is used for the pilots and quadrature phase shift keying (QPSK) for the data transmission. The main simulation parameters are summarized in Table 2.

Channel estimation accuracy is evaluated in terms of the normalized MSE

$$\text{NMSE CE} = \frac{\mathbb{E}\left[\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2\right]}{\mathbb{E}\left[\|\mathbf{H}\|_F^2\right]}. \tag{20}$$

The expected value is estimated through Monte Carlo simulation so that the NMSE is approximated by normalized average square error, but we use the acronym NMSE for clarity in the sequel. We adopt the robust MMSE receiver derived in [26, Eq. (27)], which considers that only channel estimates are known by the BS and, thus, takes into account the channel estimation error. This receiver combining structure is known for minimizing the MSE for symbol

TABLE 2. Simulation parameters.

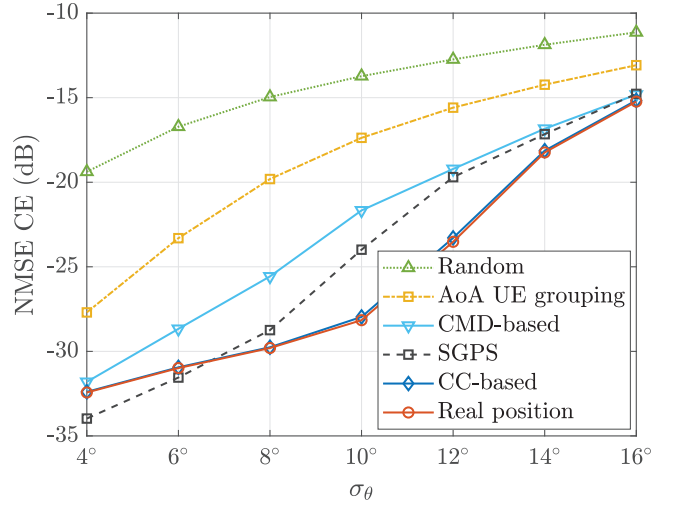|	
Cell area	1 km^2
Number of UEs (N)	512
Number of active UEs (K)	64
Number of sectors (S)	3
Number of antenna elements per ULA (M)	64
Carrier frequency (f_c)	6 GHz
Wavelength (λ)	0.05 m
Number of paths (L)	200
Normalized antenna spacing (Δ_r)	0.5
Angular standard deviation (σ_θ)	10°
Maximum antenna gain ($G_{A_{\max}}$)	0 dB
Maximum antenna attenuation ($A_{\max}$)	30 dB
Half-power beamwidth (θ_{3dB})	65°

FIGURE 6. Normalized mean square error for channel estimation versus the angular standard deviation. More directive channels, i.e., narrower angular spread captured by σ_θ, give better channel estimates because the pilots are less likely to interfere with each other.

detection (MSE SD) and maximizing the instantaneous SINR simultaneously [26], [34]. Therefore, the performance of the proposed method is also evaluated in terms of MSE SD, and average achievable uplink rate per UE

$$r_k = \mathbb{E}\big[\log_2 (1 + \gamma_k)\big], \tag{21}$$

where the UE-specific SINR γ_k, when employing the robust beamforming, is given by

$$\gamma_k = \hat{\mathbf{h}}_k^H \left(\sum_{j \neq k} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^H + \sum_{p=1}^{K} \mathbf{R}_{\tilde{\mathbf{h}}_p} + \frac{\sigma^2}{p_u} \mathbf{I}_M \right)^{-1} \hat{\mathbf{h}}_k. \tag{22}$$

Five baseline pilot assignment methods are considered:
- *"Random"*: This method assigns the orthogonal pilot sequences uniformly at random while respecting the pilot reuse factor, i.e., the same pilot sequence is reused the minimum required times when allocating one sequence to each UE.
- *"SGPS"*: The statistical greedy pilot scheduling (SGPS) developed in [26], which relies on the knowledge of the channel covariance matrices and mitigates the pilot contamination by grouping the UEs with high spatial correlation and assigns them orthogonal pilot sequences.
- *"AoA UE grouping"*: A multi-cell pilot assignment method proposed in [22] adapted to the single-cell scenario. This method relies on the exact UEs' positions to group the UEs based on their AoA. The orthogonal pilot sequences are randomly allocated within each group.
- *"Real position"*: This method assumes knowledge of the exact physical locations of UEs to compute the angular distances and then applies our nearest neighbor pilot allocation algorithm presented in Algorithm 1. Note that this method acts as a lower bound to the proposed one, since it utilizes the real positions of the UEs instead of the relative ones provided by CC.
- *"CMD"*: The CMD-aided pilot assignment method employs the nearest neighbor pilot allocation algorithm (Algorithm 1) directly in the feature matrix, **F** in (15),

without using any DR technique. This baseline demonstrates the benefit of applying a DR technique in the proposed features.

