

Massive MIMO Pilot Decontamination and Channel Interpolation via Wideband Sparse Channel Estimation

Saeid Haghighatshoar¹, Member, IEEE, and Giuseppe Caire², Fellow, IEEE

Abstract—We consider a massive MIMO system based on *time division duplexing* (TDD) and channel reciprocity, where the base stations (BSs) learn the channel vectors of their users via the pilots transmitted by the users in the *uplink* (UL). It is well-known that, in the limit of very large number of BS antennas, the system performance is limited by pilot contamination, due to the fact that the same set of orthogonal pilots is reused in multiple cells. In the regime of moderately large number of antennas, another source of degradation is channel interpolation because the pilot signal of each user probes only a limited number of *orthogonal frequency division multiplexing* (OFDM) subcarriers, and the channel must be interpolated over the other subcarriers, where no pilot symbol is transmitted. In this paper, we propose a low-complexity algorithm that uses the received UL wideband pilot snapshots in an observation window comprising several coherence blocks (CBs) to obtain an estimate of the *angle-delay power spread function* (PSF) of the received signal. This is generally given by the sum of the angle-delay PSF of the desired user and the angle-delay PSFs of the copilot users, i.e., the users re-using the same pilot dimensions in other cells/sectors. We propose supervised and unsupervised clustering algorithms to decompose the estimated PSF and isolate the part corresponding to the desired user only. We use this decomposition to obtain an estimate of the covariance matrix of the user wideband channel vector, which we exploit to decontaminate the desired user channel estimate by applying *minimum mean squared error* (MMSE) smoothing filter, i.e., the optimal channel interpolator in the MMSE sense. We also propose an effective low-complexity approximation/implementation of this smoothing filter. We use numerical simulations to assess the performance of our proposed method, and compare it with other recently proposed schemes that use the same idea of separability of users in the angle-delay domain.

Index Terms—Pilot decontamination, wideband channel model, subspace estimation, clustering, MMSE channel interpolation.

I. INTRODUCTION

CONSIDER a massive MIMO multi-cell system with M antenna per each base station (BS), per-cell processing, *Orthogonal Frequency Division Multiplexing* (OFDM),

Time Division Duplexing (TDD), and reciprocity-based channel estimation as in [2]–[4]. In such systems, time is divided into several slots, where in each slot users are scheduled to send *uplink* (UL) pilot signals in order to allow the BS to estimate their channel vectors. The BS exploits the UL-DL reciprocity and uses the resulting channel estimates to coherently detect data from the users in the UL and precode data to the users in the DL. A family of mutually orthogonal pilot sequences are obtained in the time-frequency domain by assigning to each pilot a different set of signal dimensions in the tessellation of the time-frequency plane under the OFDM [2] or, more in general, by sharing all the signal dimensions but assigning to the pilots mutually orthogonal symbol sequences across all signal dimensions (e.g., [5]). Due to limited channel coherence time, the signal dimensions in each UL-DL scheduling slot are limited. Consequently, also the number of UL pilot signal dimensions is limited, resulting in a limited number of orthogonal pilots. Therefore, to simultaneously serve several users across the whole system, pilots must be reused in multiple cells according to a specific reuse pattern [2]. As a result, the channel estimation during the UL pilot transmission is severely degraded by the interference received from the users in neighboring cells (or sectors) re-using the same pilot sequences as the users inside the cell; these users are referred to as copilot users (CPUs). Such a phenomenon is called *pilot contamination*. It is well-known that pilot contamination becomes the only limiting factor on the spectral efficiency of the system in the asymptotic limit where the number of BS antennas $M \rightarrow \infty$ but the number of users per cell K is kept finite [2], [6]. In the more realistic case of large but finite M and K with $M \gg K$, the pilot contamination still represents an important source of degradation especially for the edge users lying on the cell boundary [7]–[9].

A. Approaches to Pilot Decontamination

Several approaches have been proposed to cope with pilot contamination. In [10], it is observed that if multipath components (MPCs) of the channel vectors of the users have a limited angular spread (spatial correlation), it is possible to coordinate the pilot transmission in adjacent cells such that the channels of CPUs are confined in nearly orthogonal subspaces due to their angular diversity. However, in order to effectively separate CPUs, the covariance information

Manuscript received February 28, 2017; revised June 25, 2017 and August 18, 2017; accepted September 25, 2017. Date of publication October 13, 2017; date of current version December 8, 2017. This paper was presented at the 50th Annual Asilomar Conference on Signals, Systems, and Computers, November 2016 [1]. The associate editor coordinating the review of this paper and approving it for publication was J. Lee. (Corresponding author: Saeid Haghighatshoar.)

The authors are with the Communications and Information Theory Group, Technische Universität Berlin, 10623 Berlin, Germany (e-mail: saeid.haghighatshoar@tu-berlin.de, caire@tu-berlin.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2017.2760825

1536-1276 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

(or subspace information), i.e., the second-order statistics of users' channel vectors, must be known at the BS. A similar a priori statistical knowledge is used in [11] and [12] in the so-called JSDM scheme to reuse pilots in the same cell in order to decrease the pilot dimension overhead. More generally, it has been shown in [13] that if the covariance matrices of the users and their CPUs are available at the BS and satisfy certain mild conditions of linear independence, pilot contamination in the limit of $M \rightarrow \infty$ can be completely eliminated. However, this requires the knowledge of the user channel covariance matrices, which is itself difficult to obtain precisely due to pilot contamination.

A quite different approach is proposed in [14], in which no a priori knowledge of subspace is needed. Instead, it is noticed that when the number of BS antennas M is much larger than the number of per-cell served users K , and the power imbalance between the desired and the interfering users is above a certain threshold, the eigenvalues of the sample covariance matrix of the received signal corresponding to the desired users and that corresponding to CPUs in adjacent cells concentrate on "clusters" with disjoint supports. Thus, by distinguishing those clusters, it is possible to identify blindly the desired and the interfering signal subspaces. In contrast to [10]–[12], which work only in the presence of spatially correlated channels, the method of [14] would work also with i.i.d. (isotropically distributed) channel vectors, provided that the power imbalance between the desired and the interfering users is sufficiently large and the matrix dimension is large enough such that the eigenvalue clustering is sufficiently sharp. A combination of techniques in [10] and [14], via exploiting both the spatial correlation and power discrimination, has been used in [15].

Another method to cope with pilot contamination consists in "pilot contamination precoding" as proposed in [16]. The main idea is that, due to very large number of BS antennas, the only residual interference that matters after beamforming is the coherent interference due to pilot contamination, which can be eliminated by jointly precoding across neighboring cells (e.g., in the DL using linear precoding or non-linear dirty-paper coding and in the UL using linear interference mitigation or non-linear successive interference cancellation). Such a scheme, however, requires centralized processing of multiple cell sites in order to jointly decode/precoding the UL/DL signals; this goes against the beauty and simplicity of massive MIMO, for which single-cell processing is one of the main motivations [2].

In the recent work [17], developed independently and in parallel with our present work, a method for pilot decontamination is proposed by exploiting the fact that the channel vectors of CPUs at a given BS have typically different MPCs in the angle-delay domain. Therefore, if it is possible to identify the MPCs pertaining only to the desired user, the interference due to CPUs can be mitigated by linear space-frequency filtering, thus, mitigating the effect of pilot contamination. Our work is also based on the same idea but differs from [17] in many aspects and generally can achieve much better performance without incurring any additional pilot overhead with respect to the standard pilot schemes used in current systems

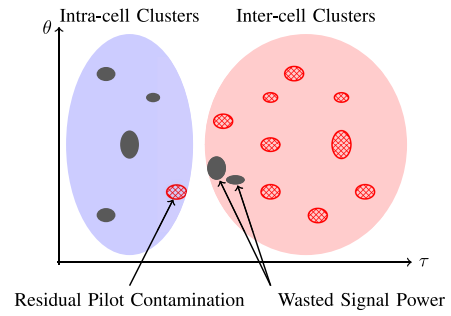


Fig. 1. Illustration of the proposed pilot decontamination scheme via exploiting the sparsity of the angle-delay scattering map of the received signal at the BS. The multipath components (MPCs) of the intended user and those of its CPUs are illustrated with gray solid and red dashed bubbles respectively. In this example, most of the MPCs of the user have shorter delays, thus, they can be identified and fairly separated from those of the CPUs via clustering in the delay domain as illustrated in the figure. The red dashed bubbles inside the signal cluster illustrate the residual copilot interference whereas the gray solid bubbles inside the interference cluster correspond to the useful signal wasted because of the delay-based clustering. It is seen that clustering eliminates the interference due to CPUs at the cost of filtering out also some components of the useful signal power (this should be optimized for designing an efficient clustering algorithm). The ultimate efficiency of the clustering for eliminating the copilot interference depends on how much MPCs of the user and its CPUs are separated in the delay domain.

(e.g., in LTE-TDD [18]). We defer a thorough comparison of [17] with our work to Section VII-H.

B. Contribution

In this paper, we pursue a new method for pilot decontamination that has the following advantages:

- Unlike [10], [13], and [15], we do not assume a priori knowledge of the channel covariance matrices or centralized coordination of pilot allocation.
- Unlike [14] and [15], we do not rely on asymptotic results in random matrix theory, which requires a) i.i.d. isotropic channel vectors (spatially correlated channel vectors along with covariance information in [15]), and b) sufficiently large power imbalance between the users inside and outside the cell.
- Unlike [16], we do not rely on joint precoding and centralized processing. Instead, we apply a strictly uncoordinated per-BS processing.

Here, we only provide an intuitive explanation of our proposed scheme and postpone the thorough description to Section V. The idea is qualitatively illustrated in Fig. 1. In a massive MIMO macrocell system, the propagation between users and BS antennas occurs through relatively *sparse* MPCs in the angle-delay domain. We exploit this underlying sparsity to estimate the *angle-delay Power Spread Function* (PSF) of each user by sampling only a small number of antennas and sending UL pilots only over a small subset of subcarriers. Then, we apply suitable algorithms to cluster the estimated PSF in the angle-delay plane to approximately separate the MPCs belonging to the desired user from those of its CPUs. This is illustrated qualitatively in Fig. 1 for a configuration where clustering is based on the difference of propagation delays, and where the interference due to CPUs can be fairly eliminated by filtering in the delay domain, at the cost of possibly filtering

out also some components of the useful signal. Furthermore, once we identified the angle-delay domain clusters pertaining to the desired user's PSF, we exploit them to obtain a very compact representation of the user wideband covariance matrix over the whole set of OFDM subcarriers. In turns, we use this information for MMSE channel estimation, obtaining at once both decontamination (i.e., the contribution of the CPUs is filtered out by the channel estimator) and channel interpolation over the whole signal bandwidth. We develop a novel computationally efficient channel interpolation method that approximates the *Minimum Mean Squared Error* (MMSE) smoothing filter. This provides a close-to-optimal MSE channel estimator under the Gaussian statistics and avoids performance degradation incurred due to imperfect instantaneous channel estimation, especially for a moderate number of antennas M [19].

C. Notation

We represent scalar constants by non-boldface letters (e.g., x or X), sets by calligraphic letters (e.g., \mathcal{X}), vectors by boldface small letters (e.g., \mathbf{x}), and matrices by boldface capital letters (e.g., \mathbf{X}). We denote the i -th row and the j -th column of a matrix \mathbf{X} with the row-vector $\mathbf{X}_{i,:}$ and the column-vector $\mathbf{X}_{:,j}$ respectively. For a $p \times q$ matrix \mathbf{X} , we represent by $\text{vec}(\mathbf{X})$ the $pq \times 1$ column-vector obtained by stacking the column of \mathbf{X} on top of each other, where we denote the resulting vector with a blackboard letter \mathbb{X} and a matrix consisting of r such vectors by $\mathbb{X} = [\mathbb{X}_1, \dots, \mathbb{X}_r]$. We indicate the Hermitian conjugate and the transpose of a matrix \mathbf{X} by \mathbf{X}^H and \mathbf{X}^T with the same notation being used for vectors and scalars. $\mathbf{X} \otimes \mathbf{Y}$ indicates the Kronecker product of the matrices \mathbf{X} and \mathbf{Y} . We denote the complex and the real inner product between two matrices (and similarly two vectors) \mathbf{X} and \mathbf{Y} by $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}^H \mathbf{Y})$ and $\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbb{R}} = \text{Re}[\langle \mathbf{X}, \mathbf{Y} \rangle]$ respectively. We use $\|\mathbf{X}\| = \langle \mathbf{X}, \mathbf{X} \rangle$ for the Frobenius norm of a matrix \mathbf{X} and $\|\mathbf{x}\|$ for the l_2 -norm of a vector \mathbf{x} . The identity matrix of order p is represented by \mathbf{I}_p . For an integer $k > 0$, we use the shorthand notation $[k]$ for $\{1, 2, \dots, k\}$.

