Exploiting Spatial Correlation for Pilot Reuse in Single-Cell mMTC

Lucas Ribeiro[†], Markus Leinonen[†], Hanan Al-Tous[⋆], Olav Tirkkonen[⋆], and Markku Juntti[†]

[†]Centre for Wireless Communications, FI-90014, University of Oulu, Finland

[⋆]Department of Communications and Networking, Aalto University, Finland

e-mail: {lucas.ribeiro, markus.leinonen, markku.juntti}@oulu.fi, {hanan.al-tous, olav.tirkkonen}@aalto.fi

Abstract—As a key enabler for massive machine-type communications (mMTC), spatial multiplexing relies on massive multiple-input multiple-output (mMIMO) technology to serve the massive number of user equipments (UEs). To exploit spatial multiplexing, accurate channel estimation through pilot signals is needed. In mMTC systems, it is impractical to allocate a unique orthogonal pilot sequence to each UE as it would require too long pilot sequences, degrading the spectral efficiency. This work addresses the design of channel features from correlated fading channels to assist the pilot assignment in multi-sector mMTC systems under pilot reuse of orthogonal sequences. In order to reduce pilot collisions and to enable pilot reuse, we propose to extract features from the channel covariance matrices that reflect the level of orthogonality between the UEs channels. Two features are investigated: *covariance matrix distance* (CMD) feature and CMD-aided *channel charting* (CC) feature. In terms of symbol error rate and achievable rate, the CC-based feature shows superior performance than the CMD-based feature and baseline pilot assignment algorithms.

I. INTRODUCTION

Massive machine-type communications (mMTC) is a class of services planned for 5G and 6G systems to provide access for enormous numbers of connected devices, possibly on the order of tens of billions [1]–[3]. Massive multiple-input multiple-output (mMIMO) technology has been proposed as a way to improve the spectral and energy efficiency of such systems by exploiting the spatial multiplexing provided by the large numbers of antennas [4], [5]. To take advantage of the boost in spectral and energy efficiency provided by large-scale antenna systems, channel state information (CSI) is required at the base station (BS). Uplink CSI is obtained through the transmission of pilot sequences from the user equipments (UEs) to the BS, which estimates the channel using the received signal by employing any estimation technique such as least squares or minimum mean square error. A time-division duplexing (TDD) protocol is likely to be implemented in mMIMO networks in order to minimize signaling overhead and to avoid downlink channel estimation by exploiting the reciprocity between the uplink and downlink channels [6].

One of the main challenges of mMTC is the design of access protocols that support the huge number of MTC devices. Due to the large number of UEs, resources have to be shared, which precludes the allocation of orthogonal pilot sequences to all UEs. *Pilot reuse* can be used to overcome the lack of resources, at the cost of leading to interference between UEs sharing the same pilot sequence, also known as *pilot contamination* [7].

While legacy systems rely on code and frequency multiplexing, mMTC can exploit spatial multiplexing through mMIMO to avoid/alleviate pilot contamination [2], [8].

Different approaches to mitigate pilot contamination have been proposed in the literature. In [9], the authors proposed a location-aware pilot assignment for multi-cell systems with Rician fading channels which requires knowledge about the large-scale fading, angle of arrival (AoA), and the channel Rician factor. In [6], a pilot assignment scheme was devised to deal with pilot interference in multi-cell mMIMO based on the UEs' large-scale coefficients, which are assumed to be known in all cells. Their objective is to assign orthogonal pilot sequences to UEs with poor channel quality. You *et al.* [10] proposed a pilot reuse strategy based on the channel covariance matrices for single-cell mMIMO systems with correlated channels. They proposed a pilot allocation algorithm that assigns orthogonal pilot sequences to UEs with similar channel covariance matrices, measured by the covariance matrix distance (CMD) metric.

Recently, we proposed a pilot reuse scheme in [11] that utilizes *channel charting* (CC) to exploit the spatial information existing in CSI, aiming to maximize the AoA distances between the UEs using the same pilot sequence. CC is a framework proposed in [12] to estimate the relative position of devices in an unsupervised manner which 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. Our results [11] showed that CSI can be utilized to allocate the pilot sequences, and improve channel estimation accuracy and symbol error rate (SER), for single-cell correlated channels systems.