Fig. 6 presents the performance of the proposed method in terms of NMSE CE for a fixed SNR value of 10 dB and the angular standard deviation σ_θ ranging from 4 to 16 degrees. As expected, a smaller angular standard deviation yields a lower channel estimation error. This occurs because the more directive the UEs' channels are, the lower is the interference one UE causes to another. Here, we can see that the proposed method works well for a moderate and large angular spread of channels, beating all benchmark methods for $\sigma_\theta \geq 8°$, but loses to SGPS algorithm when σ_θ becomes very narrow. Nonetheless, for such a small angular spread, the performance of the proposed method is as good as the real position benchmark. As reported in [61], reasonable values for the angular standard deviation are between 6° and 21°. Therefore, we opted for presenting all the subsequent results for $\sigma_\theta = 10°$. Although the CMD-based method greatly improves the performance as compared to the random pilot allocation and AoA UE grouping, it underperforms compared to the proposed method for all considered angular spreads of channels, highlighting the advantages of preprocessing the features by applying a DR technique.

Fig. 7 shows how the dimensionality of the CC affects the performance of the proposed method in terms of NMSE CE and MSE SD for $\sigma_\theta = 10°$ and pilot length $\tau = 64$. The behaviors of the NMSE CE and MSE SD are shown in Fig. 7(a) and Fig. 7(b), respectively, as a function of the threshold parameter ϵ for the residual variance in (19). Also, the average dimensionality of the interference map generated by CC is displayed on the top of the curve. In Fig. 7(a), we can see that the performance of the pilot allocation is quite robust to changes in ϵ, i.e., large changes on ϵ have small effects on the NMSE CE. Therefore, one can reduce the
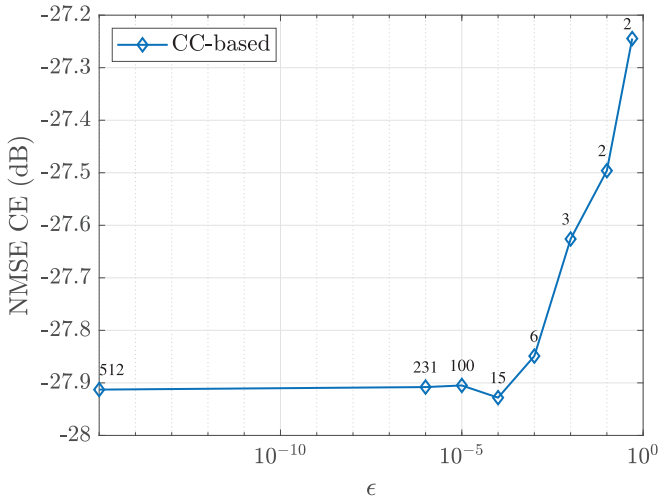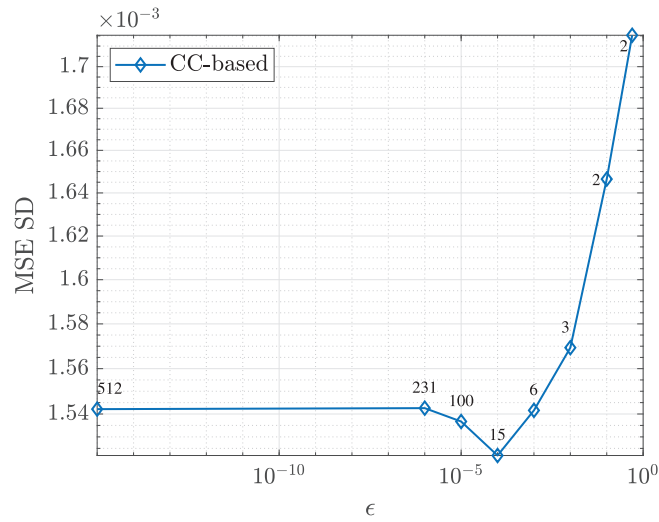
(a) Average NMSE for channel estimation versus ϵ.

