Superimposed Pilots are Superior for Mitigating Pilot Contamination in Massive MIMO
Part I: Theory and Channel Estimation

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

Abstract—Superimposed pilots are introduced, in this two-part paper, as an alternative to time-multiplexed pilot and data symbols for mitigating pilot contamination in massive multi-input multiple-output (MIMO) systems. In this part of the two-part paper, a non-iterative scheme for uplink channel estimation based on superimposed pilots is first proposed and an expression for the uplink signal-to-interference-plus-noise ratio (SINR), at the output of a matched filter employing this channel estimate, is derived. Based on this expression, it is observed that the quality of the channel estimate can be improved by reducing the interference that results from transmitting data alongside the pilots, and an iterative data-aided scheme is also proposed that reduces this component of interference. Approximate expressions for the uplink SINR are provided for the iterative data-aided method as well. In addition, it is shown 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. Numerical simulations demonstrating the superiority of the proposed channel estimation schemes based on superimposed pilots and the accuracy of the approximations are also given in this part.

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 feedback 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 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 [19–23]. 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], especially in scenarios with high user-mobility, wherein it is impractical to allocate dedicated symbols for training. Most recently, superimposed pilots have been employed in the context of multi-cell multiuser MIMO systems [24] without realizing that it provides a superior solution to the pilot contamination problem in massive MIMO.

In the context of multi-cell massive MIMO, 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 propose a novel low-complexity iterative channel estimation scheme for superimposed pilots and demonstrate its effectiveness by deriving a closed-form approximation for the UL signal-to-interference-plus-noise ratio (SINR) at the output of an MF-based detector. 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.

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. Section VII presents simulation results demonstrating the effectiveness of employing superimposed pilots for pilot decontamination. Section VIII concludes the paper. Some of the proofs are detailed in Appendix.

**Notation:** Lower case and upper case boldface letters denote column vectors and matrices, respectively. The notations \( (\cdot)^{\star}, (\cdot)^{T}, (\cdot)^{H}, \) and \( (\cdot)^{-1} \) represent the conjugate transpose, Hermitian transpose, and inverse, respectively. The notation \( \mathcal{CN}(\mu, \Sigma) \) stands for the complex normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \) 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 \( \emptyset \) denotes the empty set. The notation \( \mathbf{1}_{|S|} \) represents the indicator function over the set \( S \), whereas \( |S| \) 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 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

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

where \( \mu_{\ell,k} \) denotes the power with which the vector of symbols \( \mathbf{s}_{\ell,k} \in \mathbb{C}^{C_u \times 1} \) is transmitted by user \( k \) in cell \( \ell \). \( \mathbf{W}_j \in \mathbb{C}^{M \times C_u} \) is the additive white Gaussian noise at BS \( j \) with each column distributed as \( \mathcal{CN}(0, \sigma^2 \mathbf{I}_M) \). 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 between the antennas at BS \( j \), and user \( k \) in cell \( \ell \), and is assumed to be distributed as

\[
\mathbf{h}_{j,\ell,k} \sim \mathcal{CN}(0, \beta_{j,\ell,k} \mathbf{I}_M)
\]

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 \( \forall j, \ell, k \). These statistics of the channel vector represent a non-line-of-sight scenario with rich scattering. By virtue of their zero mean and mutual independence, the channel vectors are asymptotically orthogonal and the following equation holds almost surely, i.e.,

\[
\lim_{M \to \infty} \frac{\mathbf{h}_{j,\ell,k}^{H} \mathbf{h}_{m,n,p}}{M} = \delta_{j,\ell,k} \delta_{m,k} \delta_{n,p}, \quad \forall j, \ell, m, n, p .
\]

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 that depends on user mobility. For the sake of simplicity, the effects of shadowing are not taken into account in this paper. 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 \( \chi \).

---

\(^1\)The work in this paper is a significant extension of our relevant conference paper [23]. 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.