In this paper, we propose pilot reuse strategies to cope with the pilot contamination in multi-sector single-cell mMIMO networks under spatially correlated channels. To exploit the degree of orthogonality between the UEs' channels, granted by the spatial correlation, the pilot reuse is performed based on two features encompassing the channel correlation, especially in the angular domain: 1) the CMD metric and 2) the CC created based on the CMD metric. Utilizing the proposed features along with the pilot allocation algorithm devised in [11], the numerical results show that the angular information can be exploited also for multi-sector systems with spatially correlated channels, resulting in enhanced achievable rate and SER compared to baselines.
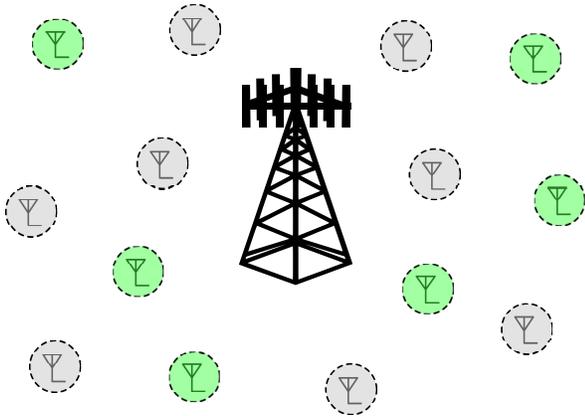
Fig. 1: Uplink mMTC scenario with K active (green) and $N - K$ inactive users (gray). The BS is equipped with $S = 3$ M-element ULAs.

II. SYSTEM MODEL

We consider an 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 system performance in mMTC, we assume that the BS knows the set of active UEs at each transmission instant[1].

The cell is divided into S sectors[2] and the transmitted signal is received through S uniform linear arrays (ULAs), each having M antenna elements, as shown in Fig. 1. The array response vector for a ULA is given by

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

where Δ_{r} is the normalized spacing between the antenna elements in wavelengths and θ is the AoA, i.e., the incident angle of the received signal on the ULA [17, Sec. 7.2.1].

We adopt the one-ring channel model [7, p. 236] which assumes that the multi-path components are concentrated around the UEs while the BS is located in an elevated position lacking scatterers close to it. Thus, the uplink channel vector for user $n \in \mathcal{N}$ associated with the ULA of sector $s \in \mathcal{S} = \{1, \ldots, S\}$ is modelled as a superposition of L propagation paths as

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

where $\alpha_{n,s,l}$ is the complex gain of the lth path assumed to be an independent and identically distributed (i.i.d.) complex Gaussian random variable with zero mean and $\mathbb{E}\{|\alpha_{n,s,l}|^2\} = 1$. In (2), the large-scale propagation effects and the BS antenna gain in the channel are captured in $\beta_{n,s,l} \in \mathbb{R}$, which

[1]This assumption can be invoked by the fact that there exists a multitude of compressed sensing based approaches, e.g., [13]–[15], to solve the UE activity detection problem. However, this is outside of the scope of this paper and is left for future work.

[2]In practical scenarios, the coverage area is usually divided into three or six sectors of $120°$ or $60°$ [16]. Therefore, we consider $S = 3$ in the examples.
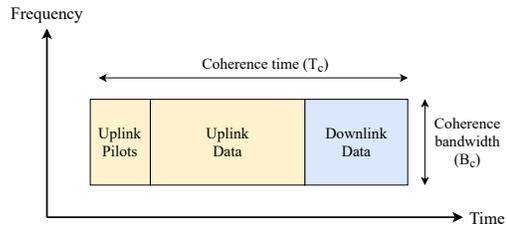
Fig. 2: The channel response is time-invariant and frequency-flat within each coherence block. At the beginning of each coherence block, τ symbols are sent to estimate the channel.

follows the free space path loss model as described in [18, Eq. (2.7)], which is defined as

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

where λ is the wavelength and d_n is the distance between UE n and the BS. In (3), $G_{\mathrm{A}}(\theta_{n,s,l}) \in \mathbb{R}$ represents the antenna gain through the lth path for user n at the ULA s, and is given in dB by [16], [19, Table 7.1-1]

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

where $G_{\mathrm{A}_{\max}}$ is the maximum antenna gain, θ_{3dB} is the half power beamwidth, $A_{\max}$ is the maximum attenuation in dB at the ULAs, and $\theta_{n,s,l}$ is the AoA of the lth path which is modelled as an i.i.d. random variable 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 user 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 incoming multi-path components arriving from user n at the ULA of sector s.