II. PROBLEM STATEMENT

A. Basic Setup

Our model and system assumptions are standard in most classical works on massive MIMO (e.g., [2]–[5], [7]–[13], [17]) and recalled here for the sake of completeness and for establishing the notation to be used later. We consider a system with a signal bandwidth of W Hz and a scheduling slot of duration T_s sec (including UL pilots, UL payload, and DL payload [2]). The underlying channel fading process has a coherence bandwidth $\Delta f_c < W$ and a coherence time $\Delta t_c \geq T_s$ [20], such that in each scheduling slot we have $\lceil \frac{W}{\Delta f_c} \rceil$ frequency sub-bands over which the channel can be considered (approximately) frequency-flat and constant in time over the whole duration of a slot. We call a frequency-time rectangle of bandwidth Δf_c and duration T_s a *coherence block* (CB). This is illustrated in Fig. 2, where it is seen that the channel is approximately constant over a CB but changes smoothly across different CBs. We denote by $\Delta \tau_{\max}$ the

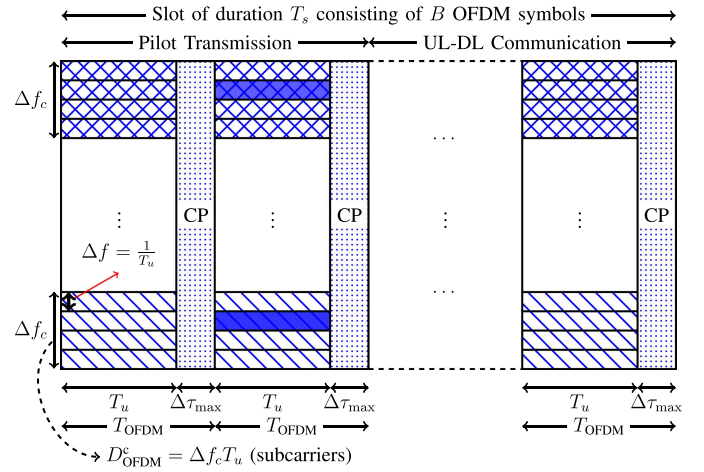


Fig. 2. A slot of duration T_s consisting of several OFDM symbols. In this example, a coherence sub-block inside an OFDM symbol contains $D_{\text{OFDM}}^c = 4$ signal dimensions (subcarriers), and two OFDM symbols are devoted to pilot transmission, thus, $D_p^c = 2D_{\text{OFDM}}^c = 8$. We also illustrate a 0-1 pilot sequence that lies on the second training OFDM symbol and probes subcarrier #3 inside each coherence sub-block.

maximum channel delay spread that the system can handle without suffering from inter-block interference between the OFDM symbols [21]. We assume that a set of B OFDM symbols are transmitted inside a time slot, each having a total duration of $T_{\text{OFDM}} = \frac{T_s}{B}$ and an effective duration of $T_u = T_{\text{OFDM}} - \Delta \tau_{\max}$ after removing the *cyclic prefix* (CP) of duration $\Delta \tau_{\max}$. The frequency spacing between the subcarriers is given by $\Delta f = \frac{1}{T_u}$, thus, each OFDM symbol has $N = \frac{W}{\Delta f} = WT_u$ subcarriers. Over each slot, we have a set of $Q = NB = WT_uB = WT_s (1 - \frac{\Delta \tau_{\max}}{T_{\text{OFDM}}})$ signal dimensions. Also, each CB is decomposed into B disjoint sub-blocks lying inside separate OFDM symbols, where each sub-block consists of $D_{\text{OFDM}}^c = \Delta f_c T_u$ subcarriers and, in total, there are $D^c = BD_{\text{OFDM}}^c$ signal dimensions in each CB. During each slot, some D_p^c out of D^c signal dimensions inside each CP are devoted to pilot transmission, while the remaining $D_d^c = D^c - D_p^c$ signal dimensions are used for UL-DL data transmission. A set of D_p^c orthogonal pilot sequences are assigned to D_p^c pilot signal dimensions in each CB. The resulting orthogonal pilots are allocated to the users in each cell/sector according to a given reuse pattern (e.g., [2], [7]), where in reuse patterns with a reuse factor $r \geq 1$, at most $K = \frac{D_p^c}{r}$ users can be simultaneously served per cell/sector with mutually orthogonal pilot sequences. In this paper, without any loss of generality, we consider 0-1 pilot sequences (see Fig. 2), where the pilot sequence of each user is transmitted over a single OFDM symbol and places a single “1” in each CB. In this way, each pilot sequence probes one subcarrier per CB (see, e.g., [5]), for a total of $n = \frac{N}{D_{\text{OFDM}}^c}$ subcarriers.

B. Pilot Contamination

We consider a reference BS called BS_0 and denote by $\text{UE}_{0,k}$ a generic user $k \in [K]$ served by BS_0 . As before, we assume that the pilot signal of $\text{UE}_{0,k}$ is transmitted over an individual OFDM symbol and probes a subset of subcarriers $\Omega_k \subset [N]$ of size $|\Omega_k| = n$. We denote by \mathcal{K}_k the set of all CPUs of

$\text{UE}_{0,k}$, i.e., the users across the whole system that transmit their pilot signal over the same pilot OFDM symbol and over the same set of subcarriers Ω_k as $\text{UE}_{0,k}$. The received signal of $\text{UE}_{0,k}$ at BS_0 during the pilot transmission is given by

$$\mathbf{y}_{k,s}[\omega] = \mathbf{h}_{k,s}[\omega] + \sum_{k' \in \mathcal{K}_k} \mathbf{h}_{k',s}[\omega] + \mathbf{z}_s[\omega], \quad \omega \in \Omega_k, \quad (1)$$

where $\mathbf{h}_{k,s}[\omega]$ and $\mathbf{h}_{k',s}[\omega]$ denote the M -dim channel vectors of $\text{UE}_{0,k}$ and its CPUs \mathcal{K}_k to the M antennas at BS_0 at time slot s and subcarrier ω , where $\mathbf{z}_s[\omega] \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$ is the *additive white Gaussian noise* (AWGN) at subcarrier ω , and where we assumed, without loss of generality, that the transmitted pilot symbols at all subcarriers $\omega \in \Omega_k$ are normalized to 1. From (1), it is seen that during the UL pilot transmission phase, the BS receives the superposition of the channel vector of $\text{UE}_{0,k}$ and that of its CPUs, thus, pilot contamination.

C. Wideband Pilot Decontamination

We denote by $\mathbf{H}_{k,s} = [\mathbf{h}_{k,s}[1], \dots, \mathbf{h}_{k,s}[N]]$ and $\mathbf{H}_{k',s} = [\mathbf{h}_{k',s}[1], \dots, \mathbf{h}_{k',s}[N]]$, $k' \in \mathcal{K}_k$, the $M \times N$ wideband channel matrices of $\text{UE}_{0,k}$ and its CPUs across N OFDM subcarriers at time slot s . We denote by $\mathbf{S}_{k,s}^{\mathbf{f}}$ an $n \times N$ matrix that has a single 1 in each row at columns corresponding to the probed subcarriers Ω_k and is 0 elsewhere. We also assume, for the sake of generality, that during the UL training phase, a subset of size m of the M BS antennas is sampled via an $m \times M$ matrix $\mathbf{S}_s^{\mathbf{a}}$. Thus, from (1), the UL pilot observation for $\text{UE}_{0,k}$ at BS_0 at time slot s can be arranged as a $m \times n$ matrix

$$\mathbf{X}_{k,s} := \mathbf{S}_s^{\mathbf{a}} \mathbf{Y}_s \mathbf{S}_{k,s}^{\mathbf{f}}^{\mathbf{H}} = \tilde{\mathbf{H}}_{k,s} + \sum_{k' \in \mathcal{K}_k} \tilde{\mathbf{H}}_{k',s} + \tilde{\mathbf{Z}}_{k,s}, \quad (2)$$

where $\mathbf{Y}_s = [\mathbf{y}_s[1], \dots, \mathbf{y}_s[N]]$ denotes the $M \times N$ wideband signal received across all the subcarriers, and where $\tilde{\mathbf{H}}_{k,s} = \mathbf{S}_s^{\mathbf{a}} \mathbf{H}_{k,s} \mathbf{S}_{k,s}^{\mathbf{f}}^{\mathbf{H}}$ contains the channel coefficients of $\text{UE}_{0,k}$ corresponding to the m sampled antennas and the n probed subcarriers, with the same interpretation holding for $\tilde{\mathbf{H}}_{k',s}$, $k' \in \mathcal{K}_k$, and $\tilde{\mathbf{Z}}_{k,s}$. We denote by $\mathbf{h}_{k,s} = \text{vec}(\mathbf{H}_{k,s})$ and $\mathbf{h}_{k',s} = \text{vec}(\mathbf{H}_{k',s})$, $k' \in \mathcal{K}_k$, the *wideband channel vectors* obtained after vectorization. Applying the vec operator and using the identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathbf{T}} \otimes \mathbf{A})\text{vec}(\mathbf{B})$, we can write (2) as

$$\mathbf{x}_{k,s} = \mathbf{S}_{k,s} \mathbf{y}_{k,s} = \tilde{\mathbf{h}}_{k,s} + \sum_{k' \in \mathcal{K}_k} \tilde{\mathbf{h}}_{k',s} + \tilde{\mathbf{z}}_{k,s}, \quad (3)$$

where $\mathbf{S}_{k,s} = \mathbf{S}_{k,s}^{\mathbf{f}} \otimes \mathbf{S}_s^{\mathbf{a}}$ and where $\tilde{\mathbf{h}}_{k,s} = \mathbf{S}_{k,s} \mathbf{h}_{k,s}$, $\tilde{\mathbf{h}}_{k',s} = \mathbf{S}_{k',s} \mathbf{h}_{k',s}$, $k' \in \mathcal{K}_k$. With this notation, the objective of pilot decontamination can be stated as follows.

Pilot Decontamination: Given the contaminated and noisy UL *wideband* pilot sketches $\{\mathbf{x}_{k,s} : s \in [\mathbf{w}]\}$ of the desired user $\text{UE}_{0,k}$ across \mathbf{w} time slots, construct an estimator for its wideband channel vector $\mathbf{h}_{k,s}$ (equivalently, its wideband channel matrix $\mathbf{H}_{k,s}$) at the next time slots $s \geq \mathbf{w} + 1$. \diamond

To explain this better, let us define the *wideband (space-frequency)* covariance matrices of $\text{UE}_{0,k}$ and of its CPUs by $\mathbf{C}_k = \mathbb{E}[\mathbf{h}_{k,s} \mathbf{h}_{k,s}^{\mathbf{H}}]$ and $\mathbf{C}_{k'} = \mathbb{E}[\mathbf{h}_{k',s} \mathbf{h}_{k',s}^{\mathbf{H}}]$, $k' \in \mathcal{K}_k$, independent of s by the WSS assumption (see Section III-A). Note

that if these covariance matrices are available at BS_0 , using the fact that the channel vectors $\mathbf{h}_{k,s}$ and $\{\mathbf{h}_{k',s} : k' \in \mathcal{K}_k\}$ are independent vector-valued stationary Gaussian random processes (see Section III-A for more details), the immediate answer to our estimation problem for pilot decontamination would be the MMSE smoothing filter, given by¹

$$\begin{aligned} \hat{\mathbf{h}}_{k,s} &= \Sigma_{\mathbf{h}_{k,s}, \mathbf{x}_{k,s}} \mathbf{C}_{\mathbf{x}_{k,s}}^{-1} \mathbf{x}_{k,s} \\ &= \mathbf{C}_k \mathbf{S}_{k,s}^{\mathbf{H}} \left(\sigma^2 \mathbf{I}_{mn} + \mathbf{S}_{k,s} (\mathbf{C}_k + \sum_{k' \in \mathcal{K}_k} \mathbf{C}_{k'}) \mathbf{S}_{k,s}^{\mathbf{H}} \right)^{-1} \mathbf{x}_{k,s}, \end{aligned} \quad (4)$$

where $\Sigma_{\mathbf{h}_{k,s}, \mathbf{x}_{k,s}} := \mathbb{E}[\mathbf{h}_{k,s} \mathbf{x}_{k,s}^{\mathbf{H}}] = \mathbf{C}_k \mathbf{S}_{k,s}^{\mathbf{H}}$, and where we used $\mathbf{S}_{k,s}^{\mathbf{H}} \mathbf{S}_{k,s} = \mathbf{I}_{mn}$. In practice, however, \mathbf{C}_k and $\mathbf{C}_{k'}$, $k' \in \mathcal{K}_k$, are not available and should be estimated from the noisy and contaminated pilot sketches $\{\mathbf{x}_{k,s} : s \in [\mathbf{w}]\}$. With this brief explanation, the problems we are addressing in this paper are as follows:

1) *How can we efficiently estimate the wideband covariance matrices of the desired user and of the CPUs from the subsampled and contaminated observations $\{\mathbf{x}_{k,s} : s \in [\mathbf{w}]\}$?* We address this question by estimating the contaminated wideband channel covariance matrix $\mathbf{C}_k + \sum_{k' \in \mathcal{K}_k} \mathbf{C}_{k'}$ via exploiting the sparsity of MPCs in the angle-delay domain (Section IV), and applying suitable clustering techniques in the angle-delay domain to decompose approximately the resulting contaminated wideband covariance matrix into its signal and interference parts \mathbf{C}_k and $\sum_{k' \in \mathcal{K}_k} \mathbf{C}_{k'}$ (Section V-B and V-C).

2) *How can we approximate the MMSE smoothing filter (4) in an efficient way, not requiring inversion of a $mn \times mn$ matrix and complicated and time-consuming matrix-matrix multiplication in (4)?* We address this complexity issue by developing computationally-efficient pilot decontamination and channel interpolation algorithms (Section VI and Appendix).

III. WIDEBAND CHANNEL MODEL

A. WSS-US Assumption

The COST 2100 channel model consists (up to some drastic simplifications) of clusters of MPCs and visibility regions [22]. The propagation between a BS and a user inside the intersection of multiple visibility regions occurs through all corresponding clusters (see Fig. 3). This implies that the statistics of the channel between a BS and a user remains constant in time and frequency as long as the user remains in the intersection of same visibility regions. As the user crosses the boundary of some region and enters a new region, the channel statistics typically undergoes a sharp transition. Since moving across the regions occurs at a time scale much larger than moving across one wavelength, it is safe to assume

¹Notice that here, knowing the *space-frequency* covariance matrices, we used only the observation at slot s to estimate the channel at slot s . In general, we can use $\mathbf{x}_{k,s}$ together with all the past observations $\{\mathbf{x}_{k,s'} : s' < s\}$ to do pilot decontamination, but this will require estimating the *space-frequency-doppler* covariance matrices of the current and past observations, which would result in even a more complex estimator. In practice, since the slot duration T_s is usually chosen to be of the same order of the channel coherence time Δt_c , the channel samples at different slots are nearly independent, and there is very little to gain from the temporal correlation of the fading. For this reason, we restrict to the common practice of estimating the channel based on the current slot UL pilot observation [2]–[4].

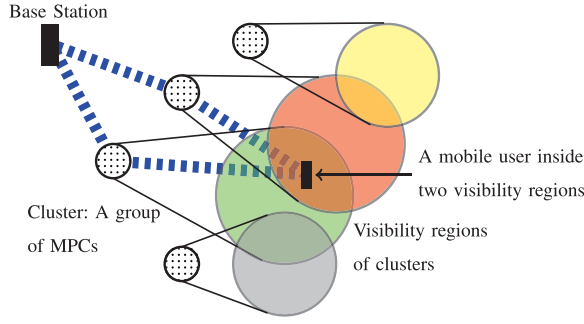


Fig. 3. A sketch of the cluster and visibility regions of the COST2100 model.

that the channel statistics is piecewise time-invariant with relatively sharp transitions at very low rate compared with the signaling rate. In this paper, for simplicity, we neglect such transitions and suppose a time-invariant second-order statistics for the channel during the whole communication interval, i.e., the channel process is assumed to remain (locally) *Wide Sense Stationary* (WSS) over time. Furthermore, the MPCs originated by different users and/or different scattering clusters are assumed to be mutually uncorrelated (US assumption). Finally, since each MPC is formed by a very large number of elementary multipath contributions, superimposing with different phases, invoking the Central Limit Theorem it is widely accepted to model the MPC coefficients as complex circularly symmetric Gaussian [20], [21].