\(^2\)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 are no longer valid.
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. Let the pilot sequences transmitted by the users be the elements of a pilot book $B$ consisting of orthogonal sequences of length $\tau$ and defined as $[18]$:

$$B \triangleq \{ \phi_1, \ldots, \phi_T \} \text{ where } \phi_n \phi_{n+\tau} = \tau \delta_{n,p}. \quad (4)$$

If $\phi_{b_i \ell \tau} \in B$ is the pilot sequence transmitted by user $k$ of cell $\ell$, where $b_i \ell \subset \{1, \ldots, \tau\}$ is the index of the transmitted pilot in the pilot book $B$, and if each pilot sequence is reused once every $r \triangleq \tau / K$ cells $[18]$, the least-squares (LS) estimate of the channel of user $m$ in cell $j$ can be obtained as $[11]$

$$\hat{h}_{j,m}^{\tau \tau} \triangleq \frac{1}{\tau \sqrt{P_u}} Y_j^{(p)} \phi_{b_i \ell \tau} = h_{j,m} + \sum_{\ell = 0}^{\tau-1} h_{j,\ell,m} + \frac{1}{\tau \sqrt{P_u}} W_j \phi_{b_i \ell \tau} \quad (5)$$

where the superscript TP indicates that the estimates are computed when using time-multiplexed pilots and $L_j(r)$ is the subset of the $L$ cells that use the same set of pilots as cell $j$. In addition, it is assumed 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 $\beta$. It can be observed in $[5]$ 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 $|L_j(r) - 1|$ cells. When $M \to \infty$, the UL SINR of user $m$ in cell $j$, at the output of a MF that uses the channel estimate in $[5]$ for detection, can be written as $[11]$

$$\text{SINR}^{\tau \tau}_{j,m} = \frac{\beta_2^{j,m}}{\sum_{\ell = 0}^{\tau-1} \beta_2^{j,\ell,m}}. \quad (6)$$

The corresponding throughput of the user using Gaussian signaling in the UL can then be expressed as $[11, 18]$

$$R_{j,m}^{\tau \tau} = \left( \frac{C_u - \tau}{C_u} \right) \log_2 \left( 1 + \text{SINR}^{\tau \tau}_{j,m} \right). \quad (7)$$

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

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., $KL \leq C_u$, then with superimposed pilots, each user can be assigned a unique orthogonal pilot $\rho_{\ell,k} \in \mathbb{C}^{x \times 1}$. The pilots are taken from the columns of an orthogonal matrix $P \in \mathbb{C}^{C_u \times C_u}$ such that $P^H P = C_u I_{C_u}$, and therefore $\rho_{\ell,k} \rho_{\ell,k}^H = C_u \delta_{k,0} \delta_{b,i}$. If $\rho_{\ell,k} x_{\ell,k} + \lambda_{\ell,k} p_{\ell,k}$ is the transmitted vector from user $k$ in cell $\ell$, then the received signal at the $j$’th BS $Y_j \in \mathbb{C}^{M \times C_u}$, when using the superimposed pilot scheme, can be written as

$$Y_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} h_{j,\ell,k} (\rho_{\ell,k} x_{\ell,k} + \lambda_{\ell,k} p_{\ell,k})^T + W_j \quad (8)$$

where $p_{\ell}^k$ and $\lambda_{\ell}^k$ 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 $\mu_{\ell,k} = \lambda_{\ell,k}^2 + p_{\ell,k}^2$.

A. Non-Iterative Channel Estimation

Treating the data symbols of all users as additive noise, the channel estimate of user $k$ in cell $\ell$ can be obtained at the $j$’th BS using the LS criterion $[24]$

$$\hat{h}_{j,\ell,k} = \arg \min_{h} \| Y_j - \lambda_{\ell,k} h \rho_{\ell,k}^T \|_F^2. \quad (9)$$

Solving $[9]$ yields

$$\hat{h}_{j,\ell,k} = Y_j (\lambda_{\ell,k}^2 \rho_{\ell,k}^T \rho_{\ell,k})^{-1} \lambda_{\ell,k} \rho_{\ell,k}^T = \frac{1}{C_u \lambda_{\ell,k}} Y_j \rho_{\ell,k}^T = h_{j,\ell,k} + \frac{1}{C_u \lambda_{\ell,k}} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \rho_{\ell,m} h_{j,m,n} x_{m,n}^T \rho_{\ell,k}^T + \frac{1}{C_u \lambda_{\ell,k}} W_j \rho_{\ell,k}^T. \quad (10)$$