III. PILOT TRANSMISSION AND CHANNEL ESTIMATION

To exploit the symmetry between downlink and uplink channels, we consider the TDD protocol, as depicted in Fig. 2. We assume that during one coherence block, the channels are time-invariant and flat-fading. At each coherence block, the set of active users transmit τ known symbols, of equal power p_{u}, 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.

We consider that 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 consider that $\tau \ll N$, i.e., global *pilot reuse* is employed in the cell area. Consequently, in the channel estimation phase, the same pilot is shared by K/τ UEs on the average. 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 available pilot sequences. User $k \in \mathcal{K}$ transmits a pilot signal $\boldsymbol{\psi}_k = \sqrt{p_{\mathrm{u}}}\boldsymbol{\phi}_{\pi_k}$, where $\pi_k \in \mathcal{T}$ is the index of the pilot sequence assigned to UE k and $\boldsymbol{\phi}_{\pi_k} \in \mathbb{C}^\tau$ is the

corresponding pilot sequence from the orthogonal pilot book $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\tau] \in \mathbb{C}^{\tau \times \tau}$. We define the set of UEs sharing the same pilot sequence as UE k, including UE k itself, as $\mathcal{G}_k = \{j \mid j \in \mathcal{N}, \pi_j = \pi_k\}$.

The received signal for channel estimation at the ULA of sector s, $\mathbf{Y}_s = [\mathbf{y}_s^1, \ldots, \mathbf{y}_s^\tau] \in \mathbb{C}^{M \times \tau}$, can be written as

$$\mathbf{Y}_s = \mathbf{H}_s \check{\mathbf{\Psi}} + \mathbf{N}_s, \tag{5}$$

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

We consider that the linear minimum mean square error (LMMSE) receiver is deployed at the BS to jointly estimate the active UEs' channel vectors across the S sectors. Therefore, the compound channel $\mathbf{h}_k \in \mathbb{C}^{MS}$ between UE $k \in \mathcal{K}$ and the BS is given by

$$\mathbf{h}_k = \begin{bmatrix} \mathbf{h}_{k,1} \\ \vdots \\ \mathbf{h}_{k,S} \end{bmatrix}, \tag{6}$$

with covariance matrix $\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^{\mathrm{H}}] \in \mathbb{C}^{MS \times MS}$. The received pilot signal, $\mathbf{Y} = [\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_S^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{MS \times \tau}$.

We assume that the channel covariance matrices of all UEs (active and inactive ones) are known at the BS. In order to be able to retrieve the UEs' individual signals from the received compound signal, the BS needs to keep track of $\mathbf{R}_n, \forall n \in \mathcal{N}$. In practice, an initial training phase is required to obtain the first estimate for the covariance matrices. After this initial phase, the BS can keep updating each $\mathbf{R}_n$ based on the estimated channel.

The LMMSE estimate of the communications channel between user k and the BS, $\mathbf{h}_k$ in (6), is given as [10],

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

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

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

where $\mathcal{I}_k = \{j \mid j \neq k, j \in \mathcal{G}_k \cap \mathcal{K}\}$ is the set of interfering users to user k, and $\mathbf{Q}_k \in \mathbb{C}^{MS \times MS}$ in (7) is the covariance matrix of the received signal, given as

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

Due to the orthogonality principle of the MMSE estimator [20, Sec. 12.4], the channel estimation error, $\tilde{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\tilde{\mathbf{h}}_k})$, is independent of $\hat{\mathbf{h}}_k$. Therefore, we can de-

compose the channel $\mathbf{h}_k$ as $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$. Thus, the error covariance matrix for user k is [12]

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

IV. UPLINK DATA TRANSMISSION

Let $\mathbf{x} = [x_1, \ldots, x_K]^{\mathrm{T}} \in \mathbb{C}^K$ be the transmitted symbol vector at a given time instant. The corresponding received signal at the BS, $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_s^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{MS}$, is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{MS \times K}$ is the channel matrix, and $\mathbf{n} = [\mathbf{n}_1^{\mathrm{T}}, \ldots, \mathbf{n}_s^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{MS}$ is the noise vector at the $M \times S$ receiver antenna elements. We assume that all UEs transmit with the same power, i.e., $p_{\mathrm{u}} = |x_k|^2$ is the transmit symbol power for user k.