B. Sparsity in the Angle-Delay Domain

Without loss of generality, we focus on a single BS-user pair and neglect the user and BS indices to simplify the notation. Also, for simplicity, we adopt a discrete multipath model [5], [23]–[26] with L MPCs, each of which is characterized by an *Angle of Arrival* (AoA) θ_l and a delay τ_l . All the results of this paper extend to the general case of mixed-type discrete-continuous scattering as long as the MPCs have a limited angle-delay support. In each time slot s , the UL channel is given by the vector impulse response

$$\check{\mathbf{h}}_s(\tau) = \sum_{l=1}^L \rho_{s,l} \mathbf{a}(\theta_l) \delta(\tau - \tau_l), \quad (5)$$

where $\mathbf{a}(\theta)$ denotes the array response at AoA $\theta \in [-\theta_{\max}, \theta_{\max}]$, whose k -th component is $[\mathbf{a}(\theta)]_k = e^{jk \frac{2\pi d}{\lambda} \sin(\theta)}$, where $\lambda = \frac{c_0}{f_0}$ denotes the wavelength (c_0 denoting the speed of light) and where f_0 is the carrier frequency. We assume that the array elements have the uniform spacing $d = \frac{\lambda}{2 \sin(\theta_{\max})}$, thus, $[\mathbf{a}(\theta)]_k = e^{jk\pi \frac{\sin(\theta)}{\sin(\theta_{\max})}}$. As said before, from the WSS-US and Gaussian assumption, we have that the discrete-time path gain processes $\{\rho_{s,l} : s \in \mathbb{Z}\}$ are stationary with respect to the (slot) time index s and independent across l . Furthermore, we assume no line of sight propagation, yielding $\rho_{s,l} \sim \mathcal{CN}(0, \sigma_l^2)$, where σ_l^2 denotes the strength of the l -th MPC (independent of s because of the WSS assumption).

In the OFDM discrete frequency domain, channel (vector) frequency response corresponding to the impulse response (5)

is given by

$$\mathbf{h}_s[\omega] = \sum_{l=1}^L \rho_{s,l} \mathbf{a}(\theta_l) e^{-j \frac{2\pi}{N} W \tau_l (\omega-1)}, \quad \omega \in [N], \quad (6)$$

such that, as anticipated in Section II, the wideband channel matrix at slot s is given by

$$\mathbf{H}_s = [\mathbf{h}_s[1], \mathbf{h}_s[2], \dots, \mathbf{h}_s[N]]. \quad (7)$$

We define the N -dim vector $\mathbf{b}(\tau)$, whose ω -th component given by $[\mathbf{b}(\tau)]_\omega = e^{j \frac{2\pi}{N} W \tau (\omega-1)}$. Thus, we can write (7) more compactly as

$$\mathbf{H}_s = \sum_{l=1}^L \rho_{s,l} \mathbf{a}(\theta_l) \mathbf{b}(\tau_l)^H. \quad (8)$$

The rows of \mathbf{H}_s correspond to the antenna elements, whereas its columns correspond to the OFDM subcarriers. The vectorized channel vector $\mathbf{h}_s = \text{vec}(\mathbf{H}_s)$ is, therefore, given by

$$\mathbf{h}_s = \sum_{l=1}^L \rho_{s,l} \mathbf{a}(\theta_l, \tau_l), \quad (9)$$

where $\mathbf{a}(\theta, \tau) = \text{vec}(\mathbf{a}(\theta) \mathbf{b}(\tau)^H) \in \mathbb{C}^{MN}$ denotes the array response in the angle-delay (θ, τ) . Since the MPC coefficients $\rho_{s,l} \sim \mathcal{CN}(0, \sigma_l^2)$ are independent and circularly symmetric Gaussian variables, from (6) it is immediate to check that the statistics of $\mathbf{h}_s[\omega]$ are invariant under circular shifts (with period N) in ω , implying stationarity in the frequency domain.

C. Antenna-Frequency Sampling

As explained in Section II, without any loss of generality, we can assume that a UL pilot sequence for each user probes its channel over a subset of subcarriers in an individual OFDM symbol. Also, the pilot corresponding to different users are sent either across different OFDM symbols (disjoint in time) or across the same OFDM symbol but on disjoint set of subcarriers (disjoint in frequency). As before, we focus on a single user and denote by $\mathcal{I}_s^f = \{c_{s,1}^f, \dots, c_{s,n}^f\}$ the indices of the subcarriers acquired for this user at slot s . In addition, we consider the general case where also the antennas may be subsampled. This is done for the sake of generality, and also because one may wish to exploit the channel spatial correlation and reduce the sampling overhead at the receiver side. We denote by $\mathcal{I}_s^a = \{c_{s,1}^a, \dots, c_{s,m}^a\}$ the indices of the antennas sampled at time slot s . We define $n \times N$ and $m \times M$ selection (or sampling) matrices² \mathbf{S}_s^a and \mathbf{S}_s^f , where $[\mathbf{S}_s^a]_{i,c_{s,i}^a} = 1$ and $[\mathbf{S}_s^f]_{j,c_{s,j}^f} = 1$, for $i \in [n]$ and $j \in [m]$. The sampled channel matrix at slot s is given by $\tilde{\mathbf{H}}_s = \mathbf{S}_s^a \mathbf{H}_s \mathbf{S}_s^{fH}$. Using the vec notation, this can be written as

$$\tilde{\mathbf{h}}_s = \mathbf{S}_s \mathbf{h}_s, \quad (10)$$

²In this paper, for simplicity, we focus on 0-1 antenna and subcarrier sampling matrices. This type of sampling is suitable for the Compressed Sensing algorithm that we develop later on in the paper to estimate the angle-delay PSF. However, our proposed method can be extended to work with more general projection matrices in the antenna and also subcarrier domain.

where $\mathbb{S}_s = \mathbf{S}_s^f \otimes \mathbf{S}_s^a$ is of dimension $mn \times MN$, and where we used the well-known identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$. Notice that $\mathbb{S}_s \mathbb{S}_s^H = \mathbf{I}_{mn}$, and that \mathbb{S}_s has only a single element equal to 1 in each row at column indices given by

$$\mathcal{I}_s := \{M(c_s^f - 1) + c_s^a : c_s^f \in \mathcal{I}_s^f, c_s^a \in \mathcal{I}_s^a\} \subseteq [MN]. \quad (11)$$

Using the above notation, the observation at the reference BS corresponding to a generic user (see (2) and (3)) can be written as $\mathbf{y}_s = \mathbb{S}_s \mathbf{y}_s$, where

$$\mathbf{y}_s = \mathbf{h}_s + \sum_{j \in \mathcal{K}} \mathbf{h}_{j,s} + \mathbf{z}_s, \quad (12)$$

where \mathcal{K} denotes the set of CPUs of a generic user in the reference cell/sector, and where, for notation simplicity, we dropped the index of the user and the copilot set \mathcal{K} (see, e.g., (1) and (2)) and indicated the channel vectors of a generic user and of its CPUs at slot s by \mathbf{h}_s and $\mathbf{h}_{j,s}$, $j \in \mathcal{K}$.

IV. ESTIMATION OF SPARSE SCATTERING CHANNEL

In this section, we propose a low-complexity algorithm to estimate the sparse geometry of the channel in the angle-delay domain as illustrated in Fig. 1. The resulting estimator is used in Section V to perform pilot decontamination and channel interpolation.

A. Low-Dim Signal Structure

Consider the reference BS-user pair with channel at slot s given by (9). The covariance matrix of \mathbf{h}_s is given by $\mathbf{C}_{\mathbf{h}} := \mathbb{E}[\mathbf{h}_s \mathbf{h}_s^H] = \sum_{l=1}^L \sigma_l^2 \mathbf{a}(\theta_l, \tau_l) \mathbf{a}(\theta_l, \tau_l)^H$. It is seen that although $\mathbf{C}_{\mathbf{h}}$ is a very large-dim $MN \times MN$ matrix, it is very low-rank (here the rank is L), due to sparse angle-delay scattering. This low-rank property still holds when the channel consists of a continuum of MPCs, provided that they have a small angle-delay support.

In the UL pilot observation model in (12), we denote by $\mathbf{d}_s = \mathbf{h}_s + \sum_{j \in \mathcal{K}} \mathbf{h}_{j,s}$ the *superposition of the channel vectors* (SCVs) of the desired user and that of its $|\mathcal{K}|$ CPUs. Because of the distance-dependent pathloss, the number of CPUs with a significant received power is quite small. In particular, all the CPUs with covariance matrices $\mathbf{C}_{\mathbf{h}_j} := \mathbb{E}[\mathbf{h}_{j,s} \mathbf{h}_{j,s}^H]$, $j \in \mathcal{K}$, for which $\frac{1}{MN} \text{tr}(\mathbf{C}_{\mathbf{h}_j}) \ll \sigma^2$ can be neglected. Hence, without loss of generality, we can restrict \mathcal{K} to include only the CPUs with significant “raise over thermal”, i.e., those whose received power at the reference BS is significantly larger than the noise level. Therefore, the covariance matrix $\mathbf{C}_{\mathbf{d}} := \mathbb{E}[\mathbf{d}_s \mathbf{d}_s^H] = \mathbf{C}_{\mathbf{h}} + \sum_{j \in \mathcal{K}} \mathbf{C}_{\mathbf{h}_j}$ of SCVs is still very low-rank.

Our goal in this section is to exploit this low-rank structure to estimate $\mathbf{C}_{\mathbf{d}}$ efficiently. To do so, we collect multiple sketches $\mathbf{z}_s = \mathbb{S}_s \mathbf{y}_s$, via $mn \times MN$ possibly time-variant sampling operators \mathbb{S}_s , inside a window of size w of training slots across w CBs. We represent these sketches by an $mn \times w$ matrix \mathbb{Z} . Recall that the sampling matrix $\mathbb{S}_s = \mathbf{S}_s^f \otimes \mathbf{S}_s^a$ consists of antenna and frequency sampling, where \mathbf{S}_s^f samples some of the subcarriers of a pilot OFDM symbol according to the UL pilot pattern (0-1 pattern) assigned to the user (see Section II), and where \mathbf{S}_s^a samples some of the antennas (pseudo)-randomly in each

slot s . The performance of our proposed subspace estimation algorithm improves if the frequency signature of the user is also non-equally spaced and (pseudo)-randomly time-varying over the slots. This can be implemented in practice by assigning a frequency-hopping pseudo-random pilot pattern to the users synchronized with the BS, analogous to what is currently done in CDMA systems. The drawback is that, in contrast with the uniform sampling scheme suggested by the classical Shannon-Nyquist sampling, the recovery of the whole instantaneous channel matrix from its nonuniform samples requires more complicated interpolation algorithms. As we will explain in Section V-C and VI-A, our proposed channel interpolation technique can be easily applied to both uniform and nonuniform sampling cases without incurring any additional complexity for the nonuniform one. The design of suitable pseudo-random frequency signatures yielding easy interpolation is itself an interesting problem, which is beyond the scope of this paper.

B. Low-Complexity Subspace Estimation

We use the low-complexity algorithm we developed in our previous work [27], [28] to estimate the signal subspace of the SCVs \mathbf{d}_s from the sketches $\{\mathbf{z}_s : s \in [w]\}$. The proposed algorithm is reminiscent of *Multiple Measurement Vectors* problem in Compressed Sensing and exploits the joint sparsity of SCVs in the angle-delay domain. We first quantize the angle-delay domain into a discrete grid $\mathcal{G} := \{(\theta_i, \tau_i)\}$, where for simplicity we use a uniform rectangular grid with $G = G^\theta G^\tau$ elements, with corresponding oversampling factors $\frac{G^\theta}{M}$ and $\frac{G^\tau}{N}$ in the angle and the delay domains, respectively. We define an $MN \times G$ quantized dictionary matrix \mathbb{A} whose i -th column is given by $\bar{\mathbf{a}}(\theta_i, \tau_i)$, where $\bar{\mathbf{a}}(\theta, \tau) = \frac{\mathbf{a}(\theta, \tau)}{\sqrt{MN}}$ denotes the normalized array response at angle-delay (θ, τ) . We define the $mn \times w$ matrix $\mathbb{X} = [\mathbf{x}_1, \dots, \mathbf{x}_w]$ that contains the sketches $\{\mathbf{z}_s : s \in [w]\}$. We assume that the noise power σ^2 in each antenna is known and normalize the received sketches by σ where, for simplicity of notation, we denote the normalized sketches $\frac{\mathbb{Z}}{\sigma}$ again by \mathbb{Z} . We use the following $l_{2,1}$ -norm regularized least squares proposed in [28] to estimate the signal subspace of the channel superposition:

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \frac{1}{2} \sum_{s=1}^w \|\tilde{\mathbb{A}}_s \mathbf{W}_{:,s} - \mathbb{Z}_{:,s}\|^2 + \sqrt{w} \|\mathbf{W}\|_{2,1}, \quad (13)$$

where $\tilde{\mathbb{A}}_s = \sqrt{\frac{MN}{mn}} \mathbb{S}_s \mathbb{A}$ is a scaled and subsampled (via \mathbb{S}_s) version of \mathbb{A} , and where $\mathbf{W} \in \mathbb{C}^{G \times w}$ is a matrix whose rows correspond to the random channel gain of the MPCs over the quantized grid \mathcal{G} across w slots. Notice that $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^G \|\mathbf{W}_{i,:}\|$ denotes the $l_{2,1}$ -norm of \mathbf{W} with $\mathbf{W}_{i,:} \in \mathbb{C}^w$ denoting the i -th row of \mathbf{W} . The sparsity of the SCVs in the angle-delay domain results in the row-sparsity of the coefficient matrix \mathbf{W} , i.e., \mathbf{W} must have only a few nonzero rows along the active grid elements $(\theta_i, \tau_i) \in \mathcal{G}$ corresponding to the MPCs.

In our previous work [28], we used a $l_{2,1}$ -norm regularizer for \mathbf{W} to promote this row-sparsity. The resulting algorithm is

recalled here since it forms a key step of the proposed channel decontamination and interpolation scheme. Consider the objective function (13). After suitable scaling, we can write (13) as the minimization of function $f(\mathbf{W}) = f_1(\mathbf{W}) + f_2(\mathbf{W})$, where $f_1(\mathbf{W}) = \frac{1}{2\zeta} \sum_{c=1}^w \|\tilde{\mathbf{A}}_c \mathbf{W}_{:,c} - \mathbf{X}_{:,c}\|^2$ with $\zeta = \sqrt{w}$ and $f_2(\mathbf{W}) = \|\mathbf{W}\|_{2,1}$. The gradient of f_1 is a $G \times w$ matrix $\nabla f_1(\mathbf{W})$ whose c -th column, $c \in [w]$, is given by $\nabla f_1(\mathbf{W})_{:,c} = \frac{1}{\zeta} \tilde{\mathbf{A}}_c^H (\tilde{\mathbf{A}}_c \mathbf{W}_{:,c} - \mathbf{X}_{:,c})$. To apply the algorithm in [28], we need to compute the Lipschitz constant of ∇f_1 , i.e., the smallest constant $\beta > 0$ such that for every $\mathbf{W}, \mathbf{W}' \in \mathbb{C}^{G \times w}$:

$$\|\nabla f_1(\mathbf{W}) - \nabla f_1(\mathbf{W}')\| \leq \beta \|\mathbf{W} - \mathbf{W}'\|. \quad (14)$$

We can check that $\beta \leq \frac{1}{\zeta} \max_{c \in [w]} \lambda_{\max} \{\tilde{\mathbf{A}}_c^H \tilde{\mathbf{A}}_c\}$, where λ_{\max} denotes the maximum singular value of a given matrix. Note that if the grid size G is sufficiently large and the grid points $\{(\theta_i, \tau_i)\}$ are distributed quite uniformly and densely over the angle-delay domain, we have that

$$\begin{aligned} \tilde{\mathbf{A}}_c^H \tilde{\mathbf{A}}_c &= \frac{1}{mn} \mathbb{S}_c \left\{ \sum_{i=1}^G \mathbf{a}(\theta_i, \tau_i) \mathbf{a}(\theta_i, \tau_i)^H \right\} \mathbb{S}_c^H \\ &\approx \frac{G}{mn} \mathbb{S}_c \mathbf{I}_{MN} \mathbb{S}_c^H = \frac{G}{mn} \mathbf{I}_{mn}, \end{aligned}$$

where we used $\mathbb{S}_c \mathbb{S}_c^H = \mathbf{I}_{mn}$. This implies that $\beta \approx \frac{G}{\zeta mn} = \frac{G}{mn\sqrt{w}}$. We also need the proximal operator of $l_{2,1}$ -norm f_2 with a scaling $\alpha > 0$ defined by $\text{prox}_{\alpha f_2} : \mathbb{C}^{G \times w} \rightarrow \mathbb{C}^{G \times w}$, whose i -th row is given by

$$(\text{prox}_{\alpha f_2}(\mathbf{W}))_{i,:} = \frac{(\|\mathbf{W}_{i,:}\| - \alpha)_+}{\|\mathbf{W}_{i,:}\|} \mathbf{W}_{i,:} \quad (15)$$

and corresponds to a *shrinkage* operator shrinking the rows of \mathbf{W} by α , where $(x)_+ := \max(x, 0)$. The algorithm proposed in [28] with the Nesterov's step-size update is given by Algorithm 1. We have also the following performance guarantee from [28].