(b) Average MSE for symbol detection versus ϵ.

FIGURE 7. NMSE CE and MSE SD versus threshold parameter ϵ that controls the dimensionality of CC. The average CC dimension is shown for each value of ϵ on the top of the curve.

dimensionality of the interference map to decrease computational complexity and storage requirements without losing much in performance. In Fig. 7, ξ assumes the same value as ϵ except for $\epsilon = 0.1$ and $\epsilon = 0.5$ where it takes the values 0.01 and 0.1, respectively. Here, one can see that both the NMSE CE and MSE SD reach a point of diminishing returns by reducing ϵ at $\epsilon = 10^{-4}$. Therefore, we use $\epsilon = 10^{-4}$ and $\xi = 10^{-4}$ in the subsequent simulations.

Fig. 8 shows the NMSE CE and MSE SD as a function of the SNR for distinct pilot lengths, namely $\tau = \{32, 64, 128\}$. One can see that the pilot allocation strategy has a stronger impact at high SNR regimes, where the interference from UEs reusing the same pilot sequence becomes stronger. Note that for large pilot reuse factors, i.e., when the same pilot sequence is reused several times, the gap between the NMSE CE of the methods decreases, like shown in Fig. 8(a). This

effect is explained by the exhaustion of the capability of reducing interference, to the extent that even the optimal pilot allocation cannot avoid pilot contamination. That happens because for cases with high reuse of the pilot sequences, the likelihood of UEs with overlapping AoA intervals sharing the same pilot sequence increases. Furthermore, a lower bound for the MSE SD is also presented for benchmark, which is achieved considering perfect CSI knowledge in the LMMSE receive combining. It can be seen that the proposed CC-based method approaches the real position baseline while beating all other methods for moderate and high pilot reuse factors, $N/\tau = 8$ and 16, respectively. On the other hand, for a small pilot reuse factor $N/\tau = 4$, SGPS outperforms the proposed method.

The effect of increasing the number of antenna elements on the NMSE CE is presented in Fig. 9. For fixed SNR and pilot length, one can see that adding more antennas at the BS does not provide additional gains after some threshold. This is due to the fact that the interference is associated with the propagation characteristics of the environment, such as ASD. Thus, increasing the resolvability of the signals via a larger antenna array at the BS may not help distinguish between the UEs' signals. In fact, it can even be prejudicial depending on the pilot assignment strategy adopted, as for the CMD-based method in this particular setup. This happens because for large antenna arrays more samples are needed to get an accurate covariance matrix [62]. By comparing CMD-based method with CC-based method, it is clear that preprocessing the features before feeding them to the proposed pilot allocation algorithm drastically improves the performance.

Fig. 10 shows the uplink achievable rate as a function of the pilot length for SNR = 10 dB. One can notice that the CC-based method performs very well, approaching the real position based pilot allocation. However, its performance slightly degrades as compared to SGPS for $\tau > 70$ symbols, i.e., when the pilot reuse factor N/τ decreases below 7. Yet, in this regime, the proposed method performs as well as the real position one.

