This is a repository copy of *Superimposed Pilots are Superior for Mitigating Pilot Contamination in Massive MIMO*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/112531/

Version: Accepted Version

Superimposed Pilots are Superior for Mitigating Pilot Contamination in Massive MIMO

Karthik Upadhya, *Student Member, IEEE*, Sergiy A. Vorobyov, *Senior Member, IEEE*,
Mikko Vehkapera, *Member, IEEE*

Abstract—In this paper, superimposed pilots are introduced as an alternative to time-multiplexed pilot and data symbols for mitigating pilot contamination in massive multiple-input multiple-output (MIMO) systems. We propose a non-iterative scheme for uplink channel estimation based on superimposed pilots and derive an expression for the uplink signal-to-interference-plus-noise ratio (SINR) at the output of a matched filter employing this channel estimate. Based on this expression, we observe that power control is essential when superimposed pilots are employed. Moreover, the quality of the channel estimate can be improved by reducing the interference that results from transmitting data alongside the pilots, and an intuitive iterative data-aided scheme that reduces this component of interference is also proposed. Approximate expressions for the uplink SINR are provided for the iterative data-aided method as well. In addition, we show that a hybrid system with users utilizing both time-multiplexed and superimposed pilots is superior to an optimally designed system that employs only time-multiplexed pilots, even when the non-iterative channel estimate is used to build the detector and precoder. We also describe a simple approach to implement this hybrid system by minimizing the overall inter and intra-cell interference. Numerical simulations demonstrating the performance of the proposed channel estimation schemes and the superiority of the hybrid system are also provided.

Index Terms—Massive MIMO, pilot contamination, superimposed pilots.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems, proposed in [1], have received significant interest in recent years as a candidate for fifth-generation mobile communication technologies [2]–[4]. These systems are a variation of multi-user MIMO (MU-MIMO) and have a large number of base station (BS) antennas that result in an improved spectral efficiency through spatial multiplexing [5], [6]. Under favorable propagation conditions [1], significant gains in throughput can be achieved by employing simple linear processing at the

BS [7], [8]. In addition, large numbers of antennas result in an improved uplink (UL) energy-efficiency [9] and render the system performance resilient to hardware impairments [10].

However, all the above mentioned benefits of a massive MIMO communication system hinge on the assumption that the BS has access to accurate estimates of the channel state information (CSI). For systems that employ either frequency division duplexing (FDD) or time division duplexing (TDD), the channel estimates are obtained using orthogonal pilot sequences. In FDD systems, each antenna at the BS transmits a pilot sequence that is orthogonal to the pilot sequences transmitted by the other antennas. Since the number of orthogonal pilot sequences required becomes proportional to the number of BS antennas, FDD is considered impractical for channel estimation in massive MIMO [5], [11]. Moreover, since the CSI corresponding to each antenna is estimated by the users, it has to be fedback from the users to the BS, consuming additional bandwidth. Consequently, massive MIMO systems are typically assumed to employ TDD with full frequency-reuse and utilize channel reciprocity to obtain CSI. In these systems, each user in a cell is assigned a different pilot sequence and these pilots are time-multiplexed with data in each coherence block. When using time-multiplexed pilots and data, the requirement for high spectral efficiency necessitates sharing of pilot sequences between adjacent cells, resulting in the channel estimates of the users in a cell being corrupted by the channel vectors of users in the adjacent cells. This phenomenon called 'pilot contamination' [12] introduces interference in both the UL and downlink (DL), degrading the overall performance of the system.

Existing methods to mitigate pilot contamination for massive MIMO are designed for the case wherein the pilots are time-multiplexed with the data, henceforth referred to as time-multiplexed pilots. In [13], it has been observed that the eigenvectors of the autocorrelation matrix of the received data correspond to the channel vectors of the desired and interfering users, and a method for channel estimation has been developed based on this observation. Pilot decontamination has been performed in [14] by projecting the contaminated channel estimate on an interference-free subspace spanned by the channel vectors of the desired users, whereas [15] derives asymptotic conditions for separability between the subspaces of the desired and interfering users. In [16], a coordinated method for pilot allocation has been proposed for separating desired and interfering users in correlated channels. A pilot decontamination method based on the array processing model has been proposed in [17] for use in parametric channels with

a finite number of discrete paths. In [18], a resource allocation approach has been proposed for optimizing the number of users scheduled in each cell in order to minimize the effect of pilot contamination. A common theme for the approaches described above, except for [16] and [18], is that they focus on decontaminating the channel estimate at the receiver, which in this case is the BS. However, since pilot contamination results from interfering pilot transmissions, there is a scope for *better separating the desired and interfering users by optimizing the pilot transmissions at the user terminal* as well.

In this paper, we propose using superimposed pilots as an alternative to, and in combination with, time-multiplexed pilots in massive MIMO systems. Methods for channel estimation based on pilots that are embedded in data, such as superimposed pilots, have been extensively studied for MIMO systems [19]–[24]. However, these papers have focused on embedded and superimposed pilots in the context of accommodating a loss in signal-to-noise ratio (SNR) in exchange for a reduced pilot transmission overhead [20], [21]. Particularly, scenarios with high user-mobility, wherein it is impractical to allocate dedicated symbols for training, have been of interest for employing superimposed pilots. In the context of multi-cell massive MIMO, provided that the number of users in the system is smaller than the number of symbols in the UL, superimposed pilots allow for each user in the system to be assigned a unique pilot sequence, enabling the receiver to estimate the channel vectors of both the desired and interfering users. In addition, these pilots mitigate pilot contamination by time-averaging over long sequences and offer a higher efficiency due to a reduced transmission overhead.

We obtain expressions for the signal-to-interference-plus-noise ratio (SINR) at the output of a matched filter (MF)-based detector when a non-iterative least-squares (LS)-based channel estimate is employed for channel estimation. Based on the SINR expression, we highlight the need for power control when superimposed pilots are employed in a massive MIMO system. Moreover, we discuss the shortcomings of the non-iterative channel estimator and propose an intuitive low-complexity iterative channel estimation scheme for superimposed pilots.[1] In addition, we introduce the concept of a hybrid system and show by means of theoretical arguments that the hybrid system is superior to its counterpart that employs only time-multiplexed pilots, even when the non-iterative channel estimator is used to obtain the channel estimate from superimposed pilots. A simple approach to design and implement this hybrid system is also detailed. Although the use of superimposed pilots requires some coordination between the BSs in assigning pilot sequences to the users and estimating their path-loss coefficients, these are minor impediments compared to the performance improvements provided by the proposed scheme.

The article in the existing literature that is closest to this paper is [26], wherein superimposed pilots have been employed in the context of multi-cell multiuser MIMO systems.

However, unlike [26], the focus of our paper is to demonstrate the superiority of superimposed pilots when used in conjunction with time-multiplexed pilots in a hybrid system. The theoretical results and simulations that have been obtained are in line with this objective.

In Section II, the system model for the massive MIMO UL is described. In Section III, time-multiplexed pilots are described and the pilot contamination problem is detailed. Section IV introduces the superimposed pilot scheme and describes the non-iterative method for channel estimation and Section V discusses the iterative data-aided scheme. In Section VI, the concept of a hybrid system that employs both time-multiplexed and superimposed pilots is introduced and in Section VII, a simple approach for implementing this hybrid system is discussed. Section VIII presents simulation results demonstrating the effectiveness of employing superimposed pilots for pilot decontamination. Section IX concludes the paper. Some of the proofs are given in Appendix.

Notation : Lower case and upper case boldface letters denote column vectors and matrices, respectively. The notations $(\cdot)^*, (\cdot)^T, (\cdot)^H$, and $(\cdot)^{-1}$ represent the conjugate, transpose, Hermitian transpose, and inverse, respectively. The notation $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\mathbb{E}\{\cdot\}$ is used to denote the expectation operator. The notation $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, and $\|\cdot\|$ and $\|\cdot\|_F$ denote the Euclidean norm of a vector and Frobenius norm of a matrix, respectively. Upper case calligraphic letters denote sets, and $\varnothing$ denotes the empty set. The notation $\mathbf{1}_{\{\mathcal{S}\}}$ represents the indicator function over the set $\mathcal{S}$, whereas $\mathrm{Card}\,(\mathcal{S})$ is used to represent its cardinality. The notation $\delta_{n,m}$ denotes the Kronecker delta function, and $\eta(\cdot)$ stands for an element-by-element decision function that replaces each element of the input vector with the constellation point that is closest in Euclidean distance to that element. The big O notation $f(x) = O(g(x))$ implies that $|f(x)|/|g(x)|$ is bounded as $x \to \infty$.

II. SYSTEM MODEL

We consider a TDD massive MIMO UL with L cells and K single-antenna[2] users per cell. Each cell has a BS with $M \gg K$ antennas. The number of symbols C, over which the channel is coherent, is assumed to be divided into C_u and C_d, which are the number of symbols in the UL and DL time slots, respectively. The matrix of received measurements $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ at BS j can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \mathbf{s}_{\ell,k}^T + \mathbf{W}_j \tag{1}$$

where $\mu_{\ell,k}$ denotes the power with which the vector of symbols $\mathbf{s}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ are transmitted by user k in cell ℓ, $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise at BS j with each column distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. Moreover, the columns of $\mathbf{W}_j$ are mutually independent of each other. The vector $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ represents the channel response

---

[1]The work in this paper is a significant extension of our relevant conference paper [25]. In addition to a detailed exposition, we have included additional results that demonstrate the superiority of massive MIMO systems that use superimposed pilots instead of time-multiplexed pilots.

[2] For training and channel estimation, users with $T > 1$ antennas can be treated as T separate single-antenna users.

between the antennas at BS j, and user k in cell ℓ, and is assumed to be distributed as[3]

$$\mathbf{h}_{j,\ell,k} \sim \mathcal{CN}(\mathbf{0}, \beta_{j,\ell,k}\mathbf{I}_M) \qquad (2)$$

where $\beta_{j,\ell,k}$ denotes the large-scale path-loss coefficient which depends on the user location in the cell. In addition, the channel vectors $\mathbf{h}_{j,\ell,k}$ are assumed to be mutually independent of each other $\forall j, \ell, k$. The aforementioned statistics of the channel vector correspond to the non-line-of-sight scenario with rich scattering [1]. By virtue of their zero mean and mutual independence, the channel vectors are asymptotically orthogonal and the following equation holds almost surely [1]

$$\lim_{M \to \infty} \frac{\mathbf{h}_{j,\ell,k}^H \mathbf{h}_{m,n,p}}{M} = \beta_{j,\ell,k}\, \delta_{j,m}\delta_{\ell,n}\delta_{k,p},\ \forall\, j, k, \ell, m, n, p\,. \qquad (3)$$

Moreover, $\mathbf{h}_{j,\ell,k}$ is assumed to be constant for the duration of C symbols, and $\beta_{j,\ell,k}$ is constant for a significantly longer duration which depends on the user mobility. For the sake of simplicity, the effects of shadowing are not taken into account in this paper.[4] The transmitted symbols $\mathbf{s}_{\ell,k}$ contain both pilots and data. The pilots could either be time-multiplexed or superimposed pilots, and the elements of the data vector $\mathbf{x}_{\ell,k}$ are assumed to be independent and identically distributed (i.i.d) random variables with zero-mean and unit variance and take values from an alphabet χ, which is a realistic assumption.

III. TIME-MULTIPLEXED PILOTS AND THE PILOT CONTAMINATION PROBLEM

With time-multiplexed pilots, each user in a cell transmits a $\tau \geq K$ length orthogonal pilot sequence for channel estimation followed by $C_u - \tau$ symbols of uplink data. In order to minimize the overhead incurred, it is necessary to reuse these pilot sequences in the adjacent cells. However, this pilot-reuse results in the channel estimates of the desired users being contaminated by the channel vectors of users in adjacent cells, causing interference and in turn, a loss in spectral efficiency.

It is assumed here that the transmission of the pilot sequences by the users in the L cells are synchronized, which corresponds to the worst-case scenario for pilot contamination.[5] Consider a matrix $\mathbf{\Phi} \in \mathbb{C}^{\tau \times \tau}$ whose columns $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\tau\}$ are the orthogonal pilot sequences that are transmitted by the users, i.e., $\boldsymbol{\phi}_n^H \boldsymbol{\phi}_p = \tau \delta_{n,p}$. If $\boldsymbol{\phi}_{b_{\ell,k}}$ is the pilot sequence transmitted by user k of cell ℓ, where $b_{\ell,k} \in \{1, \ldots, \tau\}$ is the index of the transmitted pilot, and if each pilot sequence is reused once every $r \triangleq \tau/K$ cells [18],

the LS estimate of the channel of user m in cell j can be obtained as [1]

$$\widehat{\mathbf{h}}_{j,j,m}^{\text{TP}} \triangleq \frac{1}{\tau\sqrt{p_u}}\mathbf{Y}_j^{(p)}\boldsymbol{\phi}_{b_{j,m}}^* = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell=0 \\ \ell \neq j \\ \ell \in \mathcal{L}_j(r)}}^{L-1} \mathbf{h}_{j,\ell,m} + \frac{1}{\tau\sqrt{p_u}}\mathbf{W}_j^{(p)}\boldsymbol{\phi}_{b_{j,m}}^* \qquad (4)$$

where the superscript TP indicates that the estimates are computed when using time-multiplexed pilots, the superscript p indicates that the observations are made during pilot transmission, and $\mathcal{L}_j(r)$ is the subset of the L cells that use the same set of pilots as cell j. In addition, it is assumed in (4) without loss of generality that the transmit powers are same for all users employing time-multiplexed pilots, i.e., $\mu_{\ell,k} = p_u$, $\forall \ell, k$, and any variation in the transmit power of an individual user is absorbed into the corresponding path-loss coefficient β. It can be observed from (4) that the estimates of the channel vectors of the users in cell j are contaminated by the channel vectors of the users in the remaining $\text{Card}(\mathcal{L}_j(r) - 1)$ cells. When $M \to \infty$, the UL SINR of user m in cell j, at the output of an MF that uses the channel estimate in (4) for detection, can be written as [1]

$$\text{SINR}_{j,m}^{\text{TP-ul}} = \frac{\beta_{j,j,m}^2}{\sum\limits_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2}\,. \qquad (5)$$

The corresponding throughput of the user using Gaussian signaling in the UL can then be expressed as [1]

$$\text{R}_{j,m}^{\text{TP-ul}} = \frac{(C_u - \tau)}{C}\log_2\left(1 + \text{SINR}_{j,m}^{\text{TP-ul}}\right)\,. \qquad (6)$$

From the above equation, it can be observed that the rate per user is a function of both the overhead τ as well as the loss in SINR due to pilot contamination. A larger value of r would reduce the effect of pilot contamination and increase the SINR at the cost of a reduced transmission efficiency $(C_u - \tau)/C$.