Algorithm 1 Forward-Backward Splitting With Nesterov's Update

-
- 1: **Initialization:** Fix $\mathbf{W}^{(0)}$, set $\mathbf{Z}^{(0)} = \mathbf{W}^{(0)}$, $t_0 = 1$.
 - 2: **for** $k = 0, 1, \dots$, **do**
 - 3: $\mathbf{R}^{(k)} = \mathbf{Z}^{(k)} - \frac{1}{\beta} \nabla f_1(\mathbf{Z}^{(k)})$.
 - 4: $\mathbf{W}^{(k+1)} = \text{prox}_{\frac{1}{\beta} f_2}(\mathbf{R}^{(k)})$.
 - 5: $t_{k+1} = \frac{1 + \sqrt{4t_k^2 + 1}}{2}$ and $\mu_k = 1 + \frac{t_k - 1}{t_{k+1}}$.
 - 6: $\mathbf{Z}^{(k+1)} = \mathbf{W}^{(k)} + \mu_k (\mathbf{W}^{(k+1)} - \mathbf{W}^{(k)})$.
 - 7: **end for**
-

Proposition 1 ([29, Th. 11.3.1]): Let $\{\mathbf{W}^{(k)}\}_{k=0}^\infty$ be the sequence generated by Algorithm 1 for an arbitrary initial point $\mathbf{W}^{(0)}$ and for the step-sizes according to the Nesterov's update rule. Then, for any k , we have $f(\mathbf{W}^{(k+1)}) - f(\mathbf{W}^*) \leq \frac{4\beta \|\mathbf{W}^* - \mathbf{W}^{(0)}\|^2}{(k+1)^2}$.

Let \mathbf{W}^* be the optimal solution of (13) and let $s_i = \frac{1}{mn} \|\mathbf{W}_{i,:}^*\|$ be the l_2 -norm of the i -th row of \mathbf{W}^* . The covariance matrix \mathbf{C}_d of the channel superposition can be estimated

from [28, Proposition 1] by

$$\mathbf{C}_d^* \approx \sum_{i=1}^G s_i \bar{\mathbf{a}}(\theta_i, \tau_i) \bar{\mathbf{a}}(\theta_i, \tau_i)^H. \quad (16)$$

Remark 1: By increasing the grid size G , (16) provides a more precise estimate of the covariance matrix of SCVs. However, as also mentioned in [28], since the Lipschitz constant $\beta = \frac{G}{mn\sqrt{w}}$ grows proportionally to G , it is seen from Proposition 1 that increasing G reduces the convergence speed of the algorithm. Intuitively, by increasing G and as a result β , the shrinkage operator $\text{prox}_{\frac{1}{\beta} f_2}$ in Algorithm 1 becomes softer, and the algorithm requires more iterations to converge, although at the end the resulting estimate \mathbf{C}_d^* in (16) is generally improved.

C. Computational Complexity

Each iteration of Algorithm 1 requires computing w columns of $G \times w$ gradient ∇f_1 , where the c -th column, $c \in [w]$, is given by $\nabla f_1(\mathbf{W})_{:,c} = \frac{1}{\zeta} \tilde{\mathbf{A}}_c^H (\tilde{\mathbf{A}}_c \mathbf{W}_{:,c} - \mathbf{X}_{:,c})$, evaluated at $\mathbf{W} = \mathbf{W}^{(k)}$ at iteration k . Here, we consider a special grid \mathcal{G} whose discrete AoAs θ_k belong to

$$\Theta := \left\{ \sin^{-1} \left((-1 + \frac{2(i-1)}{G}) \sin(\theta_{\max}) \right) : i \in [G^\theta] \right\},$$

in the angular range $[-\theta_{\max}, \theta_{\max}]$. We also assume that all the discrete delay elements τ_k in \mathcal{G} belong to a uniform grid in the delay domain $[0, \Delta\tau_{\max}]$ of size G^τ . For this particular choice of the grid \mathcal{G} , the gradient matrix $\nabla f_1(\mathbf{W})$ can be efficiently computed via 2D *Fast Fourier Transform* (FFT) as follows. Let $\mathcal{I}_c^a \subseteq [M]$ and $\mathcal{I}_c^f \subseteq [N]$ denote the indices of the sampled antennas and subcarriers in the OFDM symbol at $c \in [w]$. For each $c \in [w]$, we first compute $\tilde{\mathbf{A}}_c \mathbf{W}_{:,c}$. Following the MATLAB[®] notation, we first set $\mathbf{M} = \text{reshape}(\mathbf{W}_{:,c}, G^\theta, G^\tau)$, and let $\tilde{\mathbf{M}} = G^\theta G^\tau \text{fft2}(\mathbf{M})$ be the inverse 2D *Discrete Fourier Transform* (DFT) of $\mathbf{W}_{:,c}$ scaled with G . This can be efficiently computed using the FFT algorithm in $O(G \log_2(G))$ operations under mild conditions on the integers G^θ and G^τ (e.g., they may be powers of 2). Then, $\tilde{\mathbf{A}}_c \mathbf{W}_{:,c}$ is simply given by $\frac{1}{\sqrt{mn}} \text{vec}(\tilde{\mathbf{M}}(\mathcal{I}_c^a, \mathcal{I}_c^f))$. The whole complexity of this step is $O(wG \log_2(G))$. After computing $\tilde{\mathbf{A}}_c \mathbf{W}_{:,c}$, we need to calculate $\tilde{\mathbf{A}}_c^H \mathbf{R}_{:,c}$, where \mathbf{R} is an $mn \times w$ matrix with $\mathbf{R}_{:,c} = \tilde{\mathbf{A}}_c \mathbf{W}_{:,c} - \mathbf{X}_{:,c}$, for $c \in [w]$. To do this, we set \mathbf{M} to be an $M \times N$ all-zero matrix and embed $\mathbf{R}_{:,c}$ in \mathbf{M} in indices belonging to \mathcal{I}_c^a and \mathcal{I}_c^f such that $\text{vec}(\mathbf{M}(\mathcal{I}_c^a, \mathcal{I}_c^f)) = \mathbf{R}_{:,c}$, and take the 2D DFT of \mathbf{M} , which gives $\tilde{\mathbf{A}}_c^H \mathbf{R}_{:,c} = \frac{1}{\sqrt{mn}} \text{vec}(\text{fft2}(\mathbf{M}, G^\theta, G^\tau))$. The whole complexity of this step is again $O(wG \log_2(G))$.

Letting T_{conv} be the number of iterations necessary for the convergence, the whole computational complexity of our algorithm is $O(2 T_{\text{conv}} wG \log_2(G))$. Typically, T_{conv} scales proportionally to $\frac{G}{MN}$ where, as also explained in Remark 1, increasing the grid size G slows down the convergence of the algorithm. We always use $\frac{G^\theta}{M} = \frac{G^\tau}{N} = 2$. Our numerical simulations show that for this choice of the oversampling factor, Algorithm 1 runs quite fast and converges in only a few iterations even for quite large $M, N \approx 256$ and $w \approx 200$.

D. System-Level Considerations

As we will further explain in the simulations in Section VII, our proposed algorithm is able to extract the signal subspace of SCVs by gathering UL pilot observations over a time window of the order of 50 ms, over which the underlying signal subspace can be safely assumed to remain invariant. In almost all practical situations, the subspace remains stable for a time scale of the order of $1 \sim 10$ s (see Section III-A and [20], [22]), which is much larger than the time scale required for estimating the subspace. Thus, the estimated subspace can be used for many time slots. In addition, the estimation phase results in almost no system-level overhead since i) during the estimation phase the system can still work in the (standard) contaminated mode ii) the sketches of the wideband channel vectors of the users gathered for subspace estimation are also required for serving the users, so subspace estimation does not impose effectively any additional sampling or pilot transmission overhead. In practice, one can apply subspace tracking algorithms as in [28] to update the estimated subspace upon arrival of each new observation at each new time slot s . This has the additional advantage that the computational complexity of the subspace estimation is distributed across several time slots.

V. PILOT DECONTAMINATION AND CHANNEL INTERPOLATION

A. Estimating the Angle-Delay Power Spread Function

Let \mathbf{W}^* be the $G \times w$ matrix of coefficients obtained as the optimal solution of the optimization (13). We define a discrete positive measure $\gamma(\theta, \tau)$ over the angle-delay domain that assigns the weight $\gamma(\theta_l, \tau_l) = \frac{1}{mn} \|\mathbf{W}_{l,\cdot}^*\|$ to the l -th grid element $(\theta_l, \tau_l) \in \mathcal{G}$. From Section IV-B, the covariance matrix \mathbf{C}_d of SCVs is well-approximated by \mathbf{C}_d^* in (16), which can be written as $\mathbf{C}_d^* = \int \gamma(d\theta, d\tau) \mathbf{a}(\theta, \tau) \mathbf{a}(\theta, \tau)^H$ in terms of the discrete measure $\gamma(\theta, \tau)$. Consequently, we expect that $\gamma(\theta, \tau)$ be a good approximation of the *angle-delay Power Spread Function* (PSF) of SCVs. We also define the marginal measure $\gamma(\theta) = \int_0^{\Delta\tau_{\max}} \gamma(\theta, d\tau)$, which provides an approximation of the PSF in the angle domain θ .

B. Clustering in the Angle-Delay Domain

A crucial ingredient of our pilot-decontamination method is a *clustering algorithm* in the angle-delay domain. The output of such a clustering algorithm is a decomposition $(\gamma_d(\theta, \tau), \gamma_{co}(\theta, \tau)) = \text{Clst}(\gamma(\theta, \tau))$ of the angle-delay PSF $\gamma(\theta, \tau)$ into a desired signal part $\gamma_d(\theta, \tau)$, corresponding to the wideband channel vector of the desired user, and a copilot interference part $\gamma_{co}(\theta, \tau)$, corresponding to the superposition of the wideband channel vectors of CPUs. This can be done using supervised or unsupervised learning techniques. In an unsupervised scheme, Clst exploits only the a priori knowledge about the desired user and its CPUs. This is typically based on the geometric constraints of the cell in which the user signal propagates such as the location as well as the received power strengths of different delay-angle elements. For example, if it is a priori known that all the

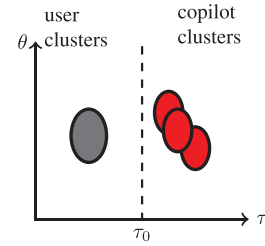


Fig. 4. Illustration of an unsupervised clustering using the a priori information in the delay domain.

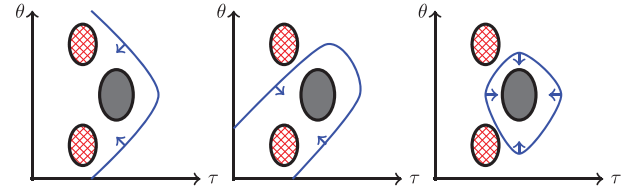


Fig. 5. Illustration of a supervised clustering algorithm. In this figure, $\gamma(\theta, \tau)$ consists of one signal cluster (solid gray) plus two interference clusters (dashed) corresponding to two CPUs.

copilot MPCs are separable in the delay domain, say by a delay threshold $\tau_0 \in [0, \Delta\tau_{\max})$, then the clustering algorithm Clst can be as simple as $\gamma_d(\theta, \tau) = \gamma(\theta, \tau) \mathbf{1}_{\{\tau \in [0, \tau_0]\}}$, and $\gamma_{co}(\theta, \tau) = \gamma(\theta, \tau) \mathbf{1}_{\{\tau \in (\tau_0, \Delta\tau_{\max}]\}}$, where $\mathbf{1}_{\mathcal{B}}$ denotes the indicator of a set \mathcal{B} , as illustrated in Fig. 4.

In a supervised scheme, Clst has access to an “oracle” provided by higher communication layers, which can be exploited to perform adaptive clustering. In brief, Clst starts from an initial clustering and refines it iteratively using the oracle response until a good partition of $\gamma(\theta, \tau)$ into $\gamma_d(\theta, \tau)$ and $\gamma_{co}(\theta, \tau)$ is obtained. A simple example of this is illustrated in Fig. 5. In this example, $\gamma(\theta, \tau)$ consists of one signal cluster and two CPU clusters, where for simplicity we have assumed that these clusters are non-overlapping. The Clst starts with the obvious initialization $\gamma_d(\theta, \tau) = \gamma(\theta, \tau)$, i.e., that there is no contamination (left figure in Fig. 5). Based on this assumption, it estimates the channel vector on all the subcarriers (using the channel interpolation scheme proposed in the following), and based on this channel estimation it attempts to decode the UL user data. In the presence of significant contamination, the effective *Signal to Interference plus Noise Ratio* (SINR) is degraded and some standard link layer control mechanism detects the data packet in error. This error detection mechanism can be exploited as an *oracle* for supervised learning. In the presence of a packet error, the Clst tries a different selection of the clusters (e.g., as in the center figure in Fig. 5). The process is repeated until the data packet is decoded correctly. Notice that in this case, although there is no guarantee that *all* the copilot interference be removed, we have the guarantee that it has been removed enough to decode the data, whenever this is possible. This means that the effective SINR for the desired user is large enough to achieve successful decoding. Of course, if no successful decoding is achieved after a fixed number of iterations, the packet is rejected, and the desired user is re-scheduled for transmission on a later slot. This is

not different from a standard “packet failure” event, which is handled by retransmission or by any suitable upper layer protocol in a completely standard manner. Notice also that Clst learns the suitable clustering without any explicit feedback from the users since the whole process is performed entirely at the BS receiver on a single packet detection. Therefore, it does not involve any additional latency with respect to a standard massive MIMO system. Interestingly, in this example, the clusters corresponding to the CPUs have smaller propagation delays than the one corresponding to the desired user. As a result, the previously mentioned unsupervised algorithm, which only exploits the propagation delay of the users, would fail to identify the signal cluster. Such a situation arises, for example, in a cell-free massive MIMO system, where copilot interference may be particularly harmful [30], [31].