The performance of the proposed method is also evaluated against the exhaustive search pilot allocation for a toy scenario in Fig. 11. Conducting the exhaustive search throughout the orthogonal pilot sequences finds the optimum pilot assignment for (14). In this example, we consider $N = 10$ UEs, all active, deployed in a cell with $M = 16$ and $S = 3$ ULAs. To illustrate the effect of the interference on the MSE, we also plot the NMSE CE of the exhaustive search method for the noise-free scenario, i.e., $\sigma_n^2 = 0$. It is clear from this result that the performance is interference-limited, conditioned on the superposition of UEs' AoA intervals. All the methods other than AoA UE grouping and random pilot allocation perform close to the exhaustive search pilot assignment. The proposed method performs very similarly to the CMD method. That happens, because, for such a small setup with few UEs, more CC dimensions are required to meet the maximum residual variance criterion, $\epsilon \leq r(\mathbf{F}, \mathbf{Z}_C)$. That means a low value of C cannot properly reflect the
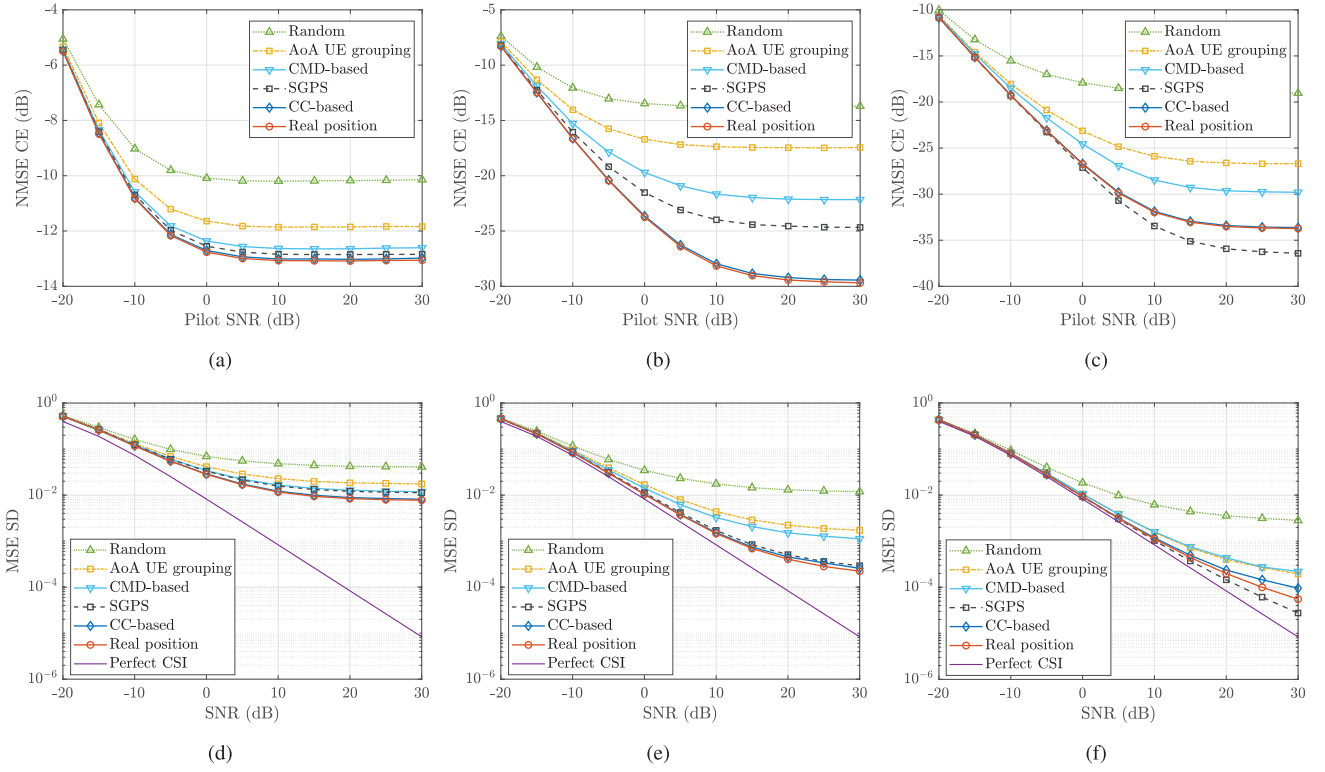
FIGURE 8. NMSE CE and MSE SD versus the SNR for distinct pilot reuse factors, $N/\tau = \{16, 8, 4\}$, respectively, and for different pilot assignment algorithms. (a) and (d) $\tau = 32$, (b) and (e) $\tau = 64$, and (c) and (f) $\tau = 128$.
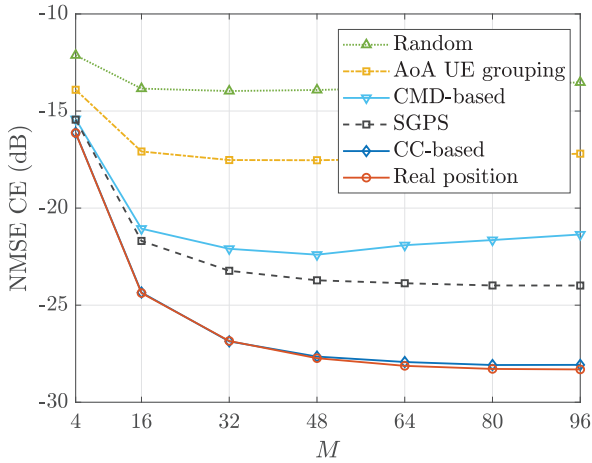