IV. SUPERIMPOSED PILOTS

With superimposed pilots, the pilot symbols are transmitted at a reduced power alongside the data symbols, and in its simplest version, the pilot and data symbols are transmitted alongside each other for the entire duration of the uplink data slot C_u. If the total number of users in the system is smaller than the number of symbols in the uplink, i.e., $KL \leq C_u$,[6] then with superimposed pilots, each user can be assigned a unique orthogonal pilot $\mathbf{p}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$. The pilots are taken from the columns of a matrix $\mathbf{P} \in \mathbb{C}^{C_u \times C_u}$ such that $\mathbf{P}^H\mathbf{P} = C_u\mathbf{I}_{C_u}$, and therefore $\mathbf{p}_{\ell,k}^H\mathbf{p}_{n,p} = C_u\delta_{\ell,n}\delta_{k,p}$. If $\rho_{\ell,k}\mathbf{x}_{\ell,k} + \lambda_{\ell,k}\mathbf{p}_{\ell,k}$ is the transmitted vector from user k in

---

[3]While, for the sake of simplicity, an environment with rich scattering is assumed, the conclusions made in this paper are independent of the channel distribution and only require the channel vectors of any pair of users to be asymptotically orthogonal.

[4]The algorithms and analysis in this paper remain the same in the presence of shadowing, provided that the users are allocated to the strongest BSs. However, the geometric interpretations that are made based on the location of the user in the cell will no longer be valid.

[5]No additional improvement in the UL performance can be gleaned by separating the pilot and data transmissions across cells [1], [9], [18].

[6] For example, using the orthogonal frequency division multiplexing (OFDM) parameters in long-term evolution (LTE) systems as in [1], i.e., $C_u = 7$ OFDM symbols, $N_{\text{smooth}} = 14$ subcarriers, and assuming the pilots are reused over $L = 7$ hexagonal cells, the maximum number of supported users in the L cells is $C_u N_{\text{smooth}} = 98$ users. Therefore, the number of users per cell is $C_u N_{\text{smooth}}/L = 14$ users. However, note that the value of $C_u = 7$ has been chosen to allow user velocities of 350 km/h [27]. For lower user speeds and with cell sectoring, larger number users can be supported and the assumption $KL \leq C_u$ will easily be satisfied.

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} = \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\rho_{\ell,k}^2 \mu_{\ell,k} \beta_{j,\ell,k}^2}{C_u \lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2} + \frac{1}{M} \left(\sum_{\substack{\ell=0 \\ \{\ell \neq j, k \neq m\}}}^{L-1} \sum_{k=0}^{K-1} \frac{\beta_{j,\ell,k} \mu_{\ell,k}}{\rho_{j,m}^2 \beta_{j,j,m}} + \sum_{\substack{\ell=0 \\ \{\ell \neq j, k \neq m\}}}^{L-1} \sum_{k=0}^{K-1} \sum_{\substack{n=0 \\ \{n \neq \ell, p \neq k\}}}^{L-1} \sum_{p=0}^{K-1} \frac{\rho_{n,p}^2 \beta_{j,\ell,k} \beta_{j,n,p} \mu_{\ell,k}}{C_u \lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2} \right) \right)^{-1} \tag{12}$$

cell ℓ, then the received signal at the j'th BS $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$, when using the superimposed pilot scheme, can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left(\rho_{\ell,k} \mathbf{x}_{\ell,k} + \lambda_{\ell,k} \mathbf{p}_{\ell,k} \right)^T + \mathbf{W}_j \tag{7}$$

where $\lambda_{\ell,k}^2$ and $\rho_{\ell,k}^2$ are the fractions of the transmit power reserved for the pilot and data symbols, respectively, and the total transmitted power $\mu_{\ell,k}$ is given as $\mu_{\ell,k} = \lambda_{\ell,k}^2 + \rho_{\ell,k}^2$.

A. Non-Iterative Channel Estimation

Treating the data symbols of all users as additive noise, the channel estimate of user k in cell ℓ can be obtained at the j'th BS using the LS criterion [26]

$$\widehat{\mathbf{h}}_{j,\ell,k} \triangleq \arg \min_{\mathbf{h}} \| \mathbf{Y}_j - \lambda_{\ell,k} \mathbf{h} \, \mathbf{p}_{\ell,k}^T \|_F^2 \ . \tag{8}$$

Solving (8) yields

$$\widehat{\mathbf{h}}_{j,\ell,k} = \mathbf{Y}_j \left(\lambda_{\ell,k}^2 \, \mathbf{p}_{\ell,k}^H \mathbf{p}_{\ell,k} \right)^{-1} \lambda_{\ell,k} \mathbf{p}_{\ell,k}^* = \frac{1}{C_u \lambda_{\ell,k}} \mathbf{Y}_j \mathbf{p}_{\ell,k}^*$$
$$= \mathbf{h}_{j,\ell,k} + \frac{1}{C_u \lambda_{\ell,k}} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \rho_{m,n} \mathbf{h}_{j,m,n} \mathbf{x}_{m,n}^T \mathbf{p}_{\ell,k}^*$$
$$+ \frac{1}{C_u \lambda_{\ell,k}} \mathbf{W}_j \mathbf{p}_{\ell,k}^* \ . \tag{9}$$

In order to estimate the data from the received observations, it is necessary to remove the term corresponding to the transmitted superimposed pilot $\lambda_{j,m} \mathbf{h}_{j,j,m} \mathbf{p}_{j,m}^T$ from the observation vector in (7). Using $\lambda_{j,m} \widehat{\mathbf{h}}_{j,j,m} \mathbf{p}_{j,m}^T$ as an estimate for this term, the estimate of $\mathbf{x}_{j,m}$ can then be obtained from the observation $\mathbf{Y}_j$ using an MF and a decision operation as follows

$$\widetilde{\mathbf{x}}_{j,j,m}^T = \frac{1}{M \rho_{j,m} \beta_{j,j,m}} \widehat{\mathbf{h}}_{j,j,m}^H \left(\mathbf{Y}_j - \lambda_{j,m} \widehat{\mathbf{h}}_{j,j,m} \mathbf{p}_{j,m}^T \right) \tag{10}$$

$$\widehat{\mathbf{x}}_{j,j,m} = \eta \left(\widetilde{\mathbf{x}}_{j,j,m} \right) \ . \tag{11}$$

The SINR of user m in cell j, at the output of an MF that employs the channel estimate in (9), is derived in Appendix A and is given in (12) (shown at the top of the page). The SINR in (12), when $M \to \infty$, can be written as

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} = \frac{\lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2}{\frac{1}{C_u} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \mu_{\ell,k} \beta_{j,\ell,k}^2} \ . \tag{13}$$

The corresponding per-user rate in the uplink when using Gaussian signaling is given as

$$R_{j,m}^{\mathrm{SP-ul}} = \frac{C_u}{C} \log_2 \left(1 + \mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} \right) \ . \tag{14}$$

B. Power Control and Choice of Parameters $\lambda_{j,m}$ and $\rho_{j,m}$

From (13), it can be seen that the SINR of a user is dependent on the product of the transmit powers and large-scale fading coefficients of the remaining $LK - 1$ users in addition to the product of its own transmit power and large-scale fading coefficient. This dependence results in a situation similar to the near-far problem in code division multiple access (CDMA) systems, wherein users that have larger values of large-scale fading coefficient β swamp users that have smaller values of β. Therefore, it becomes necessary to use power control to provide a uniform user experience.

While the parameters $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$ can be optimized by maximizing the sum-rate of all the users, i.e.,

$$\max_{\mu_{\ell,k}, \rho_{\ell,k}, \lambda_{\ell,k}} \left\{ \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} R_{\ell,k}^{\mathrm{SP-ul}} \right\} \tag{15}$$

the optimization problem is in general non-convex and requires coordination between the BSs. As an alternative, a suboptimal solution that does not involve coordination between the BSs is obtained here for the parameters $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$. This suboptimal solution will be shown to maximize a lower bound on the sum-rate, and it is as follows.

The received signal in (7) can be equivalently written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \left(\frac{\rho_{\ell,k}}{\sqrt{\mu_{\ell,k}}} \mathbf{x}_{\ell,k}^T + \frac{\lambda_{\ell,k}}{\sqrt{\mu_{\ell,k}}} \mathbf{p}_{\ell,k}^T \right) + \mathbf{W}_j$$
$$= \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \bar{\mathbf{h}}_{j,\ell,k} \left(\bar{\rho}_{\ell,k} \mathbf{x}_{\ell,k} + \bar{\lambda}_{\ell,k} \mathbf{p}_{\ell,k} \right)^T + \mathbf{W}_j \tag{16}$$

where

$$\bar{\mathbf{h}}_{j,\ell,k} \triangleq \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \sim \mathcal{CN} \left(\mathbf{0}, \bar{\beta}_{j,\ell,k} \mathbf{I}_M \right) \tag{17}$$

$$\bar{\beta}_{j,\ell,k} \triangleq \beta_{j,\ell,k} \cdot \mu_{\ell,k} \tag{18}$$

$$\bar{\rho}_{\ell,k} \triangleq \sqrt{\frac{\rho_{\ell,k}^2}{\mu_{\ell,k}}} > 0 \tag{19}$$

$$\bar{\lambda}_{\ell,k} \triangleq \sqrt{\frac{\lambda_{\ell,k}^2}{\mu_{\ell,k}}} > 0 \tag{20}$$

$$\bar{\lambda}_{\ell,k}^2 + \bar{\rho}_{\ell,k}^2 = 1 \ . \tag{21}$$

From (16), it can be seen that a system having arbitrary values of $\beta_{j,\ell,k}$, $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$, can be reduced into an equivalent system with parameters $\bar{\beta}_{j,\ell,k}$, $\bar{\rho}_{\ell,k}$, and $\bar{\lambda}_{\ell,k}$, such that $0 \leq \bar{\rho}_{\ell,k}, \bar{\lambda}_{\ell,k} \leq 1$. Substituting (18) – (21) into (12), an equivalent expression for the SINR, as shown in (22) (shown at the top of the next page) can be obtained.

To obtain the parameter $\mu_{\ell,k}$, we propose using the statistics-aware power-control approach detailed in [18], wherein user m in cell j transmits at a power $\mu_{j,m} = \omega / \beta_{j,j,m}$

$$\text{SINR}_{j,m}^{\text{SP}-\text{ul}} = \left(\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1} \frac{\bar{\rho}_{\ell,k}^2 \bar{\beta}_{j,\ell,k}^2}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}^2} + \frac{1}{M}\left(\sum_{\substack{\ell=0 \\ \{\ell\neq j, k\neq m\}}}^{L-1}\sum_{k=0}^{K-1} \frac{\bar{\beta}_{j,\ell,k}}{\bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}} + \sum_{\substack{\ell=0 \\ \{\ell\neq j, k\neq m\}}}^{L-1}\sum_{k=0}^{K-1}\sum_{\substack{m=0 \\ \{m\neq \ell, n\neq k\}}}^{L-1}\sum_{n=0}^{K-1} \frac{\bar{\rho}_{m,n}^2 \bar{\beta}_{j,\ell,k} \bar{\beta}_{j,m,n}}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}^2} \right) \right)^{-1} \tag{22}$$

$$\text{SINR}_{j,m}^{\text{SP}-\text{ul}} \geq \left(\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1} \frac{\bar{\rho}_{\ell,k}^2}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2} + \frac{LK-1}{M\bar{\rho}_{j,m}^2} + \sum_{\substack{\ell=0 \\ \{\ell\neq j, k\neq m\}}}^{L-1}\sum_{k=0}^{K-1}\sum_{\substack{m=0 \\ \{m\neq \ell, n\neq k\}}}^{L-1}\sum_{n=0}^{K-1} \frac{\bar{\rho}_{m,n}^2}{MC_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2} \right)^{-1} \tag{29}$$

where ω is a design parameter. The parameter ω is chosen such that the transmitted power from a user satisfies a maximum power constraint, and users with severely low SINRs that would need a transmit power larger than this constraint would be denied service. This power control policy results in an identical received power of ω at the j'th BS for all the users in cell j. In addition, as mentioned in [18], the ratio $0 \leq \beta_{j,\ell,k}/\beta_{\ell,\ell,k} \leq 1$ is the relative strength of the interference received at BS j from a user in cell ℓ. This ratio is at most 1, when the user is at the edge of the j'th cell, and reduces to zero as its distance from BS j increases. Therefore, setting $\mu_{\ell,k} = \omega/\beta_{\ell,\ell,k}$ and using the definitions of $\bar{\beta}_{j,\ell,k}$, $\bar{\rho}_{\ell,k}$, and $\bar{\lambda}_{\ell,k}$, and the inequality $0 \leq \beta_{j,\ell,k}/\beta_{\ell,\ell,k} \leq 1$, the following equations can be obtained

$$\bar{\beta}_{j,j,m} = \beta_{j,j,m} \cdot \mu_{j,m} = \omega \tag{23}$$
$$\bar{\beta}_{j,\ell,k} = \beta_{j,\ell,k} \cdot \mu_{\ell,k} \leq \beta_{\ell,\ell,k}\mu_{\ell,k} = \omega \quad \forall \ell \neq j \tag{24}$$
$$\rho_{j,m}^2 \beta_{j,j,m} = \bar{\rho}_{j,m}^2 \omega \tag{25}$$
$$\lambda_{j,m}^2 \beta_{j,j,m} = \bar{\lambda}_{j,m}^2 \omega \tag{26}$$
$$\rho_{\ell,k}^2 \beta_{j,\ell,k} \leq \rho_{\ell,k}^2 \beta_{\ell,\ell,k} = \bar{\rho}_{\ell,k}^2 \omega \quad \forall \ell \neq j \tag{27}$$
$$\lambda_{\ell,k}^2 \beta_{j,\ell,k} \leq \lambda_{\ell,k}^2 \beta_{\ell,\ell,k} = \bar{\lambda}_{\ell,k}^2 \omega \quad \forall \ell \neq j . \tag{28}$$

Substituting the above equations into (22), a lower bound on the SINR, as shown in (29) (shown at the top of the page), can be obtained.