C. Instantaneous Channel Estimation/Interpolation and Pilot Decontamination

Let $\gamma_d(\theta, \tau)$ and $\gamma_{co}(\theta, \tau)$ be the PSFs of the desired user and of the CPUs, obtained as described before. For channel decontamination and interpolation, we apply the MMSE smoothing filter (4) with “plug-in” covariance estimates given by $\mathbf{C}_h^* = \int \gamma_d(d\theta, d\tau) \mathbf{a}(\theta, \tau) \mathbf{a}(\theta, \tau)^H$ for the desired user channel, and by $\mathbf{C}_i^* = \int \gamma_{co}(d\theta, d\tau) \mathbf{a}(\theta, \tau) \mathbf{a}(\theta, \tau)^H$ for the superposition of the CPU channels. The resulting plug-in channel estimator-interpolator is given by

$$\begin{aligned} \hat{\mathbf{h}}_s &= \mathbf{\Sigma}_{\mathbf{h} \times} \mathbf{C}_s^{-1} \mathbf{z}_s \\ &\approx \mathbf{C}_h^* \mathbf{S}_s^H (\sigma^2 \mathbf{I}_{mn} + \mathbf{S}_s (\mathbf{C}_h^* + \mathbf{C}_i^*) \mathbf{S}_s^H)^{-1} \mathbf{z}_s, \end{aligned} \quad (17)$$

where $\mathbf{\Sigma}_{\mathbf{h} \times} = \mathbb{E}[\mathbf{h}_s \mathbf{z}_s^H]$ denotes the cross covariance matrix of \mathbf{h}_s and \mathbf{z}_s , and where $\mathbf{z}_s = \text{vec}(\mathbf{X}_s)$ is the UL pilot observation at time slot s . Under the condition that the estimated covariance matrices \mathbf{C}_h^* and \mathbf{C}_i^* are close to the true covariance matrices \mathbf{C}_h and \mathbf{C}_i , the channel estimator in (17) is close to the ideal MMSE smoothing filter (4). Notice that in the absence of copilot interference (i.e., for $\mathbf{C}_i = \mathbf{0}$) such MMSE smoothing filter implements the optimal channel interpolation in the antenna and frequency domain in the MMSE sense. In conventional implementations, “ad-hoc” channel interpolation techniques in the OFDM subcarrier domain are used in order to interpolate the unobserved columns (subcarriers) and rows (antennas) of the channel matrix \mathbf{H}_s from the instantaneous noisy UL pilot observation \mathbf{X}_s as given in (2). Typical schemes include simple piecewise constant, linear, or DFT-based (Sinc-shaped) interpolation (see [32], [33] and the refs. therein). The advantage of our proposed subspace estimation for pilot decontamination is that, as seen from (17), the channel vector \mathbf{h}_s can be directly estimated from the sketch \mathbf{z}_s , thus, we obtain per-slot channel estimation/interpolation for free.

VI. LOW-COMPLEXITY CHANNEL INTERPOLATION AND PILOT DECONTAMINATION

Computing the $MN \times MN$ covariance matrices from the estimated PSF $\gamma(\theta, \tau)$ and performing the matrix multiplication for the MMSE estimation in (17), as proposed in the

previous section, may result in a prohibitive complexity for typical massive MIMO systems (e.g., $M = 128$ antennas and $N = 128$ subcarriers). In this section, we propose two low-complexity algorithms to address this computational complexity issue. The first algorithm, explained in Section VI-A, uses a masking technique in the angle-delay domain, which yields a low-complexity approximation of the MMSE estimator proposed in (17). The second algorithm, stated in Section VI-B, has much lower complexity but, in order to guarantee to eliminate pilot contamination, requires a stronger angular separability condition as we will explain.

A. Interpolation and Pilot Decontamination by Masking

Let $\gamma(\theta, \tau)$ be the estimated PSF supported on the grid elements $(\theta_l, \tau_l) \in \mathcal{G}$ as in Section V-A. We define the mask \mathcal{M} as follows

$$\mathcal{M} := \{(\theta_l, \tau_l) : \gamma(\theta_l, \tau_l) \geq \iota\}, \quad (18)$$

where $\iota \in \mathbb{R}_+$ denotes a masking threshold in the angle-delay domain which selects only those grid elements with a significantly large received power. We assume that $\mathcal{M} = \mathcal{M}^h \cup \mathcal{M}^i$ is decomposed into disjoint signal and CPU interference masks \mathcal{M}^h and \mathcal{M}^i with $\mathcal{M}^h \cap \mathcal{M}^i = \emptyset$ via the clustering algorithm Clst. For the case of supervised clustering, Clst changes the masks \mathcal{M}^h and \mathcal{M}^i in each iteration, while keeping their union equal to \mathcal{M} as in (18), until it finds a good estimate of the true signal cluster, e.g., when the packet decoding is successful as described before.

Let \mathbf{S}_s^a and \mathbf{S}_s^f be the antenna and frequency sampling matrices at slot s . We apply joint interpolation and pilot decontamination as follows. We find an estimate of the channel matrix \mathbf{H}_s denoted by \mathbf{P} and an estimate of copilot interference denoted by \mathbf{Q} via minimizing $\|\mathbf{X}_s - \mathbf{S}_s^a (\mathbf{P} + \mathbf{Q}) \mathbf{S}_s^{fH}\|$, where $\mathbf{X}_s = \mathbf{S}_s^a \mathbf{Y}_s \mathbf{S}_s^{fH}$ denotes the subsampled observations at slot s and where \mathbf{Y}_s denotes the noisy contaminated received wideband signal. To do so, we impose the additional constraint that a significant amount of power of \mathbf{P} and \mathbf{Q} be concentrated in the mask \mathcal{M}^h and \mathcal{M}^i in the angle-delay domain, respectively. We denote by $\mathbb{F}^{\text{OVS}} : \mathbb{C}^{M \times N} \rightarrow \mathbb{C}^{G^\theta \times G^\tau}$ the oversampled 2D DFT and by $\mathbb{F} : \mathbb{C}^{G^\theta \times G^\tau} \rightarrow \mathbb{C}^{G^\theta \times G^\tau}$ the usual 2D DFT in dimension $G^\theta \times G^\tau$, where $\frac{G^\theta}{M}$ and $\frac{G^\tau}{N}$ denote the oversampling factors in the angle and delay domain respectively. Note that for an $M \times N$ matrix \mathbf{H} , we have $\mathbb{F}^{\text{OVS}}(\mathbf{H}) = \mathbb{F}(\mathbf{H}^{\text{OVS}})$ where \mathbf{H}^{OVS} denotes a $G^\theta \times G^\tau$ matrix that has \mathbf{H} in its up-left corner and is zero elsewhere. This follows from the well-known property of DFT, where an oversampling in one domain can be obtained by zero-padding in the corresponding transform domain. For simplicity, we assume that \mathbb{F} is normalized such that it is an isometry preserving the matrix Frobenius norm, i.e., $\|\mathbb{F}(\mathbf{L})\| = \|\mathbf{L}\|$ for any $G^\theta \times G^\tau$ matrix \mathbf{L} . We define the following cost function for \mathbf{P} and \mathbf{Q}

$$\begin{aligned} c(\mathbf{P}, \mathbf{Q}) &= \frac{1}{2} \|\mathbf{X}_s - \mathbf{S}_s^a (\mathbf{P} + \mathbf{Q}) \mathbf{S}_s^{fH}\|^2 \\ &\quad + \mathbb{J}_{\mathcal{M}^h}(\mathbb{F}^{\text{OVS}}(\mathbf{P})) + \mathbb{J}_{\mathcal{M}^i}(\mathbb{F}^{\text{OVS}}(\mathbf{Q})) \end{aligned} \quad (19)$$

where $\mathbb{J}_{\mathcal{M}^h}, \mathbb{J}_{\mathcal{M}^i} : \mathbb{C}^{G^\theta \times G^\tau} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ are convex regularizers penalizing those nonzero coefficients of their arguments

not belonging to the masks \mathcal{M}^{h} and \mathcal{M}^{i} respectively. A simple regularizer is the indicator function of a mask \mathcal{D} , given by:

$$\mathbb{I}_{\mathcal{D}}(\mathbf{K}) := \begin{cases} \infty & \text{if } \mathbf{K}(\sim \mathcal{D}) \neq \mathbf{0}, \\ 0 & \text{otherwise,} \end{cases} \quad (20)$$

where $\mathbf{K}(\sim \mathcal{D})$ denotes those elements of the matrix \mathbf{K} not belonging to \mathcal{D} . The cost function in (19) is convex and its globally optimal solution $(\mathbf{P}^*, \mathbf{Q}^*)$ can be found via convex optimization techniques. The optimal solution \mathbf{P}^* of (19) is an estimate of the decontaminated channel matrix \mathbf{H}_s . In the presence of antenna and frequency sampling, this technique (masking and optimization) provides an interpolation scheme to recover the whole channel matrix from its subsamples. In Appendix, we propose a low-complexity algorithm for solving (19) using *Alternating Direction Method of Multipliers* (ADMM), which estimates/interpolates the decontaminated channel matrix with a complexity $O(G \log_2(G))$, where $G = G^\theta G^\tau$ denotes the total number of points in the grid \mathcal{G} . This provides a low-complexity implementation of the MMSE smoothing filter proposed in Section V-C.

B. Low-Complexity Pilot Decontamination Under the Angular Separability Condition

In this section, we explain another pilot decontamination algorithm that has much lower complexity than the MMSE estimator (17) proposed in Section V-C but, to eliminate pilot contamination, it requires a stronger condition that marginal PSF $\gamma_d(\theta)$ and $\gamma_{co}(\theta)$ of the user and its CPUs have approximately disjoint supports in the angular domain, where we define $\gamma_d(\theta) = \int_0^{\Delta\tau_{\max}} \gamma_d(\theta, d\tau)$ with a similar definition holding for $\gamma_{co}(\theta)$. This is illustrated qualitatively in Fig. 6. Notice that the separability of $\gamma_d(\theta, \tau)$ and $\gamma_{co}(\theta, \tau)$ in the joint angle-delay domain is still necessary to successfully decompose (cluster) the PSF $\gamma(\theta, \tau)$ into its signal and interference components $\gamma_d(\theta, \tau)$ and $\gamma_{co}(\theta, \tau)$.

Let \mathbf{H}_s and $\mathbf{H}_{s,j}$, $j \in \mathcal{K}$, be the channel matrices of a user and of its CPUs, and let $\mathbf{E}_s = \sum_{j \in \mathcal{K}} \mathbf{H}_{s,j}$ be the channel matrix of CPU interference. Let $\mathbf{C}_{\text{h}}^* = \int \gamma_d(d\theta) \mathbf{a}(\theta) \mathbf{a}(\theta)^H$ be the estimated covariance matrix of the channel vector $\mathbf{h}_s = \text{vec}(\mathbf{H}_s)$ from the estimated PSF $\gamma_d(\theta, \tau)$ obtained from the clustering. It is not difficult to check that \mathbf{C}_{h}^* is a block-Toeplitz matrix, which implies that every column $\mathbf{h}_s[\omega]$ of \mathbf{H}_s is an M -dim Gaussian vector with a covariance matrix well approximated by $\mathbf{C}_{\text{h}}^* = \int \gamma_d(d\theta) \mathbf{a}(\theta) \mathbf{a}(\theta)^H$, where \mathbf{C}_{h}^* is an $M \times M$ Toeplitz matrix and corresponds to the diagonal block of \mathbf{C}_{h}^* . Similarly, every column $\mathbf{e}_s[\omega]$ of the CPU interference \mathbf{E}_s is an M -dim Gaussian vectors with a Toeplitz covariance matrix given by $\mathbf{C}_{\text{e}}^* = \int \gamma_{co}(d\theta) \mathbf{a}(\theta) \mathbf{a}(\theta)^H$. Let $\mathbf{Y}_s = \mathbf{H}_s + \mathbf{E}_s + \mathbf{Z}_s$ be the received noisy and pilot contaminated signal. For simplicity, we first assume that there is no antenna or frequency sampling and \mathbf{Y}_s is fully available. We consider the following suboptimal scheme for pilot decontamination: Instead of estimating the whole channel matrix \mathbf{H}_s from \mathbf{Y}_s , as we did for the MMSE estimation in Section V-C, we estimate each column $\mathbf{h}_s[\omega]$ of \mathbf{H}_s from the corresponding column $\mathbf{y}_s[\omega]$ of \mathbf{Y}_s individually. Since $\mathbf{y}_s[\omega] = \mathbf{h}_s[\omega] + \mathbf{e}_s[\omega] + \mathbf{z}_s[\omega]$, this

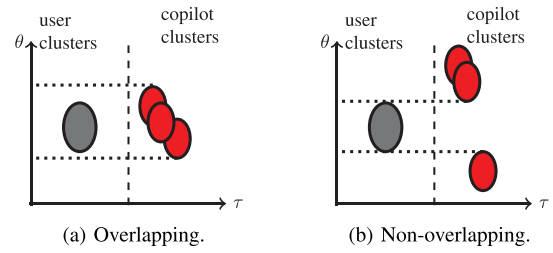


Fig. 6. Illustration of two extremes of angular overlap of the PSF of the desired user $\gamma_d(\theta)$ and that of its CPUs $\gamma_{co}(\theta)$: Overlapping (a) and Non-overlapping (b).

is a standard problem of estimating a Gaussian M -dim vector $\mathbf{h}_s[\omega]$ in an additive colored Gaussian noise $\mathbf{e}_s[\omega] + \mathbf{z}_s[\omega]$. The resulting MMSE estimator can be simply written as

$$\begin{aligned} \hat{\mathbf{h}}_s[\omega] &= \mathbf{\Sigma}_{\text{hy}} \mathbf{C}_{\text{y}}^{-1} \mathbf{y}_s[\omega] \\ &\approx \mathbf{C}_{\text{h}}^* \left(\sigma^2 \mathbf{I}_M + \mathbf{C}_{\text{h}}^* + \mathbf{C}_{\text{e}}^* \right)^{-1} \mathbf{y}_s[\omega], \end{aligned} \quad (21)$$

where $\mathbf{\Sigma}_{\text{hy}} = \mathbb{E}[\mathbf{h}_s[\omega] \mathbf{y}_s[\omega]^H]$ denotes the cross correlation matrix of $\mathbf{h}_s[\omega]$ and $\mathbf{y}_s[\omega]$. It is seen that the MMSE estimator is an $M \times M$ linear operator, which requires computing the inverse of an $M \times M$ Toeplitz matrix rather than an $MN \times MN$ block-Toeplitz matrix, as was necessary for the joint MMSE estimator in Section V-C. More importantly, since the spatial correlation of the channel is invariant with the subcarrier index ω due to the stationarity in the frequency domain, the linear estimator is the same for all the columns of the channel matrix, thus, it needs to be computed only once.