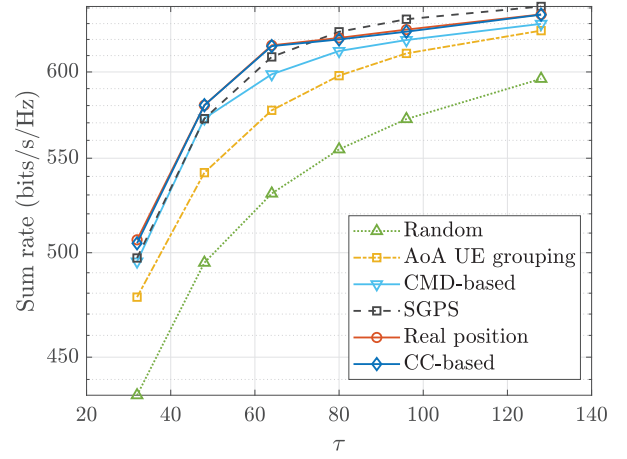**FIGURE 9.** NMSE CE versus the number of antennas M for $N = 512$, $K = 64$, $\tau = 64$, SNR= 10 dB.

FIGURE 10. Achievable sum rate for $K = 64$ active UEs versus pilot length at 10 dB SNR. For $\sigma_\theta = 10°$ and different pilot assignment algorithms.

UEs' mutual interference, which is seen here by comparing the performance of the proposed method with the CC-based method with fixed C (CC-based, $C = 2$). Thus, CC mapping keeps increasing the number of dimensions to meet the ξ and ϵ criteria, pushing the performance closer to the one presented by the CMD method.

VII. CONCLUSION
This paper addressed the pilot contamination problem in single-cell multi-sector mMIMO systems. We proposed

a novel pilot allocation algorithm based on a CC-aided interference map of UEs to cope with the pilot contamination caused by pilot reuse in mMIMO systems. The pilot sequences are assigned so that the angular distances between the UEs that share the same pilot sequence are maximized, which aims at minimizing the interference between UEs. The proposed CC-based pilot allocation algorithm showed significant improvements in terms of channel estimation error, symbol detection error, and achievable uplink rate when compared to existing pilot allocation schemes. In a toy example
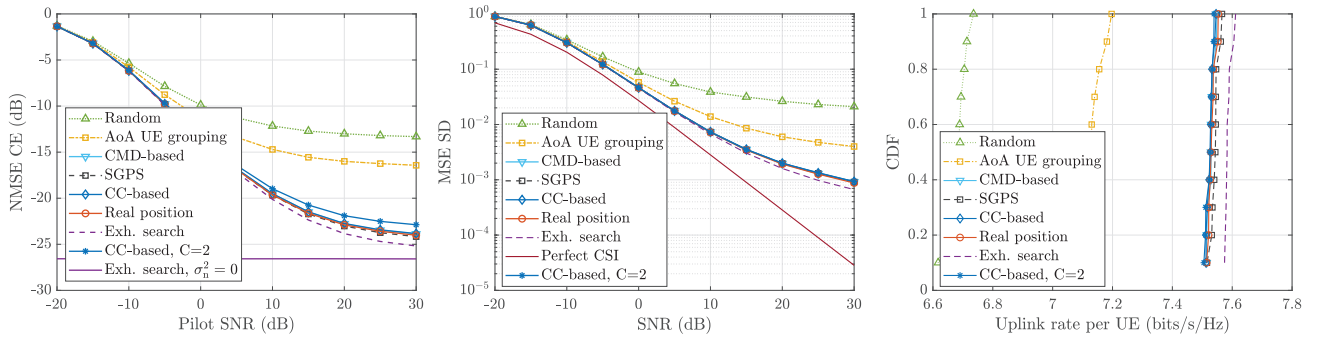
FIGURE 11. The NMSE and MSE-SD as function of SNR and the CDF of the uplink rate per UE, respectively for $N = 10$, $K = 10$, $\tau = 4$, $\sigma_\theta = 10°$, and $M = 16$.

with few UEs, the performance of the proposed method gets close to the optimal exhaustive search pilot allocation and approaches the real position benchmark for large pilot reuse ratios.