However, the maximization of the lower bound on the SINR and hence, a lower bound on the sum rate, is still a non-convex problem in the parameters $\bar{\rho}_{\ell,k}$ and $\bar{\lambda}_{\ell,k}$ and requires coordination between the BSs. To circumvent this problem, we restrict the parameters $\bar{\rho}_{\ell,k}$ and $\bar{\lambda}_{\ell,k}$ such that $\bar{\rho}_{\ell,k} = \bar{\rho}, \forall \ell,k$ and $\bar{\lambda}_{\ell,k} = \bar{\lambda}, \forall \ell,k$. The choice of this restriction is motivated by the observation from (23) that the statistics-aware power control scheme results in the same large-scale path loss coefficient for all the desired users in the cell, irrespective of their locations. As a result, from the BS's perspective, each of its users are identical, and therefore, there is no benefit in assigning different values of $\bar{\rho}_{\ell,k}$ to different users. More importantly, such a restriction renders the choice of $\bar{\rho}_{\text{opt}}$ to depend only on L, K, C_u, and M as will be shown next. Setting $\bar{\rho}_{\ell,k} = \bar{\rho}, \forall \ell,k$ and $\bar{\lambda}_{\ell,k} = \bar{\lambda}, \forall \ell,k$ in (29), we obtain

$$\text{SINR}_{j,m}^{\text{SP}-\text{ul}} \geq \left(\frac{LK}{C_u(1-\bar{\rho}^2)} + \frac{1}{M}\left(\frac{LK-1}{\bar{\rho}^2} + \frac{(LK-1)^2}{C_u(1-\bar{\rho}^2)} \right) \right)^{-1} . \tag{30}$$

Differentiating the right hand side of (30) with respect to $\bar{\rho}^2$ and setting the resulting expression to zero, the value of $\bar{\rho}^2$ that maximizes the lower bound on $\text{SINR}_{j,m}^{\text{SP}-\text{ul}}$ and the UL sum rate can be obtained as

$$\bar{\rho}_{\text{opt}}^2 = \left(1 + \sqrt{\frac{\frac{LK}{C_u} + \frac{(LK-1)^2}{MC_u}}{\frac{LK-1}{M}}} \right)^{-1} \approx \left(1 + \sqrt{\frac{M+LK}{C_u}} \right)^{-1} \tag{31}$$

and the optimal value of $\bar{\lambda}^2$ can be obtained as

$$\bar{\lambda}_{\text{opt}}^2 = 1 - \bar{\rho}_{\text{opt}}^2 \approx \left(1 + \sqrt{\frac{C_u}{M+LK}} \right)^{-1} \tag{32}$$

where the approximations in (31) and (32) have been made assuming $LK \gg 1$ in order to obtain simpler expressions.

Based on the fact established in this subsection that systems using $\rho, \lambda, \beta, \mathbf{h}$, and $\bar{\rho}, \bar{\lambda}, \bar{\beta}, \bar{\mathbf{h}}$ are equivalent, we drop the overbar for ease of notation and adopt the former set of symbols in the rest of the paper. In addition, we set $\mu_{\ell,k} = \bar{\rho}_{\ell,k}^2 + \bar{\lambda}_{\ell,k}^2 = 1, \forall \ell,k$.

C. Impact of C_u on the Performance of Superimposed Pilots

Using (13) and a fixed set of parameters r, τ, and K, the following theorem presents an important condition that guarantees the superiority of methods based on superimposed pilots over the LS estimator that is based on time-multiplexed pilots.

Theorem 1. *With fixed values of K, r, and τ and if $M \to \infty$, there exists a UL duration $\kappa_{j,m}$ beyond which a channel estimator based on superimposed pilots outperforms the LS based channel estimator that utilizes time-multiplexed pilots, in terms of the SINR performance, in any channel scenario $\{\beta_{j,\ell,m} \mid 0 \leq j, \ell \leq L-1, \ 0 \leq m \leq K-1\}$.*

Proof. If $\kappa_{j,m}$ is defined as the number of symbols in the uplink such that (5) and (13) are equal, i.e.,

$$\frac{\beta_{j,j,m}^2}{\frac{1}{\kappa_{j,m}\lambda_{j,m}^2\rho_{j,m}^2}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\rho_{\ell,k}^2\beta_{j,\ell,k}^2} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r)}}\beta_{j,\ell,m}^2} \tag{33}$$

then it is evident from (13) and (33) that $C_u > \kappa_{j,m}$ is a sufficient condition for a method that is based on superim-

posed pilots to outperform the LS method that employs time-multiplexed pilots. In addition, $\kappa_{j,m}$ is given as

$$\kappa_{j,m} \triangleq \frac{\frac{1}{\lambda_{j,m}^2 \rho_{j,m}^2} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \beta_{j,\ell,k}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2} \ . \tag{34}$$

This completes the proof. $\qquad\square$

Remark 1: An important consequence of the above theorem is that in scenarios with negligible pilot contamination, the LS method based on superimposed pilots requires a large value of C_u to outperform the LS method based on time-multiplexed pilots. As an example, consider the case when $r = 1$, $\beta_{j,j,m} = 1$, $\forall m$, $\beta_{j,\ell,m} = \beta$, $\forall \ell \neq j, m$, and $\rho_{j,m}^2 = \lambda_{j,m}^2, \forall\, j, m$. For such a scenario, $\kappa_{j,m}$ is given as

$$\kappa_{j,m} = 2K \left(1 + \frac{1}{(L-1)\beta^2} \right) \ . \tag{35}$$

Then, if the LS estimator based on superimposed pilots is required to maintain superiority over the LS estimator employing time-multiplexed pilots, C_u must scale inversely with β^2. This dependence on C_u is evident from the expression for the channel estimation error, which is given as

$$\Delta\mathbf{h}_{j,j,m} \triangleq \mathbf{h}_{j,j,m} - \widehat{\mathbf{h}}_{j,j,m} = -\frac{1}{C_u \lambda_{j,m}}$$
$$\times \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k} \mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T + \mathbf{W}_j \right) \mathbf{p}_{j,m}^* \ . \tag{36}$$

Remark 2: We build upon the discussion in [1] on grouping users based on their coherence times. While such a grouping does not offer any performance benefits to users when employing the approach in [1], the use of superimposed pilots offers low-mobility users an increase in throughput, by minimizing the channel estimation error resulting from transmitting the data alongside the pilots. This improvement in performance is a direct consequence of Theorem 1.

Remark 3: The type of pilot transmitted by a user can also be chosen based on the coherence time. While users with high-mobility or low pilot contamination would find it sufficient to use time-multiplexed pilots, users with low-mobility who suffer from significant pilot contamination due to their proximity to the cell-edge or due to shadowing would significantly benefit from employing superimposed pilots.

Remark 4: Superimposed pilots require coordination between BSs when assigning pilot sequences and synchronizing transmissions. In practical cellular networks the cells are fairly large and it can be assumed that the interference is restricted to the first tier of cells and interference from the second and higher tiers of cells can be neglected. Therefore, it is reasonable to assume that practical deployments of superimposed pilots will require pilot assignment only over the first tier of cells, implying that coordination is limited to only this first tier. This overhead is not very different from that required by time-multiplexed pilots in the presence of pilot reuse. The coordination and synchronization requirements of superimposed pilot-based systems that allocate pilots over the first tier of cells

are similar to that of time-multiplexed pilot-based systems that have a pilot reuse factor of $r = 3$ [18].

From (36), it can be seen that the error in the channel estimate includes interference resulting from transmitting data alongside the pilots. Hence, the quality of the channel estimate can be improved by eliminating the interference from the transmitted data through iterative data-aided schemes, thereby increasing the robustness of the proposed method with respect to C_u.

V. ITERATIVE DATA-AIDED CHANNEL ESTIMATION

In the iterative approach to channel estimation developed in this section, the estimated channel and data vectors of both the desired and interfering users are used in feedback in order to eliminate the first term in (36). In addition, to minimize error propagation between the channel estimates of different users, the iteration is started from the user with the highest SINR and is progressed in the decreasing order of the SINRs of the users. It has to be noted that the objective of this section is to demonstrate that iterative methods for channel estimation with superimposed pilots provide a significantly better SINR performance than their non-iterative counterparts, and hence we restrict ourselves to a simple iterative algorithm. However, there is scope for developing improved iterative algorithms in the future.

A. Algorithm

For the sake of clarity and without loss of generality, we replace the two indices k, ℓ with a single index m that lies in the range $0 \leq m \leq N - 1$, where $N \triangleq KL$. The index m is used to index the users in all the L cells. In addition, we drop the index j and implicitly assume that the channel estimation is performed at the j'th BS. Then, (7) can be rewritten as

$$\mathbf{Y} = \sum_{m=0}^{N-1} \mathbf{h}_m \left(\rho_m \mathbf{x}_m + \lambda_m \mathbf{p}_m \right)^T + \mathbf{W} \ . \tag{37}$$

Since for large M, the SINRs of the users are proportional to the users' path-loss coefficients, the users are arranged in the decreasing order of their path-loss coefficients, i.e., $\beta_0 > \beta_1 > \ldots > \beta_{N-1}$.[7] Then, using an estimate of $\rho_m \mathbf{h}_m \mathbf{x}_m^T \mathbf{p}_m^*$ for each user as a correction factor to minimize the interference from other users, the corresponding channel estimate of user m can be written as

$$\widehat{\mathbf{h}}_m^{(i)} = \frac{1}{C_u \lambda_m} \left[\mathbf{Y} - \sum_{\substack{k=0 \\ k \in \mathcal{U}_m^{(i)}}}^{m-1} \rho_k \widehat{\mathbf{h}}_k^{(i)} \left(\widehat{\mathbf{x}}_k^{(i)} \right)^T \right.$$
$$\left. - \sum_{\substack{k=m \\ k \in \mathcal{U}_m^{(i)}}}^{N-1} \rho_k \widehat{\mathbf{h}}_k^{(i-1)} \left(\widehat{\mathbf{x}}_k^{(i-1)} \right)^T \right] \mathbf{p}_m^* \tag{38}$$

---

[7]It is assumed that the BSs have access to the exact values of the path-loss coefficients β_m and that there is no false-ordering. This assumption is reasonable since for large M, the path-loss coefficients can be computed at the BS with negligible error by averaging the power of the channel coefficients over the entire array.

where $\widehat{\mathbf{h}}_m^{(0)} = \mathbf{0}$, $\forall\, m$ and $\mathcal{U}_m^{(i)}$ is the set of users whose estimated data is used in feedback in the i'th iteration to estimate the channel vector of user m. The approach to obtain $\mathcal{U}_m^{(i)}$ has been detailed in Appendix C, and involves selecting users such that the interference power, described in the next subsection, does not increase with each iteration. The channel estimate in the above equation is a modified version of the LS estimator defined in (9) with an added correction factor. Utilizing the resulting channel estimate in an MF and decision operation, similar to (10) and (11), the estimate of the data is obtained as follows

$$\left(\widetilde{\mathbf{x}}_m^{(i)}\right)^T = \frac{1}{M\rho_m\beta_m} \left(\widehat{\mathbf{h}}_m^{(i)}\right)^H \left(\mathbf{Y} - \lambda_m \widehat{\mathbf{h}}_m^{(i)} \mathbf{p}_m^T\right) \quad (39)$$

$$\widehat{\mathbf{x}}_m^{(i)} = \eta\left(\widetilde{\mathbf{x}}_m^{(i)}\right) \quad (40)$$

where $\widehat{\mathbf{x}}_m^{(0)} = \mathbf{0}$, $\forall\, m = 0, \ldots, N-1$.

Remark 5: If the matrix $\mathbf{P}$, whose columns are the superimposed pilots, is chosen as $\mathbf{P} = \text{blkdiag}\{\mathbf{P}_0, \ldots, \mathbf{P}_{L-1}\}$, where the ℓ'th block $\mathbf{P}_\ell \in \mathbb{C}^{K \times K}$ is comprised of the orthogonal pilot sequences used by the K users in cell ℓ, then the latency introduced when the non-iterative method is employed is the same as that for time-multiplexed pilots. However, when the iterative method is employed, the channel and the data vectors of the users are required and therefore, the uplink data in the entire slot will have to be aggregated before estimating the channel, which introduces a latency of C_u symbols.

Remark 6: From (9), the non-iterative method for channel estimation requires MC_u operations per user, whereas the MF and decision operations in (10) and (11) require M and C_u operations per user, respectively.

For the iterative method with ν iterations, the channel estimator, matched filter, and decision operations have a combined complexity of $O(\nu MC_u) + O(\nu M) + O(\nu C_u)$.