If in addition there is an antenna sampling via an operator \mathbf{S}_s^a , letting $\mathbf{x}_s[\omega] = \mathbf{S}_s^a \mathbf{y}_s[\omega]$ to be the m -dim sketch at subcarrier ω after antenna sampling, the MMSE estimator of $\mathbf{h}_s[\omega]$ from $\mathbf{x}_s[\omega]$ takes on the form

$$\begin{aligned} \hat{\mathbf{h}}_s[\omega] &= \mathbf{\Sigma}_{\text{hx}} \mathbf{C}_{\text{x}}^{-1} \mathbf{x}_s[\omega] \\ &\approx \mathbf{C}_{\text{h}}^* \mathbf{S}_s^{aH} \left(\sigma^2 \mathbf{I}_m + \mathbf{S}_s^a (\mathbf{C}_{\text{h}}^* + \mathbf{C}_{\text{e}}^*) \mathbf{S}_s^{aH} \right)^{-1} \mathbf{x}_s[\omega]. \end{aligned} \quad (22)$$

When the channel matrices of several users are learned over the same OFDM symbol, only a subset of columns of \mathbf{Y}_s is observed for each user. In such a case, we apply the column-wise pilot-decontamination in (21) or (22) to estimate the corresponding columns of \mathbf{H}_s . Then, we apply traditional channel interpolation methods to reconstruct the remaining columns of \mathbf{H}_s from the estimated ones (e.g., via piecewise constant, linear, or DFT-based interpolation techniques [32], [33]). The proposed suboptimal pilot decontamination reduces the implementation complexity considerably. However, the drawback is that in contrast with $\gamma_d(\theta, \tau)$ and $\gamma_{co}(\theta, \tau)$, which are usually well-separable in the joint angle-delay domain, $\gamma_d(\theta)$ and $\gamma_{co}(\theta)$ might generally overlap in the angle domain. In such a case, the dominant subspaces of \mathbf{C}_{h}^* and \mathbf{C}_{e}^* will be highly overlapping, and the suboptimal MMSE will eliminate a significant fraction of the power of the columns of the channel matrix \mathbf{H}_s lying in the interference subspace \mathbf{C}_{e}^* , which results in a poor design of the final beamforming matrix.

In practice, since the number of CPUs is typically small, if the users have a limited angular support and are quite

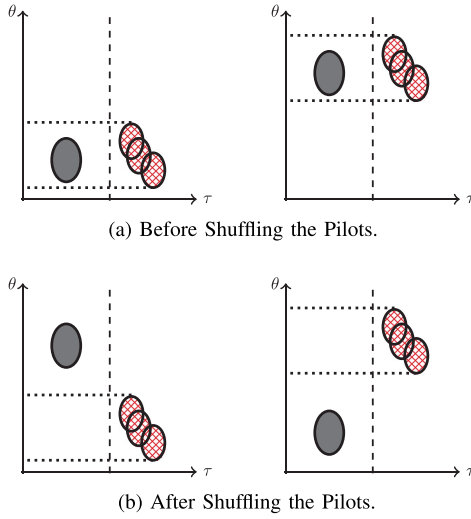


Fig. 7. Separability in the angle domain via shuffling the pilots among the users: Before shuffling the pilots (a) and after shuffling the pilots (b).

randomly distributed inside the cell, there is a high chance that the effective overlap between $\gamma_d(\theta)$ and $\gamma_{co}(\theta)$ be quite negligible for most users. Another way to make $\gamma_d(\theta)$ and $\gamma_{co}(\theta)$ non-overlapping consists of shuffling the pilots assigned to the active users across the whole system as proposed in [10]. This is illustrated qualitatively in Fig. 7, where by re-allocating the pilot of the users of interest the BS can induce angular separation with respect to the CPUs. Pilot-shuffling requires some coordination among neighboring BSs inside the system. In [10] and [13], it is assumed that the PSFs or the covariance matrices \mathbf{C}_h and \mathbf{C}_e of all the users and their CPUs are available. In contrast, in this paper we estimate the PSFs by using wideband pilots and exploiting the sparsity in the angle-delay domain, and identify the signal and interference PSFs by applying suitable clustering algorithms. Hence, our scheme to identify \mathbf{C}_h and \mathbf{C}_e directly from the pilot data can be seen as an enabler for the coordinated pilot shuffling scheme in [10] and the pilot decontamination in [13].

VII. SIMULATION RESULTS

In this section, we assess the performance of our proposed pilot decontamination and channel interpolation algorithm via numerical simulations.

A. Cellular Geometry and Antenna Model

We consider a cellular system consisting of hexagonal cells of radius $R_{\text{cell}} = 1.5 \text{ Km}$ and a maximum tolerable delay spread of $\Delta\tau_{\text{max}} = \frac{2R_{\text{cell}}}{c_0} = 10 \mu\text{s}$. For simulations, we assume the transmit/receive power decays with a power-loss exponent $\eta = 3.2$ (for large cells), where the SNR before beamforming for a user located at a distance r from the BS is given by $\text{SNR}(r) = \frac{\text{SNR}_{\text{max}}}{1 + (\frac{r}{r_0})^\eta}$, where $r_0 = 500 \text{ m}$, and where SNR_{max} is selected such that the SNR before beamforming for a user located at the cell boundary is $\text{SNR}_{\text{min}} = 5 \text{ dB}$. We repeat the simulations for $\eta = 2$ (small cells) to intensify the effect of interference, especially copilot interference, received from the

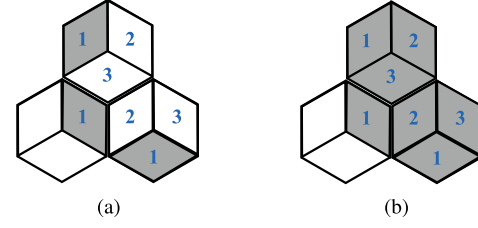


Fig. 8. Pilot distribution for a system with PR3 (a) and PR1 (b), with copilot sectors highlighted in gray.

users in adjacent cells. We normalize the SNR such that the SNR before beamforming for a user close to the BS remains the same in both scenarios.

We assume that each hexagonal cell is divided into 3 sectors as illustrated in Fig. 8. The BS uses a ULA with M antennas to serve the users inside each sector, thus, the whole BS transmitter consists of 3 ULAs (one per sector). The ULAs are well isolated in the RF domain such that each ULA only receives the signal of the users lying in its 120° angular span $[-\theta_{\text{max}}, \theta_{\text{max}}]$ with $\theta_{\text{max}} = 60$ degrees.

B. Scattering Model

We consider a one-ring scattering model for the user signal, where the transmitted signal from a user in the UL is reflected by a ring of scatterers located around the user with a radius of $R_{\text{one-ring}} = 150 \text{ meters}$. We assume that all the scatterers contribute equally in terms of scattering power to the channel vector of the user observed at the BS. Thus, all the users have an equal delay-span of $\frac{2R_{\text{one-ring}}}{c_0} = 1 \mu\text{s}$ but different angular spreads depending on their distance from the BS.

C. Physical Channel Model and OFDM Parameters

We use a physical channel model similar to LTE (Long-Term Evolution) as in [2]. We consider a slot of duration $T_s = 0.532 \text{ ms}$ and decide arbitrarily to send 7 OFDM symbols over each slot, thus, each OFDM symbol has a total duration of $\frac{T_s}{7} = 76 \mu\text{s}$ and an effective duration $T_u = \frac{T_s}{7} - \Delta\tau_{\text{max}} = 66 \mu\text{s}$ after removing the CP, corresponding to a frequency spacing of $\Delta f = \frac{1}{T_u} = 15 \text{ KHz}$ between subcarriers. We take a bandwidth of $W = 2 \text{ MHz}$ with a frequency guard-band of 80 KHz , thus, the total number of subcarriers in each OFDM symbol is given by $N = \frac{1.92 \text{ MHz}}{15 \text{ KHz}} = 128$.

Assuming a coherence bandwidth of $\Delta f_c = 150 \text{ KHz}$, the number of subcarriers in each coherence sub-block (see Fig. 2) is $D_{\text{OFDM}}^c = \Delta f_c T_u \approx 10$. Thus, the wideband channel matrix of 10 users can be simultaneously learned over an individual training OFDM symbol, which enforces a subcarrier sampling ratio $\frac{1}{10}$, i.e., we can sample only $n = \lfloor \frac{N}{10} \rfloor = 12$ out of $N = 128$ subcarriers of an OFDM symbol. We devote 3 OFDM symbols to channel estimation, where we are able to learn the channel matrix and, hence, serve up to $D_p^c = 30$ users on a single TDD slot, consistently with the LTE-TDD standard.

We simulate a sectorized cellular system with each cell consisting of 3 sectors numbered $\{1, 2, 3\}$ as illustrated in Fig. 8.

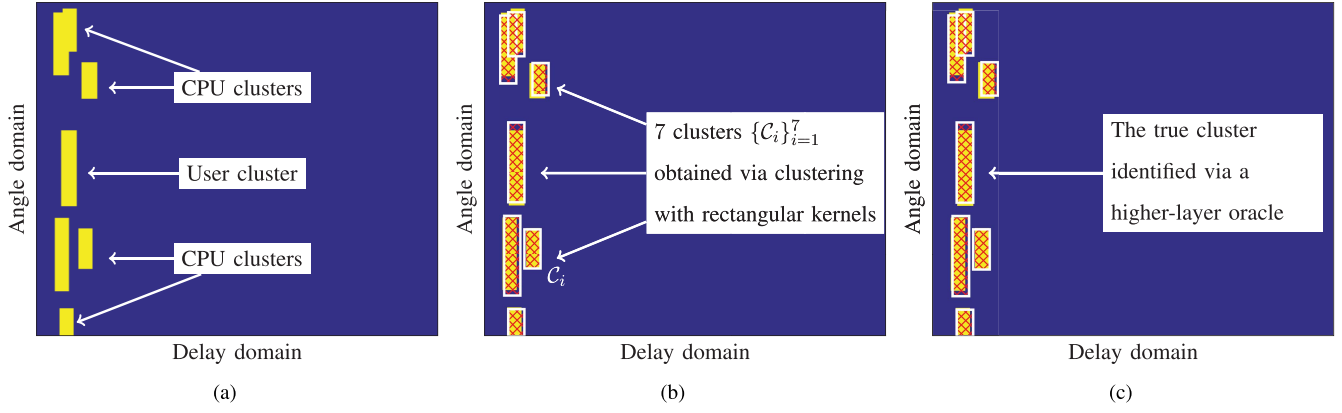


Fig. 9. Illustration of the supervised clustering in one-ring model. Fig.(a) illustrates the estimated angle-delay PSF, where it is not initially known which cluster belongs to the desired user. Fig.(b) denotes a simple clustering of the estimated PSF into 7 rectangular kernels (1 for the user and 6 for its CPUs). Fig.(c) illustrates the identification of the true signal cluster via a higher-layer communication protocol.

We consider a system with a *Pilot Reuse 3* (PR3) as illustrated in Fig. 8a, in which the set of $D_p^c = 30$ orthogonal pilots are shared among 3 sectors such that sectors with similar numbers use identical set of pilots consisting of 10 mutually orthogonal pilot sequences (i.e., $K = 10$ served users in each sector), and groups of 3 adjacent sectors with different indices (1,2,3) use collectively all the 30 orthogonal pilot sequences. We also consider a system with a *Pilot Reuse 1* (PR1) as illustrated in Fig. 8b, in which all the 30 orthogonal pilot sequences are simultaneously used in all the sectors, thus, each sector can serve up to $K = 30$ users.

D. Clustering Algorithm

Since in PR3 illustrated in Fig. 8a users and their CPUs are well-separated in the delay domain, we apply an unsupervised clustering in the delay domain as in Fig. 4 with a delay threshold $\tau_0 = \frac{R_{\text{cell}}}{c_0}$. In particular, we set $\gamma_d(\theta, \tau) = \gamma(\theta, \tau)\mathbf{1}_{\{(\theta, \tau) \in C_d^{\text{PR3}}\}}$ and $\gamma_{\text{co}}(\theta, \tau) = \gamma(\theta, \tau)\mathbf{1}_{\{(\theta, \tau) \notin C_d^{\text{PR3}}\}}$, where C_d^{PR3} is the desired signal cluster given by $C_d^{\text{PR3}} = \{(\theta, \tau) : \tau \leq \tau_0\}$. For PR1, the users and their CPUs are not generally separable in the delay domain (see, e.g., Fig. 9a). Here, we need to apply a supervised clustering algorithm to identify the desired user cluster. In the one-ring scattering model we consider for the simulations, the PSF of each user consists of a single angle-delay cluster (bubble). Since the number of CPUs is at most 6, we cluster the estimated PSF into 7 rectangular-shaped clusters illustrated in Fig. 9b. This separates approximately the clusters corresponding to the user and its CPUs but does not specify yet which cluster corresponds to the user. To identify the user cluster, we use the “oracle” provided from a higher communication layer with the following scheme. After receiving the noisy contaminated channel sketch \mathbf{z}_s during a pilot transmission slot, we obtain estimates $\hat{\mathbf{h}}_s^i$, $i = 1, 2, \dots, 7$, of the decontaminated channel vector of the user by treating the i -th cluster as the true signal cluster and the rest as CPU clusters, and applying our proposed channel interpolation algorithm in Section VI-A. During the data transmission phase, after receiving the whole OFDM symbol, we decode the received data by beamforming

along the columns of the channel matrix corresponding to $\hat{\mathbf{h}}_s^i$, $i = 1, \dots, 7$, once at a time, where we assume that there is a higher-layer oracle that selects the $\hat{\mathbf{h}}_s^i$ and the corresponding cluster that results in a successful decoding of the user data as illustrated in Fig. 9c.

E. Uplink Pilot Decontamination

For simulations, we focus on pilot decontamination in an UL scenario, where we focus on the users belonging to Sector 1 as in Fig. 8. The dominant copilot interference for each one of those users in the UL comes from its 2 nearest neighbor CPUs in PR3 and from its 6 nearest neighbor CPUs in PR1. For each user, the BS learns the superposition of the channel vector of that user plus those of its CPUs. Note that due to the orthogonality of the pilots, during the UL training phase, there is only copilot interference but no interference from the other users. During the data transmission phase (UL or DL), however, there is a coherent interference from CPUs and a noncoherent interference from all the other users. Notice also that CPUs coming from non-nearest neighbor copilot sectors are received at significantly lower power and at larger delays. Such signals are not guaranteed to be eliminated by the proposed method since the OFDM model fails due to inter-block interference (MPCs whose delays go beyond the CP interval). Nevertheless, the effect of non-nearest neighbors copilot contamination is very small.