Several avenues for further research remain. The extension of the schemes to the multi-cell setup with cooperative multi-point processing or a cell-free architecture would be of interest. The mMTC literature for grant-free access has also other solutions (such as the use of non-orthogonal pilot sequences) to solve the pilot collision problem. Sequence-design-based solutions for joint channel estimation and UE activity detection with CC could be a fruitful approach for system design.

REFERENCES

[1] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[2] G. Callebaut, S. Gunnarsson, A. P. Guevara, A. J. Johansson, L. Van Der Perre, and F. Tufvesson, "Experimental exploration of unlicensed sub-GHz massive MIMO for massive Internet-of-Things," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2195–2204, 2021.

[3] H. Yan and H. Yang, "Pilot length and channel estimation for massive MIMO IoT systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15532–15544, Dec. 2020.

[4] B. M. Lee and H. Yang, "Massive MIMO with massive connectivity for Industrial Internet of Things," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5187–5196, Jun. 2018.

[5] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

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

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

[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[9] C. Bockelmann et al., "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[10] Y. Cheng, L. Liu, and L. Ping, "Orthogonal AMP for massive access in channels with spatial and temporal correlations," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 726–740, Mar. 2021.

[11] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[12] M. T. P. Le, L. Sanguinetti, E. Björnson, and M.-G. D. Benedetto, "Code-domain NOMA in massive MIMO: When is it needed?" *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4709–4723, May 2021.

[13] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[14] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Dec. 2016.

[15] E. D. Carvalho, E. Bjornson, J. H. Sorensen, P. Popovski, and E. G. Larsson, "Random access protocols for massive MIMO," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 216–222, Jul. 2017.

[16] X. Jing, M. Li, H. Liu, S. Li, and G. Pan, "Superimposed pilot optimization design and channel estimation for multiuser massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11818–11832, Dec. 2018.

[17] L. A. Lago, Y. Zhang, N. Akbar, Z. Fei, N. Yang, and Z. He, "Pilot decontamination based on superimposed pilots assisted by time-multiplexed pilots in massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 405–417, Jan. 2020.

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

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

[20] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[21] L. S. Muppirisetty, H. Wymeersch, J. Karout, and G. Fodor, "Location-aided pilot contamination elimination for massive MIMO systems," in *Proc. IEEE Global Telecommun. Conf.*, 2015, pp. 1–5.

[22] P. Li, Y. Gao, Z. Li, and D. Yang, "User grouping and pilot allocation for spatially correlated massive MIMO systems," *IEEE Access*, vol. 6, pp. 47959–47968, 2018.

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

[24] H. T. Dao and S. Kim, "Vertex graph-coloring-based pilot assignment with location-based channel estimation for massive MIMO systems," *IEEE Access*, vol. 6, pp. 4599–4607, 2018.

[25] A. S. Al-Hubaishi, N. K. Noordin, A. Sali, S. Subramaniam, and A. M. Mansoor, "An efficient pilot assignment scheme for addressing pilot contamination in multicell massive MIMO systems," *Electronics*, vol. 8, no. 4, p. 372, 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/4/372

[26] L. You, X. Gao, X. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[27] M. Herdin, N. Czink, H. Ozcelik, and E. Bonek, "Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, 2005, pp. 136–140.

[28] T. H. Nguyen, T. V. Chien, H. Q. Ngo, X. N. Tran, and E. Björnson, "Pilot assignment for joint uplink–downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8292–8297, Aug. 2021.

[29] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[30] L. Ribeiro, M. Leinonen, H. Djelouat, and M. Juntti, "Channel charting for pilot reuse in mMTC with spatially correlated MIMO channels," in *Proc. IEEE Global Telecommun. Conf. Workshop*, 2020, pp. 1–6.

[31] L. Ribeiro, M. Leinonen, H. Al-Tous, O. Tirkkonen, and M. Juntti, "Exploiting spatial correlation for pilot reuse in single-cell mMTC," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2021, pp. 1–9.