In order to estimate the data from the received observations, it is necessary to remove the term corresponding to the transmitted superimposed pilot $\lambda_{j,m} \hat{h}_{j,j,m} \rho_{j,m}^T$ from the observation vector in $[9]$. Using $\lambda_{j,m} \hat{h}_{j,j,m} \rho_{j,m}^T$ as an estimate for this term, the estimate of $x_{j,m}$ can then be obtained from the observation $Y_j$ using an MF and a decision operation as follows

$$\hat{x}_{j,m}^T = \frac{1}{M \rho_{j,m}^2 \beta_{j,m}^2} \hat{h}_{j,m}^H \left( Y_j - \lambda_{j,m} \hat{h}_{j,m} \rho_{j,m}^T \right) \quad (11)$$

$$\hat{x}_{j,m} = \eta(\hat{x}_{j,m}) \quad (12)$$

The SINR of user $m$ in cell $j$, at the output of an MF that employs the channel estimate in $[10]$, is derived in Appendix $A$ and is given in $[13]$ (shown at the top of the next page). The SINR in $[13]$, when $M \to \infty$, can be written as $\text{SINR}^{\text{SP}}_{j,m} = \frac{\lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,m}^2}{\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \mu_{\ell,k}^2 \beta_{j,k}^2}. \quad (13)$
The corresponding per-user rate in the uplink when using Gaussian signaling is given as

$$ R_{j,m}^{\text{SP-ul}} = \frac{C_u}{C} \log_2 \left( 1 + \text{SINR}_{j,m}^{\text{SP-ul}} \right). $$ (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 $\beta$ swamp users that have smaller values of $\beta$. Therefore, it becomes necessary to use power control to provide a uniform user experience.

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

$$ \max_{\mu_{j,k}, \rho_{j,k}, \lambda_{j,k}} \left\{ \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} R_{\ell,k}^{\text{SP-ul}} \right\} $$ (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 $\mu_{j,k}$, $\rho_{j,k}$, and $\lambda_{j,k}$. This suboptimal solution will be shown to maximize a lower bound on the sum-rate, and it is as follows.

The received signal in (13) can be equivalently written as

$$ Y_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{j,k}} h_{\ell,k} \left( \frac{\rho_{j,k}}{\mu_{j,k}} x_{\ell,k}^T + \sqrt{\mu_{j,k}} \lambda_{j,k} p_{\ell,k}^T \right) + W_j $$

$$ = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \tilde{h}_{\ell,k} x_{\ell,k}^T + \lambda_{j,k} \tilde{p}_{\ell,k}^T + W_j $$ (16)

where

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

$$ \tilde{\beta}_{j,k} \triangleq \tilde{\beta}_{j,k} : \mu_{j,k} $$ (18)

$$ \tilde{\rho}_{\ell,k} \triangleq \sqrt{\frac{\rho_{j,k}}{\mu_{j,k}}} > 0 $$ (19)

$$ \tilde{\lambda}_{\ell,k} \triangleq \sqrt{\frac{\lambda_{j,k}^2}{\mu_{j,k}}} > 0 $$ (20)

$$ \tilde{\lambda}_{\ell,k}^2 + \tilde{\rho}_{\ell,k}^2 = 1. $$ (21)

From (16), it can be seen that a system having arbitrary values of $\beta_{j,k}$, $\mu_{j,k}$, $\rho_{j,k}$, and $\lambda_{j,k}$, can be reduced into an equivalent system with parameters $\tilde{\beta}_{j,k}$, $\tilde{\rho}_{\ell,k}$, and $\tilde{\lambda}_{\ell,k}$, such that $0 \leq \tilde{\rho}_{\ell,k}, \tilde{\lambda}_{\ell,k} \leq 1$. Substituting (18) – (21) into (13), an equivalent expression for the SINR, as shown in (23) (shown in the top of the next page) can be obtained. +1

To obtain the parameter $\mu_{j,k}$, we propose using the statistics-aware power-control approach detailed in [18], wherein user $m$ in cell $j$ transmits at a power $\mu_{j,m} = \omega / \beta_{j,i,m}$ where $\omega$ is a design parameter. The parameter $\omega$ 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 $\omega$ at the $j$'th BS for all the users in cell $j$. In addition, as mentioned in [18], the ratio $0 \leq \beta_{j,k,l} / \beta_{j,l,k} \leq 1$ 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 $\mu_{j,k} = \omega / \beta_{j,k,l}$ and using the definitions of $\beta_{j,k,l}$, $\tilde{p}_{\ell,k}$, and $\lambda_{j,k}$, and the inequality $0 \leq \beta_{j,k,l} / \beta_{j,l,k} \leq 1$, the following equations can be obtained

$$ \beta_{j,j,m} = \beta_{j,j,m} \cdot \mu_{j,m} = \omega $$ (22)

$$ \beta_{j,l,k} = \beta_{j,l,k} \cdot \mu_{j,k} \leq \beta_{j,l,k} \mu_{j,k} = \omega \quad \forall \ell \neq j $$ (23)

$$ \rho_{j,m}^2 \beta_{j,j,m} = \rho_{j,m}^2 \omega $$ (24)

$$ \lambda_{j,m}^2 \beta_{j,j,m} = \lambda_{j,m}^2 $$ (25)

$$ \rho_{\ell,k}^2 \beta_{j,l,k} \leq \rho_{\ell,k}^2 \beta_{j,l,k} = \rho_{\ell,k}^2 \omega \quad \forall \ell \neq j $$ (26)

$$ \lambda_{\ell,k}^2 \beta_{j,l,k} \leq \lambda_{\ell,k}^2 \beta_{j,l,k} = \lambda_{\ell,k}^2 \omega \quad \forall \ell \neq j. $$ (27)