F. Antenna Sampling and Wideband Pilot Sketches

We consider an antenna sampling ratio of 0.25, where over each OFDM training symbol only $m = 0.25 M$ of the whole number of antennas M are sampled. We assume that the sampling pattern is completely random and changes i.i.d. over time. We take a window of size $w = 100$ of sketches across w time slots to estimate the channel geometry of each user, where we assume that the channel matrices inside the window are i.i.d. since they belong to different slots (coherence times). As the whole observation takes $wT_s = 50$ ms, we can safely assume that the channel geometry remains invariant over the whole window.

G. Pilot Decontamination, Channel Interpolation, and Beamforming

After estimating the PSF of all the users, we apply the clustering algorithm explained in Section VII-D and the masking technique as in (18) to obtain the signal mask \mathcal{M}^h and the interference mask \mathcal{M}^i for each user, which we use for the rest of the time. We next simulate the communication phase, where each CB consists of a training phase to estimate the instantaneous channel vectors of the user and a data-transmission phase to send data to these users via spatial beamforming. In each training slot s , after receiving a sketch of the channel vector of each user, we apply the low-complexity channel interpolation and decontamination algorithm in Section VI-A to estimate the full channel vector of the user. For simplicity of comparison with the contaminated case, we assume no antenna sampling is applied during a training slot. We denote the decontaminated channel vectors of the K users at the reference BS by $\{\hat{\mathbf{h}}_k\}_{k=1}^K$ and the corresponding channel matrices by $\{\hat{\mathbf{H}}_k\}_{k=1}^K$, where for simplicity we dropped the dependence on the data transmission slot s . We also assume that the noise power σ^2 in each antenna is available at the BS.

In PR3, we apply the MMSE beamforming for each user in the UL, where the normalized beamforming vector for a user $k \in [K]$ at subcarrier ω is given by $\mathbf{g}_k[\omega] = \frac{\mathbf{v}_k[\omega]}{\|\mathbf{v}_k[\omega]\|}$, where

$$\mathbf{v}_k[\omega] = \left(\sigma^2 \mathbf{I}_M + \sum_{k'} \hat{\mathbf{h}}_{k'}[\omega] \hat{\mathbf{h}}_{k'}^H[\omega] \right)^{-1} \hat{\mathbf{h}}_k[\omega], \quad (23)$$

where $\hat{\mathbf{h}}_k[\omega]$ denotes the decontaminated and interpolated channel vector of the user k at subcarrier ω . Due to the sectorization, the BS in Sector 1 not only receives the pilot signal of its users but can also listen to the pilot signal of the users in adjacent sectors since they are using disjoint set of pilots. Thus, the summation over k' in (23) is taken over all the users inside the sector as well as the users in adjacent sectors. For PR1, we use a simple *conjugate beamforming* [2] given by $\mathbf{g}_k[\omega] = \frac{\hat{\mathbf{h}}_k[\omega]}{\|\hat{\mathbf{h}}_k[\omega]\|}$. We compare the performance of our

method with the case where no pilot decontamination is applied. We assume that in such a case also 1 out of 10 of columns of channel matrix of each user is observed during a training slot. We apply DFT interpolation to interpolate the unobserved columns of the channel matrix of each user and repeat similar steps, as in the decontaminated case, to design the beamforming vectors. We define the SINR of the channel of the user $k \in [K]$ at subcarrier ω by

$$\text{SINR}_k[\omega] = \frac{|\mathbf{g}_k[\omega]^H \mathbf{h}_k[\omega]|^2}{\sigma^2 + \sum_{k' \neq k} |\mathbf{g}_k[\omega]^H \mathbf{h}_{k'}[\omega]|^2}, \quad (24)$$

where the summation is taken over all the users $k' \neq k$ inside the sector as well as all other users in adjacent cells who lie in the angular span of the ULA of Sector 1 and create interference. Assuming perfect channel state information after beamforming, the *instantaneous spectral efficiency* of user k is given by $R_k = \frac{1}{N} \sum_{\omega \in [N]} \log_2(1 + \text{SINR}_k[\omega])$. We denote the achievable sum-rate in bit/s/Hz of all the users in Sector 1 by $R = \sum_{k \in [K]} R_k$, where R is a random variable depending

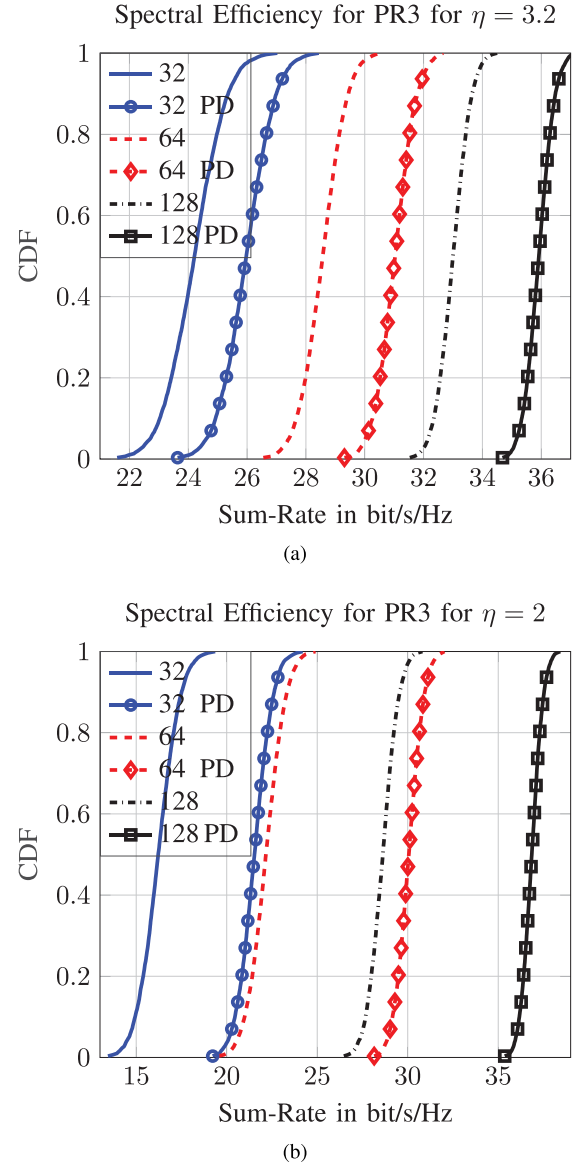


Fig. 10. CDF of Spectral Efficiency (Sum-Rate) in bit/s/Hz for different number of BS antennas $M \in \{32, 64, 128\}$ for PR3 with $\eta = 3.2$ (a) and $\eta = 2$ (b). The curves with marks and with the legend "PD" illustrate the CDF after pilot decontamination. All the plots are averaged over $N_{\text{geom}} = 30$ random user locations in the system.

on the instantaneous realizations of the channel vectors of all the users.

H. Achievable Performance and Comparison With the State of the Art

1) PR3: Fig. 10 illustrates the *Cumulative Distribution Function* (CDF) $F_R(r)$ of the achievable spectral efficiency R in bit/s/Hz before and after pilot decontamination. We average the CDFs over $N_{\text{geom}} = 30$ independent realizations of the geometry of users across the system. For each geometry realization, we run the simulations for different number of BS antennas $M \in \{32, 64, 128\}$. We also consider two different scenarios for two different power-loss exponents $\eta \in \{3.2, 2\}$ as explained in Section VII-A. It is seen that for $\eta = 3.2$ and for practical numbers of BS antennas $M \in \{32, 64, 128\}$

pilot decontamination improves the spectral efficiency by 10% – 20% for PR3, where the resulting gain increases by increasing the number of BS antennas M . For $\eta = 2$, on the other hand, our proposed scheme results in a dramatic gain in spectral efficiency.

2) *PR1*: We repeat the simulations for PR1. In this case, to pinpoint the effect of pilot contamination, rather than calculating the sum-rate averaged over all the random locations of the users, we focus on an “edge” user randomly located on the cell boundary. We expect that the spectral efficiency of such a user be affected considerably by the pilot contamination from the neighboring CPUs. As illustrated in Fig. 9, pilot decontamination for edge users in PR1 requires a supervised clustering in the angle-delay domain. Fig. 11 illustrates the simulation results. We compare the performance of our algorithm with the one proposed in [17]. In [17], the support of the MPCs of the contaminated channel vector of the user is estimated by devoting all the subcarriers in an OFDM symbol to an individual user and projecting the whole channel matrix in the 2D FFT basis. The support of the desired user is identified and separated from that of its CPUs by taking the intersection of the support obtained over several slots, where in each slot the pilots are shuffled such that the desired user collides with different CPUs at each slot. The rationale behind this idea is that in this way the support of the MPCs of the desired user remains constant over the sequence of slots, while that of the CPUs changes from slot to slot. Therefore, taking the intersection of the estimated supports over the slots should yield the MPCs of the desired user. However, in doing so, the intersection will also exclude the MPCs of the desired user that over the sequence of slots experience a deep fade, since these will be missed on some slots, and therefore will not be contained in the intersection. As a matter of fact, with time-selective fading as in our realistic setting, we could verify that the method of [17] dramatically underestimates the MPCs of the desired user.

In contrast, our proposed subspace estimation with supervised clustering is much robust to small-scale fading variations, performs much better in the presence of overlapping clusters (e.g., when a user and its CPUs have common clusters), and does not require pilot shuffling among the neighboring cells. Also, compared with [17], in our proposed scheme only a fraction (e.g., $\frac{1}{10}$) of the subcarriers in an OFDM symbol are devoted as pilot to each user, so the channel state of several users (e.g., 10 users) can be simultaneously estimated over an individual OFDM symbol, thus, much better multiplexing gain. Notice that using the whole set of N OFDM subcarriers for UL pilots is essential to the method of [17] since otherwise there is not enough resolution in the delay domain. This is because [17] makes use of simple linear projections, while our scheme estimates the PSF using the advanced $l_{2,1}$ -regularized least squares minimization described in Section IV.

Our simulation results in Fig. 11 consider the rate CDF of a single edge user randomly located near the cell boundary, thus, they do not reflect the additional multiplexing gain resulting from using a reduced pilot dimension $n \ll N$. For the simulations, we assume that the total number of users is the same in both scenarios (30 users per sector), where similar

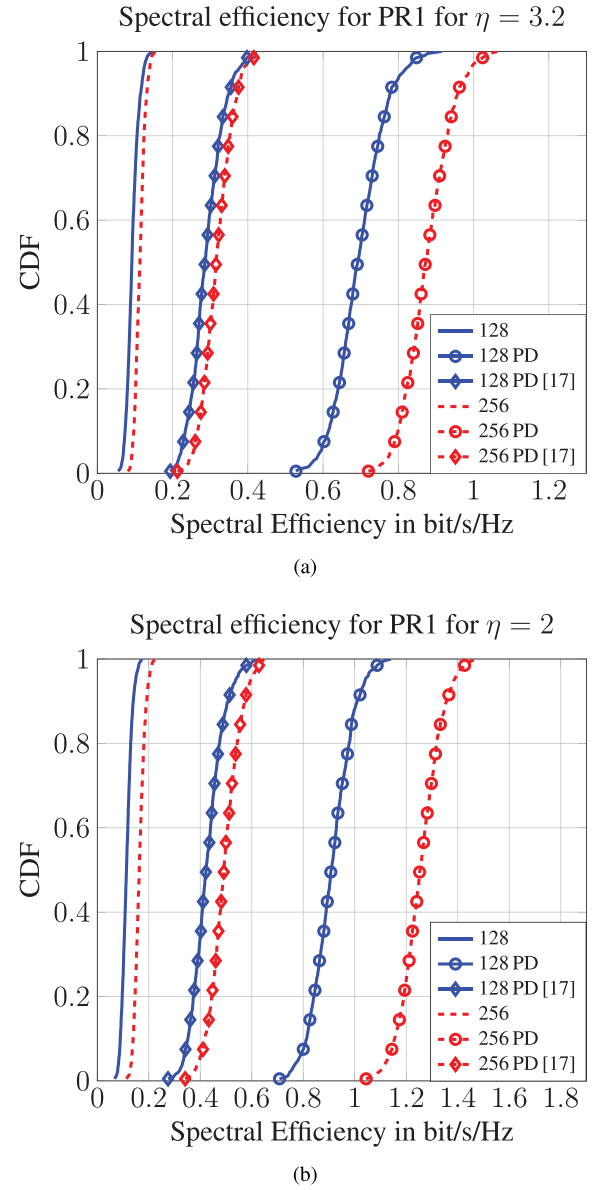


Fig. 11. CDF of Spectral Efficiency in bit/s/Hz of a specific user close to the cell boundary for different number of BS antennas $M \in \{128, 256\}$ in a system with PR1 for $\eta = 3.2$ (a) and $\eta = 2$ (b). The curves with marks illustrate the CDF after pilot decontamination: our algorithm with the legend “PD” and the algorithm in [17] with the legend “PD[17]”. All the plots are averaged over $N_{\text{geom}} = 30$ random user locations in the system.

to PR3 we average the achievable spectral efficiency of this specific edge user over $N_{\text{geom}} = 30$ independent realizations of the geometry of all the users across the whole system. From Fig. 11, it is seen that the gain in spectral efficiency obtained by our method is much more than the one proposed in [17]. In particular, the resulting gain scales much better with the number of BS antennas.

VIII. CONCLUSIONS

In this paper, we presented a novel scheme to eliminate the effect of pilot contamination on the performance of a massive MIMO wireless cellular system. We proposed a low-complexity algorithm that uses the pilot signal received

from each user inside a window containing several time slots to obtain an estimate of the angle-delay power spread function (PSF) of each user contaminated channel vectors. We used the key idea, already exploited in various ways in the recent massive MIMO literature, that the channel vectors of each user consist of sparse MPCs in the angle-delay domain. We exploited this underlying sparsity to estimate the angle-delay PSF of each user by sampling only a small subset of antennas and, more importantly, by transmitting pilots across only a subset of subcarriers compatible with LTE-TDD without incurring any pilot overhead. We proposed clustering algorithms to decompose the estimated PSF of each user into its signal and copilot interference part. We exploited this decomposition to decontaminate the channel vector of each user in the next coherence blocks. Through Monte Carlo simulations, we demonstrated the effectiveness of the proposed pilot-decontamination scheme for practical scenarios with practical user geometries, reasonable number of BS antennas $M \in \{32, 64, 128, 256\}$, and realistic fading channel statistics as in [2]. We also compared our proposed method with the competitive scheme [17] and illustrated that our method provides much better performance in terms of multiplexing gain, pilot decontamination efficiency, and scaling performance with the number of BS antennas.