B. Interference Power at the BS

Let $\mathbf{e}_m^{(i)} \triangleq \mathbf{x}_m - \widetilde{\mathbf{x}}_m^{(i)}$ be the error in the estimate of the data symbols of user m obtained from the MF in the i'th iteration. Let $\Delta\mathbf{x}_m^{(i)} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_m^{(i)}$ be the corresponding error vector after the decision operation and let $\Delta\mathbf{h}_m^{(i)} \triangleq \mathbf{h}_m - \widehat{\mathbf{h}}_m^{(i)}$ be the associated error in the channel estimate. If $\alpha_n^{(i)}$ is the variance of the elements of $\Delta\mathbf{x}_n^{(i)}$ and assuming that the elements of $\mathbf{e}_m^{(i)}$ are i.i.d. circular complex-Gaussian random variables with zero mean and variance $I_m^{(i)}$, an approximate expression for the interference power $I_m^{(i)}$ can be written as

$$I_m^{(i)} \approx \frac{1}{\beta_m^2} \left(\frac{1}{M\rho_m^2} \sum_{\substack{k=0 \\ k \neq m}}^{N-1} \beta_k \beta_m + \frac{\sigma^2 \beta_m}{M\rho_m^2} + \frac{1}{M^2\rho_m^2} \psi_m^{(i)} \right) \quad (41)$$

where the expression for $\psi_m^{(i)}$ is given in (42) on the top of the next page and $\psi_m^{(0)} = 0$, $\forall m$. The detailed derivation of $I_m^{(i)}$ can be found in Appendix B.

In deriving (41), the following simplifying assumptions have been made in order to obtain a closed form expression:

(S1) $\mathbf{e}_m^{(i)}$ is independent of $\mathbf{x}_k$ and $\mathbf{W}$, $\forall\, k, i$.
(S2) $\Delta\mathbf{x}_m^{(i)}$ is independent of $\mathbf{x}_k$, $\mathbf{W}$, and $\mathbf{h}_k$, $\forall\, k, i$.

(S3) $\Delta\mathbf{x}_m^{(i)}$ is independent of $\Delta\mathbf{x}_k^{(p)}$, $\forall p \neq i, m \neq k$ and the elements of $\Delta\mathbf{x}_m^{(i)}$ are i.i.d.
(S4) $\Delta\mathbf{h}_m^{(i)}$ is independent of $\mathbf{x}_k$, $\mathbf{W}$, and $\Delta\mathbf{x}_k^{(p)}$, $\forall\, k, p$.

In scenarios with low interference and with large M, only a few of the received symbols will be erroneous. As a result, the elements of $\Delta\mathbf{x}_m^{(i)}$ are sparse with the few non-zero elements restricted to locations that correspond to the erroneous symbols. Moreover, the vector $\mathbf{e}_m^{(i)}$ represents the error in the estimated data and in such low-interference scenarios, the elements of $\mathbf{e}_m^{(i)}$ take small values. Therefore, the simplifications (S1), (S2), and (S3) are reasonably accurate for these scenarios. Although the expression for $\Delta\mathbf{h}_m^{(i)}$, (given in (82) in Appendix B) is explicitly dependent on $\mathbf{x}_k$ and $\Delta\mathbf{x}_m^{(i)}$, we neglect the correlation between these terms since $\Delta\mathbf{h}_m^{(i)}$ is inversely proportional to C_u, and the simplification (S4) is fairly accurate when C_u is large with respect to N and when scenarios with low interference are considered. Since $\mathbf{e}_m^{(i)}$ is assumed to be a zero-mean random variable, $\Delta\mathbf{x}_k^{(i)}$ is also a zero-mean random variable, provided the constellation points in χ and their probability density functions are symmetric about the origin. This is true since by definition, $\Delta\mathbf{x}_k^{(i)}$ and $\mathbf{e}_m^{(i)}$ are related to each other through the following equation

$$\Delta\mathbf{x}_k^{(i)} = \mathbf{x}_k - \eta\left(\mathbf{x}_k - \mathbf{e}_m^{(i)}\right) . \quad (43)$$

From (43), an expression for the variance of the elements of $\Delta\mathbf{x}_k^{(i)}$, i.e., $\alpha_k^{(i)}$ can be found as

$$\alpha_k^{(i)} \triangleq \mathbb{E}\left\{ \left| \left[\Delta\mathbf{x}_k^{(i)} \right]_n \right|^2 \right\} = \int |\Delta x|^2\, p_{\Delta\mathbf{x}_k^{(i)}}(\Delta x) d\Delta x$$
$$= \int_{x \in \chi} \int |x - \eta(x-e)|^2\, p_{\mathbf{e}_k^{(i)}, \mathbf{x}_k}(e, x)\, de\, dx$$
$$= \int_{x \in \chi} \int |x - \eta(x-e)|^2\, p_{\mathbf{e}_k^{(i)}}(e)\, p_{\mathbf{x}_k}(x)\, de\, dx \quad (44)$$

where $p_{\mathbf{e}_k^{(i)}}(\cdot)$, $p_{\Delta\mathbf{x}_k^{(i)}}(\cdot)$, and $p_{\mathbf{x}_k}(\cdot)$ are the probability density functions of the elements of $\mathbf{e}_k^{(i)}$, $\Delta\mathbf{x}_k^{(i)}$, and $\mathbf{x}_k$, respectively, and $p_{\mathbf{e}_k^{(i)}, \mathbf{x}_k}(\cdot)$ is the joint density function of the random variables $\mathbf{e}_k^{(i)}$ and $\mathbf{x}_k$. The latter has been written as the product of their individual distributions in the final expression of (44), thanks to (S1).

Important example of $\alpha_m^{(i)}$: When the elements of $\mathbf{x}_m$ are uniformly distributed and take values from a unit-power P-quarternary amplitude modulation (QAM) constellation, then under the assumption that the symbol errors in $\Delta\mathbf{x}_k^{(i)}$ are dominated by the closest neighboring symbols, the expression for $\alpha_m^{(i)}$ can be written as

$$\alpha_m^{(i)} = \begin{cases} \frac{24}{\sqrt{P}(\sqrt{P}+1)} Q\left(\sqrt{\frac{\frac{3}{(P-1)}}{I_m^{(i)}}}\right), & i \geq 1 \\ 1, & i = 0 \end{cases} \quad (45)$$

where $Q(\cdot)$ is the Q-function. The detailed derivation of the above expression can be found in Appendix D.

$$\left. \psi_m^{(i)} \right|_{i \geq 1} = \frac{M^2}{C_u \lambda_m^2} \left[\sum_{k \in \mathcal{U}_m^{(i)}, k < m} \rho_k^2 \left\{ \beta_k^2 \alpha_k^{(i)} + \frac{1}{M} \sum_{n=0}^{N-1} \beta_n \beta_k \alpha_k^{(i)} + \frac{\left(1 + \alpha_k^{(i)}\right)}{M^2} \psi_k^{(i)} \right\} + \sum_{k \notin \mathcal{U}_m^{(i)}} \rho_k^2 \left\{ \beta_k^2 + \frac{1}{M} \sum_{n=0}^{N-1} \beta_n \beta_k \right\} \right.$$
$$\left. + \sum_{k \in \mathcal{U}_m^{(i)}, m \leq k \leq N} \rho_k^2 \left\{ \beta_k^2 \alpha_k^{(i-1)} + \sum_{n=0}^{N-1} \frac{1}{M} \beta_n \beta_k \alpha_k^{(i-1)} + \frac{\left(1 + \alpha_k^{(i-1)}\right)}{M^2} \psi_k^{(i-1)} \right\} + \frac{\sigma^2}{M} \left(\sum_{n=0}^{N-1} \beta_n \right) \right] . \quad (42)$$
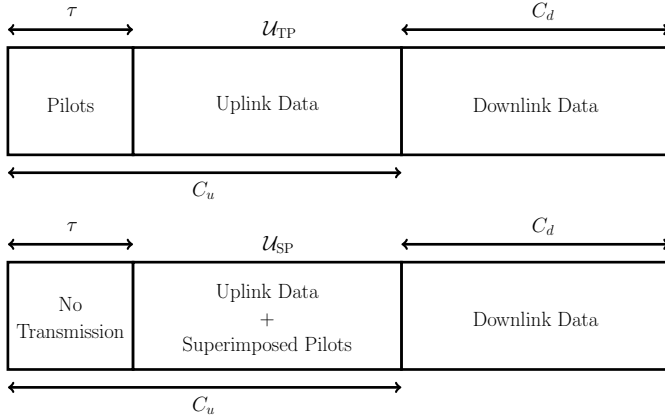

Fig. 1. Frame structure of a hybrid system with users employing time-multiplexed and superimposed pilots.

VI. HYBRID SYSTEM

One of the main advantages of superimposed pilots over time-multiplexed pilots is that it does not require a separate set of symbols for pilot transmission. This property can be used to construct a hybrid system that contains two disjoint sets of users, with the users in one of the sets employing time-multiplexed pilots, and the users in the other set employing superimposed pilots. The following theorem shows that this hybrid system has a higher throughput and supports a larger number of users than a system that employs only time-multiplexed pilots.

Theorem 2. *In a system that employs time-multiplexed pilots and is designed to maximize the UL and DL sum-rate,[8] let K be the optimal number of users per cell, L be the total number of cells in the system, $\tau > 0$ be the optimal number of symbols used for pilot training, r be the optimal pilot-reuse factor, and $C_u - \tau$ and C_d be the number of data symbols in the UL and DL slots, respectively. Then, with $M \to \infty$, there exists a hybrid system, that uses both time-multiplexed and superimposed pilots, which is capable of supporting $C_u - \tau$ additional users and offers a higher sum-rate in the UL than the optimal system that only employs time-multiplexed pilots.*

Proof. Consider the frame structure in Fig. 1, wherein there are two sets of users $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$. The users in the set $\mathcal{U}_{\mathrm{TP}}$ employ time-multiplexed pilots, with parameters selected using approaches such as in [18]. The users in the set $\mathcal{U}_{\mathrm{SP}}$ maintain radio silence during the pilot training phase of the users in $\mathcal{U}_{\mathrm{TP}}$, i.e., for τ symbols in the frame, and transmit orthogonal pilots superimposed with data during the uplink

data phase of $C_u - \tau$ symbols. Since these users maintain radio silence during the pilot training phase of τ symbols, they do not affect the quality of the channel estimates of the users in $\mathcal{U}_{\mathrm{TP}}$. As a result, under the assumption of asymptotic orthogonality of the channels, there is no interference from the users in $\mathcal{U}_{\mathrm{SP}}$ to those in $\mathcal{U}_{\mathrm{TP}}$. Therefore, the per-cell sum-rate in the UL for the users in $\mathcal{U}_{\mathrm{TP}}$ remains unchanged and can be found from (6) to be

$$\mathrm{R}_j^{\mathrm{ul}}(\mathcal{U}_{\mathrm{TP}}) = \frac{(C_u - \tau)}{C} \sum_{\substack{k=0 \\ k \in \mathcal{U}_{\mathrm{TP}}}}^{K-1} \log_2 \left(1 + \frac{\beta_{j,j,k}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,k}^2} \right) . \quad (46)$$

Assuming, for the sake of simplicity, that all the users in $\mathcal{U}_{\mathrm{SP}}$ are located in the j'th cell, the sum-rate of the users in $\mathcal{U}_{\mathrm{SP}}$ can be found using (13) and (14) as

$$\mathrm{R}_j^{\mathrm{ul}}(\mathcal{U}_{\mathrm{SP}}) = \frac{(C_u - \tau)}{C} \sum_{m \in \mathcal{U}_{\mathrm{SP}}} \log_2 \left(1 + \mathrm{SINR}_m(\mathcal{U}_{\mathrm{SP}}) \right) \quad (47)$$

$$\mathrm{SINR}_m(\mathcal{U}_{\mathrm{SP}}) \triangleq \frac{\beta_{j,j,m}^2}{\sum_{k \in \mathcal{U}_{\mathrm{SP}}} \frac{\rho_{j,k}^2 \beta_{j,j,k}^2}{(C_u - \tau)\rho_{j,m}^2 \lambda_{j,m}^2}} . \quad (48)$$

In obtaining the above expression, it has been assumed that the transmit power p_u of the users in $\mathcal{U}_{\mathrm{TP}}$ is small enough such that the interference to the users in $\mathcal{U}_{\mathrm{SP}}$ can be neglected.[9] Therefore, from (46) and (48), the combined rate $\mathrm{R}_j^{\mathrm{ul}}(\mathcal{U}_{\mathrm{SP}}) + \mathrm{R}_j^{\mathrm{ul}}(\mathcal{U}_{\mathrm{TP}})$ is strictly greater than $\mathrm{R}_j^{\mathrm{ul}}(\mathcal{U}_{\mathrm{TP}})$. In addition, since the data slot is made up of $C_u - \tau$ symbols, it is possible to allocate $C_u - \tau$ orthogonal pilots and therefore, the set $\mathcal{U}_{\mathrm{SP}}$ can contain a maximum of $C_u - \tau$ users. This concludes the proof. □

In the above theorem, given a system with users employing time-multiplexed pilots, we have shown that additional users employing superimposed pilots can always be added to the system, resulting in a hybrid system that offers a higher throughput.

In the following section, we utilize the concept of the above theorem to partition a given set of users employing time-multiplexed pilots into two disjoint subsets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ that contain users transmitting time-multiplexed pilots and superimposed pilots, respectively. There are two main benefits of performing such a partition: (i) there is an overall

---

[8]Such as the scheme described in [18].

[9]This assumption is valid since the SINR and the rate of the users in $\mathcal{U}_{\mathrm{TP}}$ are independent of the transmit power p_u when $M \to \infty$. It has to be noted that this assumption has been made for the sake of simplicity and the theorem is valid even if this assumption does not hold.

improvement in the throughput as a result of the reduced inter-cell interference; and (ii) there is a reduction in the number of users that use time-multiplexed pilots, thereby allowing for more aggressive pilot reuse since r is a function of the number of users employing time-multiplexed pilots [18].