Substituting the above equations into (23), a lower bound on the SINR, as shown in (28) (shown at the top of the next page), can be obtained. +1

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 $\tilde{\rho}_{\ell,k}$ and $\tilde{\lambda}_{\ell,k}$ and requires coordination between the BSs. To circumvent this problem, we restrict the parameters $\tilde{\rho}_{\ell,k}$ and $\tilde{\lambda}_{\ell,k}$ such that $\tilde{\rho}_{\ell,k} = \tilde{\rho}, \forall \ell, k$ and $\tilde{\lambda}_{\ell,k} = \tilde{\lambda}, \forall \ell, k$ in (30), we obtain

$$ \text{SINR}_{j,m}^{\text{SP-ul}} \simeq \left( \frac{L K}{C_u (1 - \tilde{\rho}^2)} + \frac{1}{M} \left( \frac{L K - 1}{\tilde{\rho}^2} + \frac{(L K - 1)^2}{C_u (1 - \tilde{\rho}^2)} \right) \right)^{-1}. $$ (28)

Differentiating the right hand side of (28) with respect to $\tilde{\rho}^2$
where the approximations in (29) and (30) have been made.

\[
\overline{\rho}^2 = \left(1 + \frac{L\beta}{\sqrt{\lambda}}\right)^{-1} \approx \left(1 + \sqrt{\frac{C_u}{M + Lk}}\right)^{-1}
\]

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

\[
\bar{\lambda}^2_{\text{opt}} = 1 - \overline{\rho}^2_{\text{opt}} \approx \left(1 + \sqrt{\frac{C_u}{M + Lk}}\right)^{-1}
\]

where the approximations in (29) and (30) have been made assuming \( LK \gg 1 \) in order to obtain simpler expressions.

Based on the fact established in this subsection that systems using \( \rho, \lambda, \bar{\beta}, \bar{\lambda}, \bar{\beta}, \bar{h} \) 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 \( \mu_{\ell,k} = \bar{\rho}_{\ell,k} \) and \( \bar{\lambda}_{\ell,k} = \bar{\lambda}_{\ell,k} \).

Remark 1: It can be seen that as \( M \rightarrow \infty \), \( \overline{\rho}^2_{\text{opt}} \rightarrow 0 \) and \( \bar{\lambda}^2_{\text{opt}} \rightarrow 1 \) as also pointed out in [24]. While this behavior does not significantly improve the UL SINR performance, it results in an asymptotically interference free DL, which will be described in Part II [25] of this paper.

C. Impact of \( C_u \) on the Performance of Superimposed Pilots

Using (13) and a fixed set of parameters \( r, \tau, \) 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 \( \tau \) and if \( M \rightarrow \infty \), 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 (6) and (13) are equal, i.e.,

\[
\frac{1}{\kappa_{j,m}} \sum_{\ell=0}^{L-1} k_{j,m} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \beta_{j,\ell,m}^2 = \sum_{\ell \in \mathcal{L}_j} \beta_{j,\ell,m}^2
\]

then it is evident from (13) and (31) that \( C_u > \kappa_{j,m} \) is a sufficient condition for a method that is based on superimposed pilots to outperform the LS method that employs time-multiplexed pilots. In addition, \( \kappa_{j,m} \) is given as

\[
\kappa_{j,m} \triangleq \frac{1}{\lambda_{j,m}^2} \sum_{\ell=0}^{L-1} k_{j,m} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \beta_{j,\ell,m}^2 \sum_{\ell \in \mathcal{L}_j} \beta_{j,\ell,m}^2.
\]

This completes the proof.