Given the estimated channel $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K] \in \mathbb{C}^{MS \times K}$, we use the LMMSE receiver, $\mathbf{w}_k \in \mathbb{C}^{MS}$, derived in [10],

$$\mathbf{w}_k = \left(\hat{\mathbf{H}} \hat{\mathbf{H}}^{\mathrm{H}} + \sum_{k=1}^K \mathbf{R}_{\tilde{\mathbf{h}}_k} + \frac{\sigma_{\mathrm{n}}^2}{p_u} \mathbf{I}_{MS} \right)^{-1} \hat{\mathbf{h}}_\mathbf{k}, \tag{12}$$

which takes into account the error covariance matrix $\mathbf{R}_{\tilde{\mathbf{h}}_k}$.

Therefore, the received symbol vector $\hat{\mathbf{r}} \in \mathbb{C}^{MS}$, after employing the LMMSE receiver $\mathbf{W} \in \mathbb{C}^{MS \times K}$, is given by

$$\hat{\mathbf{r}} = \mathbf{W}^{\mathrm{H}} \mathbf{y}. \tag{13}$$

Given the communication model depicted in Fig. 2, the uplink spectral efficiency is lower bounded by [7, Th. 4.1]

$$R_k^{\mathrm{up}} = \left(1 - \frac{\tau}{T_c} \right) R_k^{\mathrm{ach,up}}, \tag{14}$$

where T_c is the coherence time and the corresponding achievable uplink rate for user k, $R_k^{\mathrm{ach,up}}$, is expressed as

$$R_k^{\mathrm{ach,up}} = \mathbb{E} \left\{ \log_2 \left(1 + \gamma_k^{\mathrm{up}} \right) \right\}, \tag{15}$$

where the expectation is taken over the channel realizations, and γ_k^{up} is the *instantaneous* uplink signal-to-interference-and-noise ratio (SINR), given as

$$\gamma_k^{\mathrm{up}} = \frac{|\mathbf{w}_k^{\mathrm{H}} \hat{\mathbf{h}}_k|^2}{\mathbf{w}_k^{\mathrm{H}} \left(\sum_{j \neq k} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^{\mathrm{H}} + \sum_{n=1}^K \mathbf{R}_{\tilde{\mathbf{h}}_n} + \frac{\sigma_{\mathrm{n}}^2}{p_u} \mathbf{I}_{MS} \right) \mathbf{w}_k}. \tag{16}$$

V. PILOT REUSE ALGORITHM

The coordination of pilot assignment is crucial for mMTC. Due to the massive amount of UEs in such systems, the reuse of the orthogonal pilot sequences in the same cell becomes inevitable. As highlighted by Björnson in [7, p. 246], the strongest interference usually originates from UEs in the same cell. Therefore, a proper assignment of the pilots is essential to mitigate the pilot contamination.

Next, we tackle the pilot contamination problem by proposing pilot reuse strategies that exploit the spatial information present in CSI through the second-order statistics of the radio environment. In particular, we propose two features encompassing the channel correlation, especially in the angular domain, for the pilot reuse: 1) the CMD metric and 2) the CC

created based on the CMD metric. These features are fed to a modified version of our greedy pilot allocation algorithm proposed in [11] (Algorithm 1 detailed in Sec. V-C), which completes the pilot assignment, aiming at minimizing the pilot interference between the users sharing the same pilot.