VII. A Simple Implementation of the Hybrid System

Given a set of K users per cell in L cells with channel gains $\beta_{j,\ell,k}, \forall j, \ell = 1, \ldots, L$, and $k = 1, \ldots, K$, the problem of partitioning users into disjoint sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ can be accomplished by minimizing the overall UL inter-cell and intra-cell interference. This choice of objective function is motivated by Theorem 1, wherein it is observed that users at the cell edge cause significant pilot contamination and benefit from being assigned superimposed pilots, whereas users that are close to the BS cause negligible interference and could be assigned time-multiplexed pilots that are potentially shared with users in neighboring cells.

A. Framework

If the users in $\mathcal{U}_{\mathrm{TP}}$ transmit pilots with unit power and data at a power p_u, then the received signal from the hybrid system in the UL phase at BS j can be written as

$$\mathbf{Y}_j = \mathbf{Y}_j^{\mathrm{TP}} + \mathbf{Y}_j^{\mathrm{SP}} + \mathbf{W}_j \tag{49}$$

where $\mathbf{Y}_j^{\mathrm{TP}}$ and $\mathbf{Y}_j^{\mathrm{SP}}$ are the received signals from the users in $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$, respectively. From Fig. 1, $\mathbf{Y}_j^{\mathrm{TP}}$ and $\mathbf{Y}_j^{\mathrm{SP}}$ can be written as

$$\mathbf{Y}_j^{\mathrm{TP}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{TP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left[\boldsymbol{\phi}_{\ell,k}^T, \sqrt{p_u}\mathbf{x}_{\ell,k}^T \right] \tag{50}$$

$$\mathbf{Y}_j^{\mathrm{SP}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left[\mathbf{0}_{1\times\tau}, \rho\mathbf{x}_{\ell,k}^T + \lambda\mathbf{p}_{\ell,k}^T \right] \tag{51}$$

where the tuple (ℓ,k) is used to denote user k in cell ℓ.

If user (j,m) is a member of $\mathcal{U}_{\mathrm{TP}}$, then the LS estimate of its channel can be written as [1]

$$\widehat{\mathbf{h}}_{j,j,m} = \frac{1}{\tau}\mathbf{Y}_j\mathbf{b}_{j,m}^{\mathrm{TP}}\mathbf{h}_{j,j,m} + \sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\mathrm{TP}}}} \mathbf{h}_{j,\ell,m} + \frac{1}{\tau}\mathbf{W}_j\mathbf{b}_{j,m}^{\mathrm{TP}} \tag{52}$$

where $\mathbf{b}_{j,m}^{\mathrm{TP}} \triangleq \left[\boldsymbol{\phi}_{j,m}^H, \mathbf{0}_{(1\times(C_u-\tau))} \right]^T$. If $M \gg K$, the SINR in the UL when using the channel estimate in (52) can be obtained similar to (5) as

$$\mathrm{SINR}_{j,m}^{\mathrm{TP-ul}} \approx \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\mathrm{TP}}}} \beta_{j,\ell,m}^2} \tag{53}$$

where the approximations in (53) is made for the sake of simplicity and is valid when M is sufficiently large.

If user (j,m) is a member of $\mathcal{U}_{\mathrm{SP}}$, then the LS estimate of its channel can be written as

$$\widehat{\mathbf{h}}_{j,j,m} = \frac{1}{(C_u-\tau)\lambda}\mathbf{Y}_j\mathbf{b}_{j,m}^{\mathrm{SP}}$$

$$= \mathbf{h}_{j,j,m} + \frac{\rho}{(C_u-\tau)\lambda} \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k}\mathbf{x}_{\ell,k}^T\mathbf{p}_{j,m}^*$$

$$+ \sqrt{p_u} \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{TP}}}}^{L-1} \sum_{k=0}^{K-1} \frac{\mathbf{h}_{j,\ell,k}\mathbf{x}_{\ell,k}^T\mathbf{p}_{j,m}^*}{(C_u-\tau)\lambda} + \mathbf{W}_j\mathbf{b}_{j,m}^{\mathrm{SP}} \tag{54}$$

where $\mathbf{b}_{j,m}^{\mathrm{SP}} \triangleq \left[\mathbf{0}_{(1\times\tau)}, \mathbf{p}_{j,m}^H \right]^T$. Since it can be seen from (53) that the UL SINR of the users in $\mathcal{U}_{\mathrm{TP}}$ is independent of the UL transmit power p_u, we assume that p_u is small enough with respect to the transmit powers of the users in $\mathcal{U}_{\mathrm{SP}}$. As a result, the users in $\mathcal{U}_{\mathrm{SP}}$ do not experience significant interference during the data transmission phase of the users in $\mathcal{U}_{\mathrm{TP}}$ and result in the transmissions of $\mathcal{U}_{\mathrm{SP}}$ and $\mathcal{U}_{\mathrm{TP}}$ becoming independent of each other.[10] Then (54) simplifies as

$$\widehat{\mathbf{h}}_{j,j,m} \approx \mathbf{h}_{j,j,m} + \frac{\rho}{\lambda} \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \frac{\mathbf{h}_{j,\ell,k}\mathbf{x}_{\ell,k}^T\mathbf{p}_{j,m}^*}{(C_u-\tau)} + \mathbf{W}_j\mathbf{b}_{j,m}^{\mathrm{SP}} . \tag{55}$$

Then the SINR in the UL for the users in $\mathcal{U}_{\mathrm{SP}}$ can be obtained from (13) as

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} \approx \frac{\beta_{j,j,m}^2}{\frac{1}{(C_u-\tau)\lambda^2} \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \beta_{j,\ell,k}^2} \tag{56}$$

where, similar to (53), the approximation in (56) is made for the sake of simplicity and is valid when M is sufficiently large.

B. Algorithm to Obtain $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$.

The goal in this subsection is to obtain an algorithm for partitioning users into the sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ by minimizing the total UL inter-cell and intra-cell interference. In order to accomplish this, we quantify the amount of interference caused by a user that is assigned to either of the sets $\mathcal{U}_{\mathrm{TP}}$ or $\mathcal{U}_{\mathrm{SP}}$.

Let $I_{j,m}^{\mathrm{TP-ul}}$ or $I_{j,m}^{\mathrm{SP-ul}}$ be the contributions of user (j,m) to the total UL inter/intra-cell interference power when assigned to $\mathcal{U}_{\mathrm{TP}}$ or $\mathcal{U}_{\mathrm{SP}}$, respectively. If users (j,m) and (ℓ,k) are members of $\mathcal{U}_{\mathrm{TP}}$, then from the denominator of (53), the amount of interference that user (j,m) causes to user (ℓ,k) in the UL is $\beta_{\ell,j,k}^2\delta_{m,k}$. Likewise, from (56), if both users are members of $\mathcal{U}_{\mathrm{SP}}$, then the amount of interference that user (j,m) causes to user (ℓ,k) in the UL is $\beta_{\ell,j,m}^2/\left((C_u-\tau)\lambda^2\right)$. Therefore, $I_{j,m}^{\mathrm{TP-ul}}$ and $I_{j,m}^{\mathrm{SP-ul}}$ can be obtained as

$$I_{j,m}^{\mathrm{TP-ul}} = \sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,k)\in\mathcal{U}_{\mathrm{TP}}}} \sum_{k=0}^{K-1} \beta_{\ell,j,k}^2\delta_{m,k} = \sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\mathrm{TP}}}} \beta_{\ell,j,m}^2 \tag{57}$$

---

[10]This assumption is made for the sake of clarity and simplicity. In the absence of this assumption, the BS will have to estimate and remove $\mathbf{Y}_j^{\mathrm{TP}}$ from $\mathbf{Y}_j$ before estimating the channels of the users in $\mathcal{U}_{\mathrm{SP}}$.

$$I_{j,m}^{\mathrm{SP-ul}} = \frac{1}{(C_u - \tau)\,\lambda^2} \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \beta_{\ell,j,m}^2 \ . \tag{58}$$

From the above equations, the total cost due to UL inter/intra-cell interference can be expressed as

$$I\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right) = \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1} \Big(I_{\ell,k}^{\mathrm{TP-ul}}\mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\mathrm{TP}}\}}$$
$$+\ I_{\ell,k}^{\mathrm{SP-ul}}\mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\mathrm{SP}}\}} \Big) \tag{59}$$

Using (59) as the objective function, the sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ can be obtained as the solution of the following optimization problem

$$\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right) = \arg\min_{\substack{\mathcal{U}_{\mathrm{TP}}\subseteq\mathcal{U} \\ \mathcal{U}_{\mathrm{SP}}\subseteq\mathcal{U}}} I\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right)$$
$$\text{subject to} \quad \mathcal{U}_{\mathrm{TP}}\cup\mathcal{U}_{\mathrm{SP}} = \mathcal{U}$$
$$\mathcal{U}_{\mathrm{TP}}\cap\mathcal{U}_{\mathrm{SP}} = \varnothing \tag{60}$$

where $\mathcal{U}$ is the set of all users in the L cells. However, the optimization problem in (60) is combinatorial in nature with $2^{\mathrm{Card}(\mathcal{U})}$ possible choices for $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$, making it computationally hard to obtain the optimal solution. A workaround is to employ a greedy approach to partition $\mathcal{U}$ into $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$. At each step of this algorithm, given $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$, a user $\left(\tilde{\ell},\tilde{k}\right)$ in $\mathcal{U}_{\mathrm{TP}}$ is chosen as

$$\left(\tilde{\ell},\tilde{k}\right) = \arg\max_{(\ell,k)\in\mathcal{U}_{\mathrm{TP}}} I_{\ell,k}^{\mathrm{TP-ul}} \ . \tag{61}$$

Setting $\mathcal{U}_{\mathrm{TP}}' = \mathcal{U}_{\mathrm{TP}}\setminus\left(\tilde{\ell},\tilde{k}\right)$ and $\mathcal{U}_{\mathrm{SP}}' = \mathcal{U}_{\mathrm{SP}}\cup\left(\tilde{\ell},\tilde{k}\right)$, user $\left(\tilde{\ell},\tilde{k}\right)$ is added to $\mathcal{U}_{\mathrm{SP}}$ if

$$I\left(\mathcal{U}_{\mathrm{TP}}',\mathcal{U}_{\mathrm{SP}}'\right) \leq I\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right) \ . \tag{62}$$

The algorithm is initialized with $\mathcal{U}_{\mathrm{TP}} = \mathcal{U}$ and is terminated when (62) is no longer satisfied or when $\mathcal{U}_{\mathrm{TP}}$ is empty. The approach described above is summarized in Algorithm 1.

The complexity of the greedy algorithm used for designing the hybrid system can be obtained as follows. The terms $I^{\mathrm{TP-ul}}$ and $I^{\mathrm{SP-ul}}$ require a maximum of $\mathrm{Card}\left(\mathcal{U}\right)$ operations to compute, and therefore, computing $I\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right)$ requires $\mathrm{Card}\left(\mathcal{U}\right)^2$ operations. Assuming that the greedy algorithm runs till the condition $\mathcal{U}_{\mathrm{TP}} = \varnothing$ is satisfied, then an upper bound on the computational complexity of the greedy algorithm is $\mathrm{Card}\left(\mathcal{U}\right)^3$ operations. Moreover, an overhead of $2\mathrm{Card}\left(\mathcal{U}\right)$ data transmissions is required for sending the large-scale path-loss coefficients to a central node and receiving the sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$.

It has to be noted that Algorithm 1 is sub-optimal, but it is useful for illustrating the concept of the hybrid system. Partitioning algorithms that offer superior performance compared to Algorithm 1 with lower coordination overhead are left as topics for future research.

Algorithm 1 Greedy algorithm to select $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$

Data: $\beta_{j,\ell,k},\ \forall j, \ell = 0,\ldots,L-1,\ \ k = 0,\ldots,K-1$
Initialize: $\mathcal{U}_{\mathrm{TP}} \leftarrow \mathcal{U},\, \mathcal{U}_{\mathrm{SP}} \leftarrow \varnothing$
1: Compute $\left(\tilde{\ell},\tilde{k}\right)$ as in (61)
2: Set $\mathcal{U}_{\mathrm{TP}}' \leftarrow \mathcal{U}_{\mathrm{TP}}\setminus\left(\tilde{\ell},\tilde{k}\right)$ and $\mathcal{U}_{\mathrm{SP}}' \leftarrow \mathcal{U}_{\mathrm{SP}}\cup\left(\tilde{\ell},\tilde{k}\right)$
3: **if** $\mathcal{U}_{\mathrm{TP}} \neq \varnothing$ **and if** $I\left(\mathcal{U}_{\mathrm{TP}}',\mathcal{U}_{\mathrm{SP}}'\right) \leq I\left(\mathcal{U}_{\mathrm{TP}},\mathcal{U}_{\mathrm{SP}}\right)$ **then**
4: $\quad \mathcal{U}_{\mathrm{TP}} := \mathcal{U}_{\mathrm{TP}}',\, \mathcal{U}_{\mathrm{SP}} := \mathcal{U}_{\mathrm{SP}}'$
5: $\quad$ Return to Step (1).
6: **else**
7: $\quad$ STOP
8: **end if**

VIII. SIMULATION RESULTS

We compare the UL SINR and UL bit-error rate (BER) performance of the LS-based and (in some examples) eigen-value decomposition (EVD)-based methods that use time-multiplexed pilots to the performance of the channel estimator that uses superimposed pilots, at the output of a MF that employs these channel estimates. Two scenarios are considered for this comparison.

Scenario 1: The users are uniformly distributed in hexagonal cells of radius 1km with the BS at the center. In addition, users are located at a distance of at least 100m from the BS.

Scenario 2: Users in both the reference and interfering cells are in a fixed configuration and are equally spaced on a circle of a given radius with the BS in the center. The size of the hexagonal cell is 1km and unless otherwise specified, the users are on a circle of radius 800m.