[32] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47682–47698, 2018.

[33] L. Ribeiro, M. Leinonen, I. Rathnayaka, H. Al-Tous, and M. Juntti, "Channel charting aided pilot allocation in multi-cell massive MIMO mMTC networks," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, 2022, pp. 1–5.

[34] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017. doi: 10.1561/2000000093.

[35] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[36] H. Djelouat, M. Leinonen, L. Ribeiro, and M. Juntti, "Joint user identification and channel estimation via exploiting spatial channel covariance in mMTC," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 887–891, Apr. 2021.

[37] J. García-Morales, G. Femenias, and F. Riera-Palou, "Higher order Sectorization in FFR-aided OFDMA cellular networks: Spectral- and energy-efficiency," *IEEE Access*, vol. 7, pp. 11127–11139, 2019.

[38] H. Yu, L. You, W. Wang, and X. Yi, "Active channel sparsification for uplink massive MIMO with uniform planar array," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6018–6032, Sep. 2021.

[39] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[40] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[41] "Technical specification group radio access network; study on 3-D channel model for LTE," 3GPP, Sophia Antipolis, France, Rep. 3GPP TR 36.873 V12.7.0, Dec. 2017.

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

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

[44] A. W. J. Kolen and J. K. Lenstra, "Combinatorics in operations research," in *Handbook of Combinatorics*, R. L. Graham, M. Grötschel, and L. Lovász, Eds. Cambridge, MA, USA: MIT Press, 1996, pp. 1875–1910, ch. 35.

[45] T. Vredeveld and J. K. Lenstra, "On local search for the generalized graph coloring problem," *Oper. Res. Lett.*, vol. 31, no. 1, pp. 28–34, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167637702001657

[46] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is 'nearest neighbor' meaningful?" in *Database Theory—ICDT*, C. Beeri and P. Buneman, Eds. Berlin, Germany: Springer, 1999, pp. 217–235.

[47] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 5, no. 5, pp. 363–387, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.11161

[48] E. Lei, O. Castañeda, O. Tirkkonen, T. Goldstein, and C. Studer, "Siamese neural networks for wireless positioning and channel charting," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2019, pp. 200–207.

[49] P. Huang et al., "Improving channel charting with representation-constrained Autoencoders," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, 2019, pp. 1–5.

[50] P. Agostini, Z. Utkovski, and S. Stańczak, "Channel charting: An Euclidean distance matrix completion perspective," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5010–5014.

[51] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[52] C. K. I. Williams, "On a connection between kernel PCA and metric multidimensional scaling," *Mach. Learn.*, vol. 46, pp. 11–19, Jan. 2002.

[53] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. [Online]. Available: https://science.sciencemag.org/content/290/5500/2319

[54] L. Van Der Maaten et al., "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, nos. 66–71, p. 13, 2009.

[55] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

[56] R. W. Floyd, "Algorithm 97: Shortest path," *Commun. ACM*, vol. 5, no. 6, p. 345, Jun. 1962.

[57] J. Milton and J. Arnold, *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*. New York, NY, USA: McGraw-Hill, 2003.

[58] T. Ponnada, H. Al-Tous, O. Tirkkonen, and C. Studer, "An out-of-sample extension for wireless Multipoint channel charting," in *Proc. Cogn. Radio Orient. Wireless Netw.*, 2019, pp. 208–217.

[59] V. Silva and J. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in *Advanced Neural Information Processing System*, vol. 15, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA, USA: MIT Press, 2003. [Online]. Available: https://papers.nips.cc/paper/2002/file/5d6646aad9bcc0be55b2c82f69750387-Paper.pdf

[60] J. Tzeng, H. H.-S. Lu, and W.-H. Li, "Multidimensional scaling for large genomic data sets," *BMC Bioinf.*, vol. 9, no. 1, pp. 1–17, 2008.

[61] K. Pedersen, P. Mogensen, and B. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 437–447, Mar. 2000.

[62] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

LUCAS RIBEIRO (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Technology-Paraná, Brazil, in 2018 and 2019, respectively. He is currently pursuing the Ph.D. degree in communications engineering with the Centre for Wireless Communications, University of Oulu, Finland. His research interests include signal processing for wireless networks as well as massive machine-type communications.