The goal is to assign orthogonal pilot sequences to users with overlapping AoA intervals to avoid pilot contamination. To capture the spatial information embedded in the second-order statistics and reveal the angular domain relationship among the UEs, we use the CMD [21] metric. Thus, we compute a dissimilarity matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$ and define the dissimilarity measure associated with UE n and j as

$$d_{n,j} = 1 - \frac{\operatorname{tr}\left(\mathbf{R}_n^{\mathrm{H}} \mathbf{R}_j\right)}{\|\mathbf{R}_n\|_{\mathrm{F}} \|\mathbf{R}_j\|_{\mathrm{F}}}. \tag{17}$$

Next, we elaborate on two different features that utilize the CMD metric in (17) as a basis. The features will serve as input to our *Nearest Neighbor Pilot Assignment* Algorithm 1.

A. CMD-aided Pilot Assignment

Since the dissimilarity matrix $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_N]$ carries the spatial information needed to allocate the pilot sequences, we propose to directly utilize the distance vectors, $\mathbf{d}_n = [d_{n,1}, \ldots, d_{n,N}]^{\mathrm{T}} \in \mathbb{R}^N$, as input features, $\mathbf{f}_n \in \mathbb{R}^N$, to the proposed pilot allocation algorithm as detailed in V-C. Thus, we form $\mathbf{f}_n$ by setting $\mathbf{f}_n = \mathbf{d}_n$. After getting $\mathbf{f}_n$ we apply the Algorithm 1 to allocate the pilots.

B. CC-aided Pilot Assignment

CC is as a framework developed in [12] to generate unsupervised radio environment mappings utilizing the CSI. The idea behind CC is to find a suitable feature from CSI and then apply a dimensionality reduction (DR) technique to get a lower dimensional embedding, which preserves the relative position of the UEs. Given the CMD distance in (17), we propose to utilize it as an input to CC framework. The target is to obtain a feature that well characterizes the angular domain structures in the multi-sector multi-user scenario, in order to feed the pilot allocation Algorithm 1.

As shown in [11], for single-cell scenarios, CC can be used to coordinate the pilot assignment under pilot reuse regimes, to achieve better network performance than other methods from the literature. In [11], we have utilized CC to retrieve the angular relationship between the UEs by applying the discrete Fourier transform on the channel covariance matrices and then taking the absolute value. Here, we differently propose to use the CMD dissimilarity matrix $\mathbf{D}$ in (17) as an input to the DR technique, to construct the CC. Thus, we apply a function $\mathcal{C}$ that generates CC by mapping these high-dimensional features $\mathbf{d}_n \in \mathbb{R}^N$ to a low-dimensional domain, i.e.,

$$\mathcal{C} : \mathbf{d}_n \mapsto \mathbf{f}_n, \quad n \in \mathcal{N}, \tag{18}$$

where $\mathbf{f}_n \in \mathbb{R}^C$ is the point in the C-dimensional CC corresponding to $\mathbf{d}_n$, where typically $C = 2$ or $C = 3$.

Several unsupervised DR techniques have been proposed to map the extracted features into a lower dimension embedding [12], [22]–[24]. We propose the use of Laplacian Eigenmaps

---

Algorithm 1: Nearest neighbor pilot assignment

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 $\mathbf{f}_n$, $n \in \mathcal{N}$.

Output: A pilot assignment $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N]^{\mathrm{T}}$.

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., $\boldsymbol{\psi}_n = \phi_p$, that satisfies $n = \arg\min_{n \in \mathcal{N}^{\mathrm{un}}} \|\mathbf{f}_n - \mathbf{f}_{n'}\|^2$.

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**

(LE) as a DR technique as it aims to preserve the local structure of the high-dimensional embedding by minimizing the distances between data points and its ν nearest neighbors, where ν is a design parameter. Note that the CC method is not limited to utilize LE, and different DR techniques could be used as well. Clearly, as the low-dimensional embedding is not unique, they may lead to different pilot assignments, thus different network performance.

C. Nearest Neighbor Pilot Assignment Algorithm

After extracting the CMD and CMD + CC features, we apply a modified version of the low-complexity pilot algorithm proposed in [11] to account for the different input features and incorporate the inactive users as well, as shown in Algorithm 1. The main premise of the algorithm is that the features preserve the angular relationship among UEs. Thus, the greedy algorithm allocates the orthogonal pilot sequences in an *ordered* way $\phi_1, \phi_2, \ldots, \phi_\tau, \phi_1, \phi_2, \ldots$ by finding the unassigned UE with the small Euclidean distance between the features, aiming at maximizing the distances between the same pilot sequences.