Unless otherwise specified, the following parameters are used in both scenarios. The channel estimation methods are tested with $L = 7$ cells and $K = 5$ users per cell. A P-QAM constellation is employed and the path-loss coefficient is assumed to be 3. The simulations for the superimposed pilots-based iterative channel estimation scheme have been performed for 4 iterations. The number of symbols in the uplink time slot C_u is set to 100, and for computing the rate, C is set to 200 symbols. The values of ρ and λ are computed from (31) and (32), respectively, and ω is set to 1, where ω is the design parameter in the statistics-aware power control scheme. The signal-to-noise ratio (SNR), i.e., ω/σ^2 is set to 10dB. The methods based on time-multiplexed pilots have been simulated with $r = 1$ and $p_u = 1$. In addition, the chosen channel estimation methods have been observed to perform better with the statistics aware-power control scheme, and therefore, this power control scheme has been employed for both time-multiplexed and superimposed pilots. The plots in Scenario 1 are generated by averaging over 10^4 realizations of user locations across the cell. For each realization of user location, the channel vectors are generated and 200 bits are transmitted per user. The BER is computed by counting the bit errors for all the users in the reference cell. Similarly, the plots in Scenario 2 are generated for a fixed user location by averaging over 10^4 channel realizations with 200 bits transmitted per user for each realization.
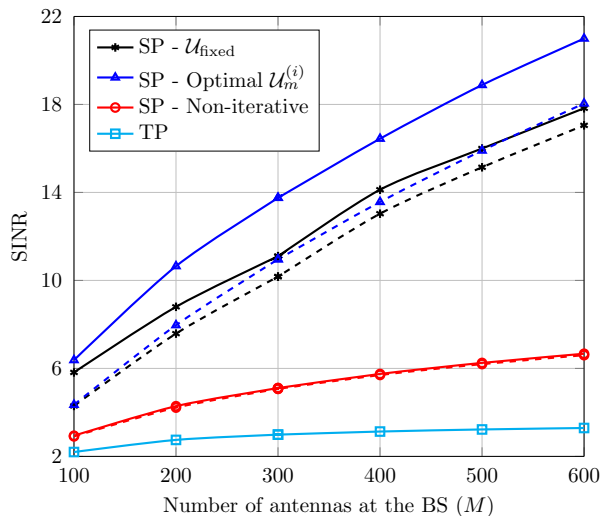
Fig. 2. The UL SINR of a user in the reference BS vs. M in Scenario 2. The values of ρ and λ are computed from (31) and (32), respectively, and since they are approximations, they result in a non-smooth SINR behavior for the iterative methods. The solid and dashed lines represent simulated and theoretical curves, respectively.
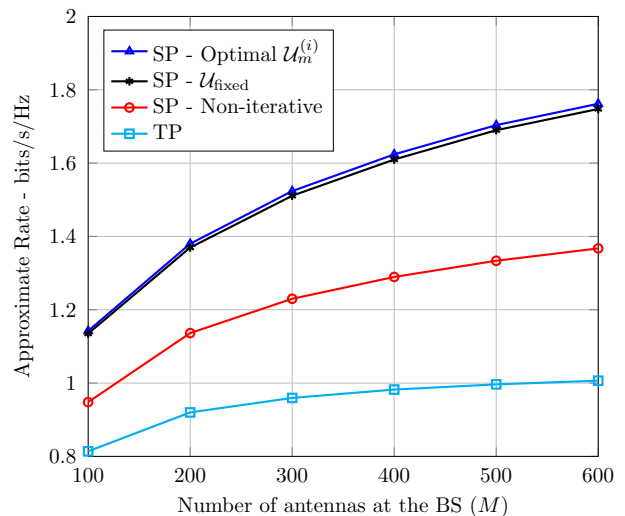

Fig. 3. Approximate per-user UL rate obtained using a 16-QAM constellation vs. M in the reference BS in Scenario 2. The maximum UL rate that can be achieved with the 16-QAM constellation, with half the symbols in a coherence block used for UL transmission, is 2 bps/Hz.
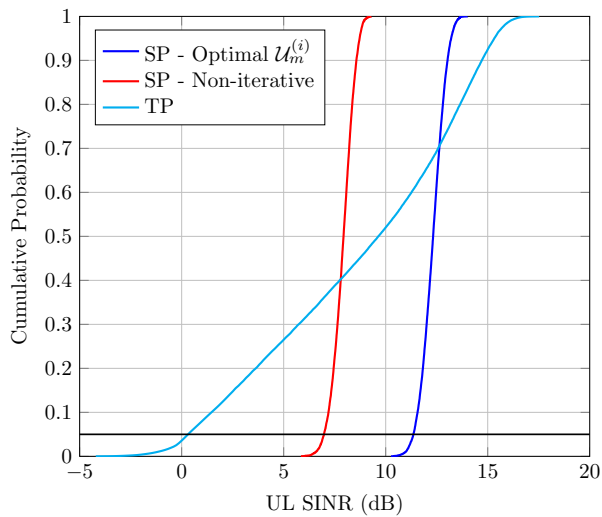

Fig. 4. Cumulative distribution of the UL SINR in dBs for users in Scenario 1 with $M = 300$ antennas. The black line indicates SINRs with probability ≥ 0.95.

Fig. 5. BER in the UL vs. K in Scenario 1 with $M/K = 50$ and $C_u = 70$ symbols.

Fig. 2 shows the variation of the UL SINR of an arbitrary user with respect to M in Scenario 2, whereas in Fig. 3, the approximate rate of an arbitrary user, calculated using 16-QAM constellation, is plotted for the same scenario. We compute the achievable rate for 16-QAM signaling, modeling a practical scenario where highly mobile users are requesting moderate-to-high data rates. The SINR when the proposed method is employed, is shown to linearly increase in the number of antennas, whereas the SINR performance is observed to saturate for the LS-based method that uses time-multiplexed pilots. This trajectory of the proposed method could be potentially maintained using techniques such as adaptive modulation and coding, thereby implying that the effects of pilot contamination can be eliminated.
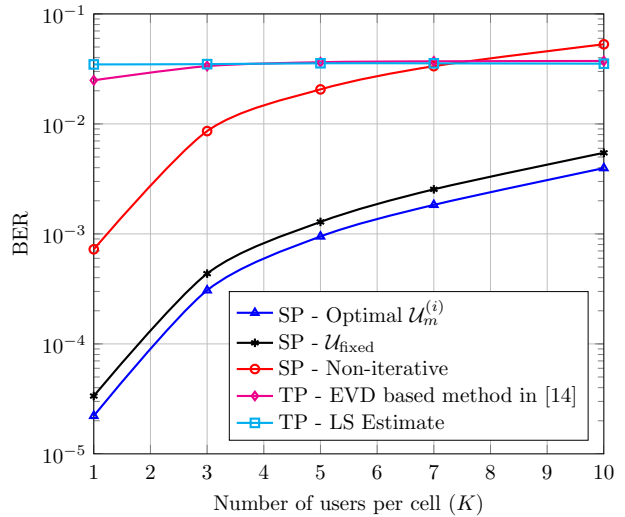
In Fig. 4, the cumulative distribution of the UL SINR in Scenario 1 is plotted. The interference power is averaged over 100 channel and data realizations for each realization of user location. While the LS-based method employing time-multiplexed pilots offers a higher SINR than the LS method employing superimposed pilots with a probability of approximately 0.6, the latter method can be seen to offer a significantly higher minimum SINR compared to the former method. Moreover, the users employing superimposed pilots have a smaller variation in their SINR than those employing time-multiplexed pilots. This is because the SINR of a user when superimposed pilots are employed is limited by the interference from the other users in the same cell, and the statistics-aware power control scheme renders the intra-cell

interference power independent of the user location within the cell. The iterative method based on superimposed pilots is observed to offer a remarkably higher SINR performance with respect to its non-iterative counterpart and the LS-based method employing time-multiplexed pilots.

In Fig. 5, the BER is plotted against the number of users per cell in Scenario 1, with K ranging from 1 to 10 and $C_u = 70$ symbols. Since $L = 7$ cells, $K = 10$ implies that the superimposed pilot-based system cannot support any new users without sharing pilots across cells. The ratio M/K is set to 50. While the non-iterative channel estimator based on superimposed pilots performs better in the UL at lower values of K than the estimators based on time-multiplexed pilots, the non-iterative estimator performs poorly at higher values of K. This is because the data transmitted alongside the pilots causes self-interference and this interference power increases with the number of users in the system. Therefore, it is necessary to resort to iterative techniques to mitigate this additional interference and it can be seen that the iterative methods offer a better performance than methods based on time-multiplexed pilots when LK is close to C_u.

In Fig. 6, the users are distributed as in Scenario 2 and the distance of the users from the BS is varied between 0.2 and 0.9 km. For the chosen range of user distance, the total rate in the UL is plotted against the corresponding received signal-to-interference ratio (SIR). The received SIR of an arbitrary user m in cell j is defined as

$$\mathrm{SIR}_j^{\mathrm{Rx}} \triangleq \frac{\omega}{\sum\limits_{\ell \neq j} \sum\limits_{k} \beta_{j,\ell,k}^2} \ . \tag{63}$$

We assume $L = 19$ hexagonal cells, i.e., a central cell with two tiers of interfering cells. Each cell has $M = 1000$ antennas, $K = 5$ users, and the value of C_u is chosen as 40 symbols. Although L is set to 19, the optimization described in Algorithm 1 and the computation of the performance metrics is performed over 7 cells which consist of the central and the first tier of cells. The value of ω for users in $\mathcal{U}_{\mathrm{SP}}$ is set to 10 and p_u for the users in $\mathcal{U}_{\mathrm{TP}}$ is set to 1. The data symbols are Gaussian distributed and the sum rate in Fig. 6 is obtained by averaging over 10^3 realizations of the channel and data symbols.

In Fig. 6, high and low values of SIR correspond to users located close to the BS and at the cell-edge, respectively. It can be observed that channel estimation methods based only on superimposed pilots (even the non-iterative formulation) are better in high interference scenarios, i.e., when the interfering users are at the cell-edge, whereas time-multiplexed pilots are better in low-interference scenarios. This behavior is a direct consequence of Theorem 1 since higher interference scenarios have smaller values of κ, resulting in superimposed pilots outperforming time-multiplexed pilots. However, at smaller values of user radius, the impact of pilot contamination is low but the self-interference in superimposed pilots resulting from transmitting the data alongside the pilots leads to a poorer performance compared to methods based on time-multiplexed pilots. In addition, it can be seen that the hybrid system adapts to the level of inter and intra-cell interference and offers a
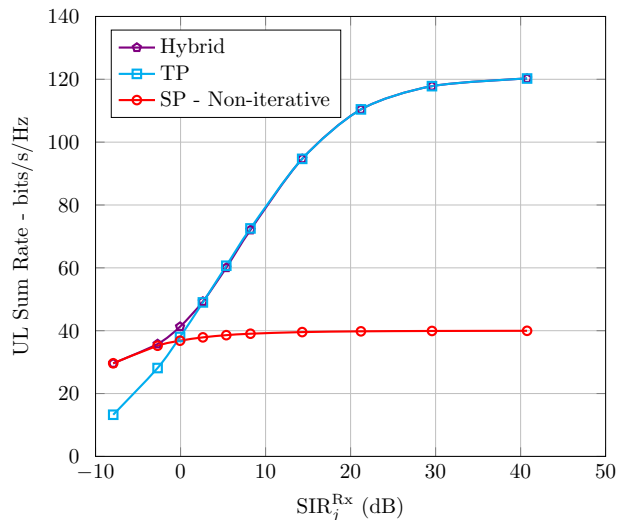

Fig. 6. UL sum rate vs. $\mathrm{SIR}_j^{\mathrm{Rx}}$ in Scenario 2 with $M = 1000$ antennas.

performance that is resilient to the location of the user within the cell.

IX. CONCLUSION AND DISCUSSION

We have proposed superimposed pilots as a superior alternative to time-multiplexed data and pilots for uplink channel estimation in massive MIMO. In the limit of an infinite number of antennas, a hybrid system using both superimposed pilots and time-multiplexed data and pilots offers a higher UL rate and supports larger number of users than the optimal system that utilizes only time-multiplexed data and pilots. The resilience to pilot contamination can be significantly improved with superimposed pilots through the use of an iterative data-aided channel estimation scheme that utilizes the data symbols of both the desired and interfering users in the feedback loop. Computer simulations in both a realistic scenario, in which users are distributed uniformly over the entire cell, and a high-interference scenario, in which users are concentrated at the cell edge, show that channel estimation methods using superimposed pilots offer a significant performance improvement over those that use time-multiplexed pilots.

The objective of this paper is to advocate superimposed pilots for practical use in massive MIMO systems by showing their superiority through theoretical and simulation based investigations. In standard MIMO communications, superimposed pilots are typically argued to be useful only for the scenario with high user mobility, and therefore, have not found practical application. On the contrary, in massive MIMO, superimposed pilots in a hybrid system provide superior performance in general. Therefore, there is a strong reason for superimposed pilots to make their way to practical use.

The proposed iterative data-aided channel estimation scheme and the greedy algorithm for partitioning users are suboptimal algorithms for corresponding non-convex problems. Algorithms that offer performance close to the optimal at low computational complexities and overheads are of interest for future research. Moreover, the downlink performance of superimposed pilots is another topic of practical importance, which we have partially addressed in [28].