MARKUS LEINONEN (Member, IEEE) received the B.Sc. (Tech.) and M.Sc. (Tech.) degrees in electrical engineering and the D.Sc. (Tech.) degree in communications engineering from the University of Oulu, Finland, in 2010, 2011, and 2018, respectively, where he joined the Centre for Wireless Communications in 2010. He received the Academy of Finland Postdoctoral Researcher position from 2021 to 2024. He was a Guest Researcher with the Technical University of Munich, Germany, in 2013, and a Visiting Postdoctoral Researcher with the University of California at San Diego in 2020. He has published over 50 journal and conference papers, a book, and a book chapter in the areas of wireless communications, mathematical optimization, and signal and information processing. His current research interests are time-critical networking and sparsity-aware signal processing for wireless communications. The research covers optimization and analysis of information freshness in status update systems and design of sparse signal recovery methods for channel estimation, user activity detection, and signal detection in wireless communication systems. He served as the Local Arrangements Chair of the IEEE SPAWC 2022. He was appointed as an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2020.

HANAN AL-TOUS (Senior Member, IEEE) received the B.Sc. degree in electrical engineering and the M.Sc. degree in communication engineering from the University of Jordan, Amman, Jordan, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering from United Arab Emirates University, Al Ain, UAE, in 2014. She worked as a Lecturer with the Electrical Engineering Department, Al-Ahliyya Amman University, Amman, from 2000 to 2010. She worked as a Postdoctoral Fellow with the United Arab Emirates University from 2014 to 2018. She is currently a Research Fellow with Aalto University, Finland. Her research interests include cooperative communications, energy harvested sensor networks, resource allocation for wireless communications, game theory, compressive sensing, and machine learning.

OLAV TIRKKONEN (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in theoretical physics from the Helsinki University of Technology in 1990 and 1994, respectively. He is Full Professor of Communication Theory with Aalto University, Finland, where he has been holding a faculty position since 2006. After postdoctoral positions at UBC, Vancouver, Canada, and NORDITA, Copenhagen, Denmark, from 1999 to 2010, he was with Nokia Research Center, Helsinki, Finland. From 2016 to 2017, he was Visiting Associate Professor with Cornell University, Ithaca, NY, USA. He has published some 300 papers and is the inventor of some 85 families of patents and patent applications which include 1% of all patents declared essential for the first standardized version of 4G LTE. His current research interests are in coding for random access and quantization, quantum computation, and machine learning for cellular networks. He served as the General Chair of 2022 IEEE International Symposium on Information Theory. He is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

MARKKU JUNTTI (Fellow, IEEE) received the M.Sc. and Dr.Sc. degrees in electrical engineering from the University of Oulu, Oulu, Finland, in 1993 and 1997, respectively.

He was with the University of Oulu from 1992 to 1998, a Visiting Scholar with Rice University, Houston, Texas, from 1994 to 1995, and a Senior Specialist with Nokia Networks, Oulu, from 1999 to 2000. Since 2000, he has been a Professor of Communications Engineering with the Centre for Wireless Communications (CWC), University of Oulu, where he leads the Communications Signal Processing Research Group. He also serves as the Head of CWC–Radio Technologies Research Unit. His research interests include signal processing for wireless networks as well as communication and information theory. He is an author or coauthor of almost 500 papers published in international journals and conference records as well as in books *Wideband CDMA for UMTS* (2000–2010), *Handbook of Signal Processing Systems* (2013–2018), and *5G Wireless Technologies* (2017). He is also an Adjunct Professor with the Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA.

Dr. Juntti has been a Secretary of the Technical Program Committee of the 2001 IEEE International Conference on Communications, and the Chair or Co-Chair of the Technical Program Committee of several conferences, including 2006 and 2021 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, the Signal Processing for Communications Symposium of IEEE Globecom 2014, the Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems of IEEE GlobalSIP 2016, the ACM NanoCom 2018, and the 2019 International Symposium on Wireless Communication Systems. He is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and served previously in similar role in IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was a Secretary of IEEE Communication Society Finland Chapter from 1996 to 1997, and the Chair from 2000 to 2001. He has also served as the General Chair of 2011 IEEE Communication Theory Workshop and the 2022 IEEE Workshop on Signal Processing Advances in Wireless Communications.