The first step of Algorithm 1 is to allocate the first pilot sequence, ϕ_1, to one of the users. Then, it greedily allocates the next orthogonal pilot sequence, ϕ_2, to the unassigned UE with the smallest distance to the previously allocated user n', i.e., it finds n that minimizes $\|\mathbf{f}_n - \mathbf{f}_{n'}\|^2$. Then, it repeats this process until all orthogonal pilot sequences have been allocated. Once the last orthogonal pilot sequence ϕ_τ has been allocated, after τ iterations, it starts to reuse pilot sequences by allocating the first sequence ϕ_1 to the closest unassigned
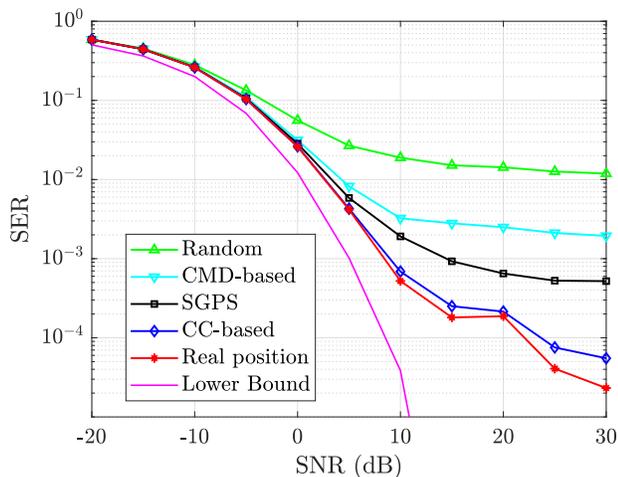
(a) SER as a function of SNR.
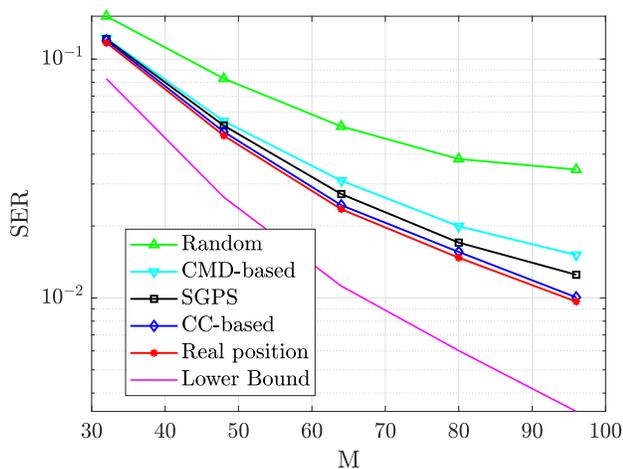
(b) SER as a function of the number of antennas per ULA M.

Fig. 3: Symbol error rate (SER) for a fixed pilot length of $\tau = 64$. (a) shows the effect of the SNR and (b) the effect of the number of antennas for a fixed SNR=0 dB.

(a) CDF.

(b) Achievable sum rate.

Fig. 4: Achievable sum rate at 10 dB SNR. (a) CDF of users' uplink achievable rate for $\tau = 64$. (b) Achievable sum rate vs. pilot length.

UE from the precedent allocated user. It repeats this process until all UEs have been assigned pilot sequences.

VI. NUMERICAL AND SIMULATION RESULTS

We consider $N = 512$ UEs uniformly distributed in a cell with $K = 64$ active UEs at any given time, and a BS equipped with $S = 3$ M-element ULAs, with $M = 64$ critically spaced, $\Delta_r = 0.5$, elements. The propagation channel between each user and the BS consists of $L = 200$ paths with a fixed angular standard deviation $\sigma_\theta = 15°$. For the antenna parameters in (4), we set the maximum antenna gain as $G_{A_{max}} = 0$ dB, the maximum attenuation as $A_{max} = 30$ dB, and the half-power beamwidth as $\theta_{3dB} = 65°$ [19]. We use binary phase shift keying (BPSK) for the channel estimation and quadrature phase shift keying (QPSK) for the data transmission.

We consider three baseline pilot assignment methods: 1) A random pilot assignment scheme; 2) A real position method, which relies on the exact UEs' positions to compute the angular separation between the UEs, and then deploys Algorithm 1.