APPENDIX A

Uplink SINR of the Non-Iterative Channel Estimation Method

Using the notation described in Section V-A, (7) can be rewritten as

$$\mathbf{Y} = \sum_{m=0}^{N-1} \mathbf{h}_m \left(\rho_m \mathbf{x}_m + \lambda_m \mathbf{p}_m \right)^T + \mathbf{W} \tag{64}$$

From (9), the estimation error of the channel estimate can be obtained as

$$\Delta \mathbf{h}_m \triangleq \mathbf{h}_m - \widehat{\mathbf{h}}_m = -\frac{1}{C_u \lambda_m} \left(\sum_{k=0}^{N-1} \rho_k \mathbf{h}_k \mathbf{x}_k^T + \mathbf{W} \right) \mathbf{p}_m^* . \tag{65}$$

From (10) and (65), the estimate of the received data after MF with the estimated channel can be written as

$$\begin{aligned}
\widetilde{\mathbf{x}}_m^T &= \frac{1}{M \rho_m \beta_m} \widehat{\mathbf{h}}_m^H \left(\mathbf{Y} - \lambda_m \widehat{\mathbf{h}}_m \mathbf{p}_m^T \right) \\
&= \frac{1}{M \rho_m \beta_m} \left(\mathbf{h}_m^H - \Delta \mathbf{h}_m^H \right) \\
&\quad \times \left(\sum_{k=0}^{N-1} \mathbf{h}_k \left(\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k \right)^T + \mathbf{W} \right. \\
&\quad \left. -\lambda_m \left(\mathbf{h}_m - \Delta \mathbf{h}_m \right) \mathbf{p}_m^T \right) = \mathbf{g}^T + \mathbf{i}^T
\end{aligned} \tag{66}$$

where $\mathbf{g}$ and $\mathbf{i}$ are the signal and interference components of the matched filtered signal, respectively, which can be written as

$$\mathbf{g} \triangleq \frac{\|\mathbf{h}_m\|^2}{M \beta_m} \mathbf{x}_m \tag{67}$$

$$\mathbf{i} \triangleq \sum_{n=1}^{5} \mathbf{i}_n \tag{68}$$

$$\mathbf{i}_1 \triangleq \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \frac{\mathbf{h}_m^H \mathbf{h}_n}{M \rho_m \beta_m} \left(\lambda_n \mathbf{p}_n + \rho_n \mathbf{x}_n \right) + \frac{\left(\mathbf{h}_m^H \mathbf{W} \right)^T}{M \rho_m \beta_m} \tag{69}$$

$$\mathbf{i}_2 \triangleq \frac{\lambda_m}{M \rho_m \beta_m} \mathbf{h}_m^H \Delta \mathbf{h}_m \mathbf{p}_m \tag{70}$$

$$\mathbf{i}_3 \triangleq -\frac{1}{M \beta_m} \Delta \mathbf{h}_m^H \mathbf{h}_m \mathbf{x}_m \tag{71}$$

$$\mathbf{i}_4 \triangleq -\frac{1}{M \rho_m \beta_m} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \Delta \mathbf{h}_m^H \mathbf{h}_n \left(\lambda_n \mathbf{p}_n + \rho_n \mathbf{x}_n \right) - \frac{\left(\Delta \mathbf{h}_m^H \mathbf{W} \right)^T}{M \rho_m \beta_m} \tag{72}$$

$$\mathbf{i}_5 \triangleq -\frac{\lambda_m}{M \rho_m \beta_m} \|\Delta \mathbf{h}_m\|^2 \mathbf{p}_m . \tag{73}$$

The average interference power can be found as

$$\mathbb{E} \left\{ \|\mathbf{i}\|^2 \right\} = \mathbb{E} \left\{ \left\| \sum_{n=1}^{5} \mathbf{i}_n \right\|^2 \right\} . \tag{74}$$

Then, using the definitions of $\mathbf{i}_n, \forall n$ in (69) – (73) and the definition of $\Delta \mathbf{h}_m$ in (65), the following expressions can be easily obtained

$$\mathbb{E} \left\{ \|\mathbf{i}_1\|^2 \right\} \approx \frac{C_u}{M \rho_m^2 \beta_m} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \beta_n \mu_n \tag{75}$$

$$\begin{aligned}
\mathbb{E} \left\{ \|\mathbf{i}_2\|^2 + \|\mathbf{i}_3\|^2 + \|\mathbf{i}_4\|^2 \right\} &\approx \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \sum_{\substack{k=0 \\ k \neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M \lambda_m^2 \rho_m^2 \beta_m^2} \\
&\quad + \sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \sum_{\substack{p=0 \\ p \neq m \\ n \neq p}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_p \beta_n}{C_u \rho_m^2 \lambda_m^2 \beta_m^2}
\end{aligned} \tag{76}$$

$$\mathbb{E} \left\{ \|\mathbf{i}_5\|^2 \right\} \approx \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \sum_{\substack{p=0 \\ p \neq m \\ n \neq p}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_n \beta_p}{C_u \lambda_m^2 \rho_m^2 \beta_m^2} \tag{77}$$

$$\mathbb{E} \left\{ \left(\mathbf{i}_3^H + \mathbf{i}_4^H \right) \mathbf{i}_5 \right\} \approx - \sum_{\substack{p=0 \\ p \neq m}}^{N-1} \sum_{\substack{n=0 \\ n \neq m \\ n \neq p}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_n \beta_p}{C_u \lambda_m^2 \rho_m^2 \beta_m^2} \tag{78}$$

where the approximation errors in (75) – (78) are proportional to either N/M, N/C_u, or C_u/M. In addition, the remaining terms of the form $\mathbf{i}_n^H \mathbf{i}_p, \forall\, n \neq p$ in the expansion of (74) are proportional to N/M or N/C_u. If M is large with respect to N and C_u, then the approximation errors and terms that are proportional to N/M and N/C_u can be neglected. Similarly, error terms that are proportional to N/C_u can also be dropped, and if $\sigma^2 \ll C_u$, then the effect of noise can also be neglected. Then, substituting (75) – (78) into the expansion of (74), the interference power is obtained as

$$\begin{aligned}
\mathbb{E} \left\{ \|\mathbf{i}\|^2 \right\} &\approx \sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \frac{C_u \beta_n \mu_n}{M \rho_m^2 \beta_m} \\
&\quad + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \sum_{\substack{k=0 \\ k \neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M \lambda_m^2 \rho_m^2 \beta_m^2} .
\end{aligned} \tag{79}$$

Using (79), the SINR can be obtained as

$$\begin{aligned}
\mathrm{SINR}_m^{\mathrm{SP-ul}} &\triangleq \frac{\mathbb{E} \left\{ \|\mathbf{g}\|^2 \right\}}{\mathbb{E} \left\{ \|\mathbf{i}\|^2 \right\}} \\
&= \frac{C_u}{\sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \frac{C_u \beta_n \mu_n}{M \rho_m^2 \beta_m} + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \sum_{\substack{k=0 \\ k \neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M \lambda_m^2 \rho_m^2 \beta_m^2}} .
\end{aligned} \tag{80}$$

It completes the derivation of (12).

APPENDIX B

Interference Power of the Iterative Method

To derive the SINR, using the definition of $\Delta \mathbf{x}_m^{(i)} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_m^{(i)}$, the channel estimate in (38) can be simplified as

$$\begin{aligned}
\widehat{\mathbf{h}}_m^{(i)} = \mathbf{h}_m + \frac{1}{C_u \lambda_m} &\left(\sum_k \rho_k \mathbf{h}_k \mathbf{x}_k^T - \sum_{k \in \mathcal{U}_j, k < m} \rho_k \widehat{\mathbf{h}}_k^{(i)} \left(\widehat{\mathbf{x}}_k^{(i)} \right)^T \right. \\
&\left. - \sum_{k \in \mathcal{U}_j, m \leq k \leq N} \rho_k \widehat{\mathbf{h}}_k^{(i-1)} \left(\widehat{\mathbf{x}}_k^{(i-1)} \right)^T + \mathbf{W}_j \right) \mathbf{p}_m^*
\end{aligned} \tag{81}$$

where

$$\Delta \mathbf{h}_m^{(i)} = -\frac{1}{C_u \lambda_m} \left(\sum_{k \in \mathcal{U}_j, k < m} \rho_k \left\{ \mathbf{h}_k \left(\Delta \mathbf{x}_k^{(i)} \right)^T + \Delta \mathbf{h}_k^{(i)} \mathbf{x}_k^T \right. \right.$$
$$\left. - \Delta \mathbf{h}_k^{(i)} \left(\Delta \mathbf{x}_k^{(i)} \right)^T \right\} + \sum_{k \in \mathcal{U}_j, m \le k \le N} \rho_k \left\{ \mathbf{h}_k \left(\Delta \mathbf{x}_k^{(i-1)} \right)^T + \Delta \mathbf{h}_k^{(i-1)} \mathbf{x}_k^T \right.$$
$$\left. - \Delta \mathbf{h}_k^{(i-1)} \left(\Delta \mathbf{x}_k^{(i-1)} \right)^T \right\} + \left. \sum_{k \notin \mathcal{U}_j} \rho_k \mathbf{h}_k \mathbf{x}_k^T + \mathbf{W} \right) \mathbf{p}_m^* . \tag{82}$$

The received symbols after MF in (39) are then given as

$$\widehat{\mathbf{x}}_m^T = \frac{1}{M \rho_m} \left(\mathbf{h}_m^H - \left(\Delta \mathbf{h}_m^{(i)} \right)^H \right) \left(\sum_{k=0}^{N-1} \mathbf{h}_k (\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k)^T \right.$$
$$\left. + \mathbf{W} - \lambda_m \left(\mathbf{h}_m - \Delta \mathbf{h}_m^{(i)} \right) \mathbf{p}_m^T \right)$$
$$= \frac{1}{M} \|\mathbf{h}_m\|^2 \mathbf{x}_m^T + \sum_{k=1}^{7} \mathbf{a}_k^T \tag{83}$$

where

$$\mathbf{a}_1 \triangleq \frac{1}{M \rho_m} \sum_{\substack{k=0 \\ k \ne m}}^{N-1} \mathbf{h}_m^H \mathbf{h}_k (\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k) \tag{84}$$

$$\mathbf{a}_2 \triangleq \frac{1}{M \rho_m} \left(\mathbf{h}_m^H \mathbf{W} \right)^T \tag{85}$$

$$\mathbf{a}_3 \triangleq \frac{\lambda_m}{M \rho_m} \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \mathbf{p}_m \tag{86}$$

$$\mathbf{a}_4 \triangleq -\frac{1}{M} \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \mathbf{x}_m \tag{87}$$

$$\mathbf{a}_5 \triangleq -\frac{1}{M \rho_m} \sum_{\substack{k=0 \\ k \ne m}}^{N-1} \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_k (\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k) \tag{88}$$

$$\mathbf{a}_6 \triangleq -\frac{1}{M \rho_m} \left(\left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{W} \right)^T \tag{89}$$

$$\mathbf{a}_7 \triangleq -\frac{\lambda_m}{M \rho_m} \left\| \Delta \mathbf{h}_m^{(i)} \right\|^2 \mathbf{p}_m . \tag{90}$$

Under the assumption that the interference power at each of the received symbols is the same, the average interference power of the m'th user at the j'th cell is given as

$$I_m^{(i)} = \frac{1}{C_u} \mathbb{E} \left\{ \left\| \sum_{k=1}^{7} \mathbf{a}_k \right\|^2 \right\} \approx \frac{1}{C_u} \left[\mathbb{E} \left\{ \sum_{k=1}^{5} \|\mathbf{a}_k\|^2 \right\} \right] \tag{91}$$

where the terms $\mathbf{a}_6$, $\mathbf{a}_7$, and $\mathbf{a}_p^H \mathbf{a}_q$, $\forall p, q$ have been dropped. Further, it can be shown straightforwardly that

$$\mathbb{E} \left\{ \|\mathbf{a}_1\|^2 \right\} = \frac{C_u}{M \rho_m^2} \sum_{\substack{k=0 \\ k \ne m}}^{N-1} \beta_k \beta_m \tag{92}$$

$$\mathbb{E} \left\{ \|\mathbf{a}_2\|^2 \right\} = \frac{C_u \sigma^2 \beta_m}{M \rho_m^2} . \tag{93}$$

Moreover, $\mathbb{E} \left\{ \|\mathbf{a}_3\|^2 \right\}$, $\mathbb{E} \left\{ \|\mathbf{a}_4\|^2 \right\}$, and $\mathbb{E} \left\{ \|\mathbf{a}_5\|^2 \right\}$ can be written as

$$\mathbb{E} \left\{ \|\mathbf{a}_3\|^2 \right\} = \frac{\lambda_m^2}{M^2 \rho_m^2} \mathbb{E} \left\{ \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \mathbf{p}_m^T \mathbf{p}_m^* \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \right\}$$
$$= \frac{C_u \lambda_m^2}{M^2 \rho_m^2} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \right\} \tag{94}$$

$$\mathbb{E} \left\{ \|\mathbf{a}_4\|^2 \right\} = \frac{1}{M^2} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \mathbf{x}_m^T \mathbf{x}_m^* \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \right\}$$
$$= \frac{1}{M^2} \mathbb{E} \left\{ \mathbf{x}_m^T \mathbf{x}_m^* \right\} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \right\}$$
$$= \frac{C_u}{M^2} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta \mathbf{h}_m^{(i)} \right\} \tag{95}$$

and

$$\mathbb{E} \left\{ \|\mathbf{a}_5\|^2 \right\} = \frac{1}{M^2 \rho_m^2} \sum_{\substack{\ell=0 \\ \ell \ne m}}^{N-1} \sum_{\substack{k=0 \\ k \ne m}}^{N-1} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \mathbf{h}_\ell \mathbf{h}_k^H \Delta \mathbf{h}_m^{(i)} \right\}$$
$$\times \mathbb{E} \left\{ (\rho_\ell \mathbf{x}_\ell + \lambda_\ell \mathbf{p}_\ell)^H (\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k) \right\}$$
$$= \frac{C_u}{M^2 \rho_m^2} \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \left(\sum_{\substack{k=0 \\ k \ne m}}^{N-1} \mathbf{h}_k \mathbf{h}_k^H \right) \Delta \mathbf{h}_m^{(i)} \right\} . \tag{96}$$

Summing up (94), (95), and (96), we obtain

$$\mathbb{E} \left\{ \sum_{k=3}^{5} \|\mathbf{a}_k\|^2 \right\} = \frac{C_u}{M^2 \rho_m^2}$$
$$\times \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \left(\sum_{k=0}^{N-1} \mathbf{h}_k \mathbf{h}_k^H \right) \Delta \mathbf{h}_m^{(i)} \right\} . \tag{97}$$

Now, let $\psi_m^{(i)}$ be defined as the second term in (97), i.e.,

$$\psi_m^{(i)} \big|_{i \ge 1} \triangleq \mathbb{E} \left\{ \left(\Delta \mathbf{h}_m^{(i)} \right)^H \left(\sum_{n=0}^{N-1} \mathbf{h}_n \mathbf{h}_n^H \right) \Delta \mathbf{h}_m^{(i)} \right\} . \tag{98}$$

Using (82) and the simplifications (S1) to (S4), (98) can be simplified to obtain (42). Substituting (92), (93), (97), and (42) into (91), $I_m^{(i)}$ can be obtained as

$$I_m^{(i)} \approx \frac{1}{M \rho_m^2} \sum_{\substack{k=0 \\ k \ne m}}^{N-1} \beta_k \beta_m + \frac{\sigma^2 \beta_m}{M \rho_m^2} + \frac{1}{M^2 \rho_m^2} \psi_m^{(i)} . \tag{99}$$

It completes the derivation of (41).