APPENDIX

LOW-COMPLEXITY INTERPOLATION USING ADMM

Consider the following cost function as in (19):

$$c(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \|\mathbf{X}_s - \mathbf{S}_s^a(\mathbf{P} + \mathbf{Q})\mathbf{S}_s^{\text{FH}}\|^2 + \mathbb{I}_{\mathcal{M}^h}(\mathbf{F}^{\text{OVS}}(\mathbf{P})) + \mathbb{I}_{\mathcal{M}^i}(\mathbf{F}^{\text{OVS}}(\mathbf{Q})). \quad (25)$$

In this section, we assume that the convex regularizers $\mathbb{I}_{\mathcal{M}^h}$ and $\mathbb{I}_{\mathcal{M}^i}$ are the indicator functions of \mathcal{M}^h and \mathcal{M}^i defined similarly to (20). We first introduce the auxiliary variables $\mathbf{P}_f = \mathbf{F}^{\text{OVS}}(\mathbf{P})$ and $\mathbf{Q}_f = \mathbf{F}^{\text{OVS}}(\mathbf{Q})$ of dimension $G^\theta \times G^\tau$ and define

$$c(\mathbf{P}, \mathbf{Q}, \mathbf{P}_f, \mathbf{Q}_f) = \frac{1}{2} \|\mathbf{X}_s - \mathbf{S}_s^a(\mathbf{P} + \mathbf{Q})\mathbf{S}_s^{\text{FH}}\|^2 + \mathbb{I}_{\mathcal{M}^h}(\mathbf{P}_f) + \mathbb{I}_{\mathcal{M}^i}(\mathbf{Q}_f). \quad (26)$$

Thus, minimizing $c(\mathbf{P}, \mathbf{Q})$ in (25) can be equivalently written as minimizing $c(\mathbf{P}, \mathbf{Q}, \mathbf{P}_f, \mathbf{Q}_f)$ under the additional linear constraints $\mathbf{P}_f = \mathbf{F}^{\text{OVS}}(\mathbf{P})$, $\mathbf{Q}_f = \mathbf{F}^{\text{OVS}}(\mathbf{Q})$, which is still a convex optimization problem. We use *Alternating Direction Method of Multipliers* (ADMM) to solve this optimization problem. We introduce the Lagrange variables $\mathbf{\Lambda}_p$ and $\mathbf{\Lambda}_q$ of dimension $G^\theta \times G^\tau$ and the augmented Lagrangian function

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|\mathbf{X}_s - \mathbf{S}_s^a(\mathbf{P} + \mathbf{Q})\mathbf{S}_s^{\text{FH}}\|^2 + \mathbb{I}_{\mathcal{M}^h}(\mathbf{P}_f) + \mathbb{I}_{\mathcal{M}^i}(\mathbf{Q}_f) \\ & + \langle \mathbf{\Lambda}_p, \mathbf{P}_f - \mathbf{F}^{\text{OVS}}(\mathbf{P}) \rangle_{\mathbb{R}} + \langle \mathbf{\Lambda}_q, \mathbf{Q}_f - \mathbf{F}^{\text{OVS}}(\mathbf{Q}) \rangle_{\mathbb{R}} \\ & + \frac{v}{2} \|\mathbf{P}_f - \mathbf{F}^{\text{OVS}}(\mathbf{P})\|^2 + \frac{v}{2} \|\mathbf{Q}_f - \mathbf{F}^{\text{OVS}}(\mathbf{Q})\|^2, \quad (27) \end{aligned}$$

where $v \in \mathbb{R}_+$ is the ADMM parameter to be set. The ADMM iteration can be written as follows:

$$(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}) = \arg \min_{\mathbf{P}, \mathbf{Q}} \mathcal{L}(\mathbf{P}, \mathbf{Q}, \mathbf{P}_f^k, \mathbf{Q}_f^k, \mathbf{\Lambda}_p^k, \mathbf{\Lambda}_q^k), \quad (28)$$

$$(\mathbf{P}_f^{k+1}, \mathbf{Q}_f^{k+1}) = \arg \min_{\mathbf{P}_f, \mathbf{Q}_f} \mathcal{L}(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}, \mathbf{P}_f, \mathbf{Q}_f, \mathbf{\Lambda}_p^k, \mathbf{\Lambda}_q^k), \quad (29)$$

$$\mathbf{\Lambda}_p^{k+1} = \mathbf{\Lambda}_p^k + v(\mathbf{P}_f^{k+1} - \mathbf{F}^{\text{OVS}}(\mathbf{P}^{k+1})), \quad (30)$$

$$\mathbf{\Lambda}_q^{k+1} = \mathbf{\Lambda}_q^k + v(\mathbf{Q}_f^{k+1} - \mathbf{F}^{\text{OVS}}(\mathbf{Q}^{k+1})), \quad (31)$$

A. Updating $\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}$

Using the vectorization and denoting by $\mathbf{p} = \text{vec}(\mathbf{P})$, $\mathbf{q} = \text{vec}(\mathbf{Q})$, $\mathbf{x}_s = \text{vec}(\mathbf{X}_s)$, $\mathbf{p}_t = \mathbf{F}^{\text{OVS}-1}(\mathbf{P}_f^k + \frac{1}{v}\mathbf{\Lambda}_p^k)$, $\mathbf{q}_t = \mathbf{F}^{\text{OVS}-1}(\mathbf{Q}_f^k + \frac{1}{v}\mathbf{\Lambda}_q^k)$, $\mathbf{p}_t = \text{vec}(\mathbf{P}_t(1:M, 1:N))$, $\mathbf{q}_t = \text{vec}(\mathbf{Q}_t(1:M, 1:N))$, we can write (28) as the following cost function to be minimized for $MN \times 1$ vectors \mathbf{p} and \mathbf{q} :

$$\frac{1}{2} \|\mathbf{x}_s - \mathbf{S}_s(\mathbf{p} + \mathbf{q})\|^2 + \frac{v}{2} \|\mathbf{p} - \mathbf{p}_t\|^2 + \frac{v}{2} \|\mathbf{q} - \mathbf{q}_t\|^2, \quad (32)$$

where $\mathbf{S}_s = \mathbf{S}_s^f \otimes \mathbf{S}_s^a$ denotes the sampling operator at slot s . The optimal solution of (32) is given by

$$\begin{aligned} \mathbf{p}^{k+1} = & \mathbf{p}_t + \frac{\check{\mathbf{x}}_s}{v} \\ & + \mathcal{S}_s(2\mathcal{S}_s + v\mathcal{I})^{-1}(\mathbf{p}_t + \mathbf{q}_t + \frac{2\check{\mathbf{x}}_s}{v}), \quad (33) \end{aligned}$$

$$\begin{aligned} \mathbf{q}^{k+1} = & \mathbf{q}_t + \frac{\check{\mathbf{x}}_s}{v} \\ & + \mathcal{S}_s(2\mathcal{S}_s + v\mathcal{I})^{-1}(\mathbf{p}_t + \mathbf{q}_t + \frac{2\check{\mathbf{x}}_s}{v}), \quad (34) \end{aligned}$$

where $\check{\mathbf{x}}_s = \mathbf{S}_s^H \mathbf{x}_s$, $\mathcal{S}_s = \mathbf{S}_s^H \mathbf{S}_s$, and where $\mathcal{I} = \mathbf{I}_{MN}$ denotes the identity matrix of order MN . Since in this paper we always use 0-1 antenna and frequency sampling matrices, (33) and (34) can be further simplified. Using the properties of the vec operator, we have that

$$\begin{aligned} \mathcal{S}_s = & \mathbf{S}_s^H \mathbf{S}_s = (\mathbf{S}_s^{\text{FH}} \otimes \mathbf{S}_s^{\text{aH}})(\mathbf{S}_s^f \otimes \mathbf{S}_s^a) \\ = & (\mathbf{S}_s^{\text{FH}} \mathbf{S}_s^f) \otimes (\mathbf{S}_s^{\text{aH}} \mathbf{S}_s^a). \quad (35) \end{aligned}$$

Note that, due to 0-1 sampling, $\mathbf{S}_s^{\text{FH}} \mathbf{S}_s^f$ and $\mathbf{S}_s^{\text{aH}} \mathbf{S}_s^a$ are diagonal matrices of dimension $M \times M$ and $N \times N$ with 1s in the diagonal elements corresponding to the index sets \mathcal{I}_s^f and \mathcal{I}_s^a and 0 elsewhere, where $\mathcal{I}_s^a \subseteq [M]$ and $\mathcal{I}_s^f \subseteq [N]$ denote the indices of antennas and subcarriers sampled at slot s as explained in Section III-C. This implies that \mathcal{S}_s is a 0-1 diagonal matrix of dimension $MN \times MN$ where the locations of 1s in the diagonal is given as in (11) by

$$\mathcal{I}_s = \{M(c_s^f - 1) + c_s^a : c_s^f \in \mathcal{I}_s^f, c_s^a \in \mathcal{I}_s^a\} \subseteq [MN].$$

As a result, the matrix $\mathcal{S}_s(2\mathcal{S}_s + v\mathcal{I})^{-1}$ in (33) and (34) is a diagonal matrix with a value $\frac{1}{v+2}$ at the diagonal elements belonging to \mathcal{I}_s and 0 elsewhere. Moreover, no matrix-vector multiplication is needed for computing $\check{\mathbf{x}}_s = \mathbf{S}_s^H \mathbf{x}_s$ since $\check{\mathbf{x}}_s$ is simply given by an $MN \times 1$ vector that contains the components of \mathbf{x}_s in the indices corresponding to \mathcal{I}_s and is 0 elsewhere. This implies that \mathbf{p}^{k+1} and \mathbf{q}^{k+1} can be easily

computed from (33) and (34), from which we obtain \mathbf{P}^{k+1} and \mathbf{Q}^{k+1} via inverse vec operation. The whole computational complexity of this step comes from calculating \mathbf{p}_t and \mathbf{q}_t , which requires $O(G \log_2(G))$ operations where $G = G^\theta G^\tau$ denotes the grid size as before.

B. Updating $\mathbf{P}_f^{k+1}, \mathbf{Q}_f^{k+1}$

We first derive the update equation for \mathbf{P}_f^{k+1} in (29). To find \mathbf{P}_f^{k+1} , we need to optimize the following function with respect to \mathbf{P}_f :

$$\frac{v}{2} \|\mathbf{P}_f + \frac{1}{v} \mathbf{A}_p^k - \mathbf{F}^{\text{ovs}}(\mathbf{P}^{k+1})\|^2 + \|\mathcal{M}^h(\mathbf{P}_f)\|. \quad (36)$$

The optimal solution of (36) is given by setting \mathbf{P}_f^{k+1} equal to $\mathbf{F}^{\text{ovs}}(\mathbf{P}^{k+1}) - \frac{1}{v} \mathbf{A}_p^k$ at those elements belonging to the mask \mathcal{M}^h while setting the remaining components equal to zero. Similarly, \mathbf{Q}_f^{k+1} is given by $\mathbf{F}^{\text{ovs}}(\mathbf{Q}^{k+1}) - \frac{1}{v} \mathbf{A}_q^k$ over the mask \mathcal{M}^i and zero elsewhere. The whole computational complexity of this step is also $O(G \log_2(G))$ for computing $\mathbf{F}^{\text{ovs}}(\mathbf{P}^{k+1})$ and $\mathbf{F}^{\text{ovs}}(\mathbf{Q}^{k+1})$. Overall, the computational complexity of each ADMM iteration is $O(G \log_2(G))$.

REFERENCES

- [1] S. Haghighatshoar and G. Caire, "Pilot decontamination through compressive wideband channel estimation," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 979–983.
- [2] 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.
- [3] C. Shepard *et al.*, "Argos: Practical many-antenna base stations," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 53–64.
- [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, Feb. 2014.
- [5] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.
- [6] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [7] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprasad, "Achieving 'massive MIMO' spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.
- [8] 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, Feb. 2013.
- [9] 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, Feb. 2016.
- [10] 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.
- [11] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [12] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
- [13] E. Björnson, J. Hoydis, and L. Sanguinetti, (2016). "Pilot contamination is not a fundamental asymptotic limitation in massive MIMO." [Online]. Available: <https://arxiv.org/abs/1611.09152>
- [14] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.
- [15] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.
- [16] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for interference reduction in large scale antenna systems," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2013, pp. 226–232.
- [17] Z. Chen and C. Yang, "Pilot decontamination in wideband massive mimo systems by exploiting channel sparsity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5087–5100, Jul. 2016.
- [18] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. Hoboken, NJ, USA: Wiley, 2011.
- [19] H. Shirani-Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2713–2723, Sep. 2009.
- [20] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [21] A. F. Molisch, *Wireless Communications*, vol. 34. Hoboken, NJ, USA: Wiley, 2012.
- [22] L. Liu *et al.*, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.
- [23] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*. San Francisco, CA, USA: Academic, 2013.
- [24] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wide-band correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.
- [25] G. Auer, "3D MIMO-OFDM channel estimation," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 972–985, Apr. 2012.
- [26] B. H. Fleury, "First- and second-order characterization of direction dispersion and space selectivity in the radio channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2027–2044, Jun. 2000.
- [27] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Jan. 2017.
- [28] S. Haghighatshoar and G. Caire, (2016). "Low-complexity massive MIMO subspace estimation and tracking from low-dimensional projections." [Online]. Available: <https://arxiv.org/abs/1608.02477>
- [29] A. Nemirovski, *Efficient Methods in Convex Programming* (Technical Report Lecture Notes). Haifa, Israel: Technion, Israel Institute of Technology, 1994.
- [30] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 201–205.
- [31] O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "RRH based massive MIMO with 'on the Fly' pilot contamination control," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.
- [32] J. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2083–2088, Sep. 2005.
- [33] A. A. Hutter, R. Hasholzner, and J. S. Hammerschmidt, "Channel estimation for mobile OFDM systems," in *Proc. IEEE VTS 50th Veh. Technol. Conf. (VTC-Fall)*, vol. 1, Sep. 1999, pp. 305–309.



Saeid Haghighatshoar (S'12–M'15) received the B.Sc. degree in electrical engineering (electronics) and the M.Sc. degree in electrical engineering (communication systems) from the Sharif University of Technology, Tehran, Iran, in 2007 and 2009, respectively, and the Ph.D. degree in computer and communication sciences from the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2014. Since 2015, he has been a Post-Doctoral Researcher with the Communications and Information Theory Group, Technische Universität Berlin, Berlin, Germany. His research interests lie in information theory, communication systems, wireless communication, optimization theory, and compressed sensing.



Giuseppe Caire (S'92–M'94–SM'03–F'05) was born in Turin, Italy, in 1965. He received the B.Sc. degree in electrical engineering from the Politecnico di Torino, Italy, in 1990, the M.Sc. degree in electrical engineering from Princeton University in 1992, and the Ph.D. degree from the Politecnico di Torino in 1994. He was a Post-Doctoral Research Fellow with the European Space Agency, Noordwijk, The Netherlands, from 1994 to 1995, an Assistant Professor in telecommunications with the Politecnico di Torino, an Associate Professor with the University of Parma, Italy, a Professor with the Department of Mobile Communications, the Eurecom Institute, Sophia-Antipolis, France, and a Professor of electrical engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA. He is currently

an Alexander von Humboldt Professor with the Electrical Engineering and Computer Science Department, Technical University of Berlin, Germany.

His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, respectively, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, and the Vodafone Innovation Prize in 2015. He served on the Board of Governors for the IEEE Information Theory Society from 2004 to 2007 and was an officer from 2008 to 2013. He was the President of the IEEE Information Theory Society in 2011. He served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1998 to 2001 and an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003.