Note that this baseline acts as a lower bound to the CC-based method, since it uses the real position instead of the relative position estimated via CC; 3) The statistical greedy pilot scheduling (SGPS) developed in [10] that, similarly to our method, relies on the knowledge of the channel covariance matrices. Differently from [10], we extract designed features from the channel covariance matrices and then apply Algorithm 1; this also obviates the need to store the channel covariance matrices at the BS.

Fig. 3 shows how SER of the different pilot assignment strategies changes with the SNR and the number of antennas M. A fixed pilot length of $\tau = 64$ is considered, which gives a pilot reuse factor of $N/\tau = 8$, meaning that each pilot sequence is reused 8 times across all the UEs connected to the BS. For benchmark, we also show the lower bound for SER, which is achieved with perfect CSI knowledge. From Fig. 3(a) we see that the performance of the CC-based pilot allocation approaches the real position method, beating all the

other baseline methods. The CMD-based approach, although implementing the same pilot allocation algorithm, shows poor performance as compared to CC-based, which highlights the importance of choosing a good feature to the pilot allocation Algorithm 1. Fig. 3(b) presents the effect of changing the number of ULA elements M for SNR=0 dB. By increasing the number of antennas, the performance increases because the UEs' channels become more orthogonal.

We also evaluate the performance of the proposed features in terms of the uplink achievable rate. The cumulative distribution function (CDF) for the average uplink rate per UE is depicted in Fig. 4(a) for a fixed pilot length of $\tau = 64$, at 10 dB SNR. It can be seen that, on average, at least 90 % of the UEs using the CMD-based pilot allocation strategy have a higher uplink achievable rate than the UE with the highest rate employing the random pilot allocation. On the other hand, this percentage grows to 100 % when the UEs are employing CC-based or SGPS schemes.

Fig. 4(b) plots the uplink sum rate against the pilot length. One can notice that the CC-based method performs very well, approaching the real position based pilot allocation. However, it slightly degrades its performance for $\tau > 100$ symbols, i.e., when the pilot reuse factor N/τ decreases below 5. Yet, in this regime, the SGPS method performs as well as the real position one. Although the CMD-based method greatly improves the performance as compared to the random pilot allocation, it underperforms when compared to the other methods.

VII. Conclusions

This paper investigated the reuse of orthogonal pilot sequences in a single-cell multi-sector network for mMIMO systems. In these systems, the transmission rate performance is interference limited due to pilot contamination. We proposed to apply new features capturing the angular information present in CSI in the algorithm developed in [11] to accommodate the multi-sector framework and to improve the average uplink rate for the UEs. The CC-based feature was shown to approach the real position method, outperforming the other baselines in terms of SER and achievable uplink rate. The CMD-based feature performs close to SGPS and CC-based pilot allocation, whilst having lower complexity. The proposed features showed to fit well into the multi-sector framework by presenting a competitive performance for SER and achievable rate as compared to existing pilot allocation schemes.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] 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, 2017.

[3] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: physical and mac-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, 2016.

[4] 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, 2014.

[5] 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, 2016.

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

[7] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Proc.*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[8] 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, 2020.

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

[10] 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, 2015.

[11] 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. Worksh.*, 2020, pp. 1–6.

[12] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel Charting: Locating users within the radio environment using channel state information," *IEEE Acc.*, vol. 6, pp. 47 682–47 698, 2018.

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

[14] 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, 2018.

[15] 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.*, pp. 1–1, 2021.

[16] J. García-Morales, G. Femenias, and F. Riera-Palou, "Higher order sectorization in FFR-aided OFDMA cellular networks: Spectral- and energy-efficiency," *IEEE Acc.*, vol. 7, pp. 11 127–11 139, 2019.

[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[18] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[19] 3GPP, "Technical specification group radio access network; study on 3D channel model for LTE," V12.7.0.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall International, 1993.

[21] 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 Vol. 1.

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

[23] P. Huang, O. Castañeda, E. Gönültaş, S. Medjkouh, O. Tirkkonen, T. Goldstein, and C. Studer, "Improving channel charting with representation-constrained autoencoders," in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.*, 2019, pp. 1–5.

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