APPENDIX C
Choice of the Set of Users $\mathcal{U}_m^{(i)}$

Let $\mathcal{S}$ be a set of the KL users in the system and let $\mathcal{P}(\mathcal{S})$ be its power set. In addition, for the sake of clarity, let the additional argument $\mathcal{U}_m^{(i)}$ be added to the functions $I_m^{(i)}$ and $\psi_m^{(i)}$ in this section. Now, the optimal set $\mathcal{U}_m^{(i)}$ can be obtained by solving the following optimization problem

$$\mathcal{U}_m^{(i)} = \arg \min_{\mathcal{U} \in \mathcal{P}(\mathcal{S})} \left\{ I_m^{(i)} (\mathcal{U}) \right\} . \tag{100}$$

Substituting (41) into (100) yields

$$\mathcal{U}_m^{(i)} = \arg \min_{\mathcal{U} \in \mathcal{P}(\mathcal{S})} \left\{ \psi_m^{(i)} (\mathcal{U}) \right\} . \tag{101}$$

Now, $\psi_m^{(i)}(\mathcal{U})$ can be rewritten as

$$\psi_m^{(i)}(\mathcal{U}) = c + \sum_{n=0}^{N-1} \left\{ \xi_n \mathbf{1}_{\{n \notin \mathcal{U}\}} + \epsilon_n^{(i)}(\mathcal{U}) \mathbf{1}_{\{n \in \mathcal{U}, n < m\}} \right.$$
$$\left. + \epsilon_n^{(i-1)}(\mathcal{U}) \mathbf{1}_{\{n \in \mathcal{U}, n \geq m\}} \right\} \quad (102)$$

where c, ξ_n, and $\epsilon_n^{(i)}(\mathcal{U})$ are defined as

$$c \triangleq \frac{M\sigma^2}{C_u \lambda_m^2} \left(\sum_{k=0}^{N-1} \beta_k \right) \quad (103)$$

$$\xi_n \triangleq \frac{M^2 \rho_n^2}{C_u \lambda_m^2} \left\{ \beta_n^2 + \frac{1}{M} \sum_{k=0}^{N-1} \beta_k \beta_n \right\} \quad (104)$$

$$\epsilon_n^{(i)}(\mathcal{U}) \triangleq \frac{M^2 \rho_n^2}{C_u \lambda_m^2} \left\{ \beta_n^2 \alpha_n^{(i)} + \sum_{k=0}^{N-1} \frac{1}{M} \beta_k \beta_n \alpha_n^{(i)} \right.$$
$$\left. + \frac{\left(1 + \alpha_n^{(i)}\right)}{M^2} \psi_n^{(i)}(\mathcal{U}) \right\} . \quad (105)$$

It can be seen from (102) that the optimization problem (101) is separable over the user indices, implying that the decision to include user n in $\mathcal{U}_m^{(i)}$ is independent of the other $N-1$ users. Therefore, the channel and data estimates of user n are used in the i'th iteration if the following condition is satisfied

$$n \in \mathcal{U}_m^{(i)} \text{ iff } \psi_m^{(i)} \big|_{n \in \mathcal{U}_m^{(i)}} < \psi_m^{(i)} \big|_{n \notin \mathcal{U}_m^{(i)}} . \quad (106)$$

From (102) and (106), the set $\mathcal{U}_m^{(i)}$ is obtained as

$$\mathcal{U}_m^{(i)} = \left\{ n \in \mathbb{N} \,\middle|\, \epsilon_n^{(i)}(\mathcal{U}) < \xi_n \text{ when } n < m \right.$$
$$\left. \text{and } \epsilon_n^{(i-1)}(\mathcal{U}) < \xi_n \text{ when } n \geq m \right\} . \quad (107)$$

Equivalently, using (104) and (105), the above expression simplifies to

$$\mathcal{U}_m^{(i)} = \left\{ n \in \mathbb{N} \,\middle|\, \alpha_n^{(i)} < \gamma_n^{(i)} \text{ when } n < m \right.$$
$$\left. \text{and } \alpha_n^{(i-1)} < \gamma_n^{(i-1)} \text{ when } n \geq m \right\} \quad (108)$$

where

$$\gamma_n^{(i)} \triangleq \frac{\left\{ \beta_n^2 + \frac{1}{M} \sum_{k=0}^{N-1} \beta_n \beta_k - \frac{\psi_n^{(i)} \big|_{n \in \mathcal{U}_m^{(i)}}}{M^2} \right\}}{\left\{ \beta_n^2 + \frac{1}{M} \sum_{k=0}^{N-1} \beta_n \beta_k + \frac{\psi_n^{(i)} \big|_{n \in \mathcal{U}_m^{(i)}}}{M^2} \right\}} . \quad (109)$$

If $\mathbf{x}_m$ takes values from the P-QAM constellation, then substituting (45) into (108), the set $\mathcal{U}_m^{(i)}$ can be obtained as

$$\mathcal{U}_m^{(i)} = \left\{ n \in \mathbb{N} \,\middle|\, I_n^{(i)} < f_n^{(i)} \text{ when } n < m \right.$$
$$\left. \text{and } I_n^{(i)} < f_n^{(i-1)} \text{ when } n \geq m \right\} \quad (110)$$

where $f_n^{(i)}$ is defined as

$$f_n^{(i)} \triangleq \frac{3}{P-1} Q^2 \left(\frac{\sqrt{P}\left(\sqrt{P}+1\right)\gamma_n^{(i)}}{24} \right) . \quad (111)$$

However, since the decision rules are based on approximate SINR expressions, it is worth commenting that the reliability of the decision rule in (108) decreases with increasing user and iteration indices. Alternatively, a fixed and conservative decision rule can be used to obtain $\mathcal{U}$ as follows

$$\mathcal{U}_{\text{fixed}} = \left\{ m \in \mathbb{N} \,\middle|\, I_m^{(2)}\left(\{m\}\right) < I_m^{(2)}\left(\varnothing\right) = I_m^{(1)}\left(\varnothing\right) \right\} . \quad (112)$$

The decision rule in (112) results in a set $\mathcal{U}_{\text{fixed}}$ that is computed at the beginning of the first iteration and is left unchanged for the subsequent iterations.

APPENDIX D
Derivation of $\alpha_m^{(i)}$ for a P-QAM constellation

For P-QAM constellation and $i \geq 1$, the integral over $\mathbf{x}_m$ in (44) reduces to a summation, which can be written as

$$\alpha_m^{(i)} = \sum_{x \in \chi} \int |x - \eta(x-e)|^2 p_{\mathbf{e}_m^{(i)}}(e) p_{\mathbf{x}_m}(x) \, de . \quad (113)$$

Since the P symbols are equally likely, $p_{\mathbf{x}_m}(x) = 1/P$, $\forall\, x$ and under the assumption that the errors $x - \eta(x-e)$ are dominated by the closest neighboring symbols, the above equation reduces to

$$\alpha_m^{(i)} = \frac{1}{P} \sum_{x \in \chi} d_x^2 k_x Q \left(\frac{\frac{d_x}{2}}{\sqrt{\frac{I_m^{(i)}}{2}}} \right) \quad (114)$$

where d_x is the distance between the symbol x and its closest neighbor and k_x is the number of symbols at a distance of d_x from x. The Q-function in the above equation results from the assumption on the statistics of $\mathbf{e}_m^{(i)}$. For a unit-power P-QAM constellation, $d_x = \sqrt{6/P-1}$, $\forall x$ [29]. In addition, it can be easily verified that $k_x = 2$ for the 4 corner symbols, $k_x = 3$ for the $(\sqrt{P}-2)4$ symbols on the outer edges, and $k_x = 4$ for the remaining $P - 4\sqrt{P} + 4$ symbols. Substituting these values into (114) yields

$$\alpha_m^{(i)}\big|_{i \geq 1} = \frac{24}{\sqrt{P}\left(\sqrt{P}+1\right)} Q \left(\sqrt{\frac{\frac{3}{(P-1)}}{I_m^{(i)}}} \right) . \quad (115)$$

Moreover, since $\Delta\mathbf{x}_{\ell,m}^{(0)} = \mathbf{x}_{\ell,m}$, the value of $\alpha_m^{(0)}$ is 1. It completes the derivation of (45).

REFERENCES

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

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

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

[4] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[8] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

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

[10] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[11] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[12] J. Jose, A. Ashikhmin, T. 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.

[13] H. Q. Ngo and E. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Kyoto*, Mar. 2012, pp. 3249–3252.

[14] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[15] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3421–3425.

[16] 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.

[17] K. Upadhya and S. A. Vorobyov, "An array processing approach to pilot decontamination for massive MIMO," in *Proc. IEEE 6th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Dec. 2015, pp. 453–456.

[18] 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.

[19] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[20] S. He, J. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, Jun. 2007.

[21] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[22] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.

[23] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM), St. Louis*, vol. 4, Dec. 2005, pp. 5 pp.–2233.

[24] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.

[25] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, Mar. 2016, pp. 3366–3370.

[26] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[27] C. Cox, *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*, 2nd ed. Chichester, UK: John Wiley & Sons, 2012.

[28] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination," in *Proc. IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, Washington D.C., Dec. 2016, pp. 665–669.

[29] J. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, 2008.

Karthik Upadhya received the B.E. degree in electronics and communication engineering from Visvesvaraya Technological University, India in 2007 and the M.Tech degree in communication systems from Indian Institute of Technology Madras, India in 2011. He is currently a Ph.D. student at Aalto University, Finland. His research interests include wireless communications and array signal processing. Before joining the Ph.D. program, he worked as a Member of Technical Staff at Saankhya Labs, India from 2011 to 2013 and as a research assistant at the Indian Institute of Science (IISc), India from 2013 to 2014.

Sergiy A. Vorobyov (M'02-SM'05) received the M.Sc. and Ph.D. degrees in systems and control from Kharkiv National University of Radio Electronics, Ukraine, in 1994 and 1997, respectively. He is a Professor with the Department of Signal Processing and Acoustics, Aalto University, Finland. He has been previously with the University of Alberta, Alberta, Canada as an Assistant Professor from 2006 to 2010, Associate Professor from 2010 to 2012, and became Full Professor there in 2012. Since his graduation, he also held various research and faculty positions at Kharkiv National University of Radio Electronics, Ukraine; the Institute of Physical and Chemical Research (RIKEN), Japan; McMaster University, Canada; Duisburg-Essen University and Darmstadt University of Technology, Germany; and the Joint Research Institute between Heriot-Watt University and Edinburgh University, U.K. His research interests include optimization and liner algebra methods in signal processing and communications; statistical and array signal processing; sparse signal processing; estimation and detection theory; sampling theory; and multi-antenna, very large, cooperative, and cognitive systems.

Dr. Vorobyov is a recipient of the 2004 IEEE Signal Processing Society Best Paper Award, the 2007 Alberta Ingenuity New Faculty Award, the 2011 Carl Zeiss Award (Germany), the 2012 NSERC Discovery Accelerator Award, and other awards. He is serving as Area Editor for IEEE Signal Processing Letters since 2016. He served as an Associate Editor for the IEEE Transactions on Signal Processing in 2006–2010 and the IEEE Signal Processing Letters in 2007–2009. He was a member of the Sensor Array and Multi-Channel Signal Processing and Signal Processing for Communications and Networking Technical Committees of the IEEE Signal Processing Society in 2007–2012 and 2010–2016, respectively. He has served as the Track Chair for Asilomar 2011, Pacific Grove, CA; the Technical Co-Chair for IEEE CAMSAP 2011, Puerto Rico; the Tutorial Chair for ISWCS 2013, Ilmenau, Germany; and the Technical Co-Chair for IEEE SAM 2018, Sheffield, U.K.

Mikko Vehkapera (M'10) received the Ph.D. degree from Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2010. Between 2010–2013 he was a post-doctoral researcher at School of Electrical Engineering, and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Sweden, and 2013–2015 an Academy of Finland Postdoctoral Researcher at Aalto University School of Electrical Engineering, Finland. He is now an assistant professor (lecturer) at University of Sheffield, Department of Electronic and Electrical Engineering, United Kingdom. He held visiting appointments at Massachusetts Institute of Technology (MIT), US, Kyoto University and Tokyo Institute of Technology, Japan, and University of Erlangen-Nuremberg, Germany. His research interests are in the field of wireless communications, information theory and signal processing. Dr. Vehkapera was a co-recipient for the Best Student Paper Award at IEEE International Conference on Networks (ICON2011) and IEEE Sweden Joint VT-COM-IT Chapter Best Student Conference Paper Award 2015.