Remark 2: 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 \( r = 1, \beta_{j,\ell,m} = 1, \forall m, \beta_{j,\ell,m} = \beta, \forall \ell \neq j, m, \) and \( \rho_{j,m} = \lambda_{j,m}^2, \forall j, m \). For such a scenario, \( \kappa_{j,m} \) is given as

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

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 \( \beta^2 \). This dependence on \( C_u \) is evident from the expression for the channel estimation error, which is given as

\[
\Delta h_{j,m} \equiv h_{j,m} - \hat{h}_{j,m} = -\frac{1}{C_u \lambda_{j,m}^2} \sum_{\ell=0}^{L-1} k_{j,m} \rho_{\ell,k} h_{j,\ell,k} x_{\ell,k}^* W_j p_{j,m}.
\]

Remark 3: We build upon the discussion in [11] 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 [11], 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 4: In addition, 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.

From (34), 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-ADDED CHANNEL ESTIMATION

In the iterative approach to channel estimation that we develop 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 (34). In addition, to minimize error propagation between the channel estimates of different users, the iteration is started from the user with the higher SINR and is progressed in the decreasing order of the SINRs of the users.

A. Algorithm

For the sake of clarity and without loss of generality, we replace the two indices $k, \ell$ with a single index $m$ that lies in the range $0 \leq m \leq N - 1$, where $N \equiv K L$. 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, (8) can be rewritten as

$$Y = \sum_{m=0}^{N-1} h_m (\rho_m x_m + \lambda_m p_m)^T + W.$$

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., $\beta_0 > \beta_1 > \ldots > \beta_{N-1}$. Then, using an estimate of $\rho_m h_m x_m^T p_m$ 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

$$\tilde{h}_m^{(i)} = \frac{1}{C_u \lambda_m} \left[ Y - \sum_{k=0}^{m-1} \rho_k \hat{h}_k^{(i)} (\hat{x}_k^{(i)})^T 
- \sum_{k=m}^{N-1} \rho_k \hat{h}_k^{(i-1)} (\hat{x}_k^{(i-1)})^T \right] p_m^*$$

where $\hat{h}_m^{(0)} = 0$, $\forall m$ and $U_m^{(i)}$ 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 channel estimate in the above equation is a modified version of the LS estimator defined in (10) with an added correction factor. Utilizing the resulting channel estimate in an MF and decision operation, similar to (11) and (12), the estimate of the data is obtained as follows

$$\hat{x}_m^{(i)} = \frac{1}{M \rho_m \beta_m} \left( \tilde{h}_m^{(i)} \right)^H \left( Y - \lambda_m \hat{h}_m^{(i)} p_m^T \right)$$

$$\hat{x}_m^{(i)} = \eta \hat{x}_m^{(i)}$$

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

Remark 5: From (10), the non-iterative method for channel estimation has a per-user complexity of $O(C_u)$, whereas the MF and hard-decision operations in (11) and (12) have complexities of $O(M)$ and $O(C_u)$, respectively. The complexity for the non-iterative method is similar to that of the LS method for time-multiplexed pilots.

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

B. Interference Power at the BS

Let $e_m^{(i)} = x_m - \hat{x}_m^{(i)}$ be the error in the estimate of the data symbols of user $m$ obtained from the MF in the $i$’th iteration. Let $\Delta x_m^{(i)} = x_m - \hat{x}_m^{(i)}$ be the corresponding error vector after the hard-decision operation and let $\Delta h_m^{(i)} = h_m - \hat{h}_m^{(i)}$ be the associated error in the channel estimate. If $\alpha_m^{(i)}$ is the variance of the elements of $\Delta x_m^{(i)}$ and assuming that the elements of $e_m^{(i)}$ 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

$$I_m^{(i)} \approx \frac{1}{\beta_m^2} \left( \frac{1}{M \rho_m^2} \sum_{k=0}^{N-1} \beta_k \beta_m + \frac{\sigma^2 \beta_m}{M \rho_m^2} + \frac{1}{M^2 \rho_m^2} \psi^{(i)} \right)$$

where the expression for $\psi^{(i)}$ is given in (47) on the top of the next page and $\psi^{(0)} = 0$, $\forall m$. The detailed derivation of $I_m^{(i)}$ can be found in Appendix B.

+1 In deriving (39), the following simplifying assumptions have been made in order to obtain a closed form expression:

(S1) $e_m^{(i)}$ is independent of $x_k$ and $W$, $\forall k, i$.

(S2) $\Delta x_k^{(i)}$ is independent of $x_k$, $W$, and $h_k$, $\forall k, i$.

(S3) $\Delta x_k^{(i)}$ is independent of $\Delta x_k^{(p)}$, $\forall p \neq i, m \neq k$ and the elements of $\Delta x_m^{(i)}$ are i.i.d.

(S4) $\Delta h_k^{(i)}$ is independent of $x_k$, $W$, and $\Delta x_k^{(p)}$, $\forall k, p$.

In scenarios with low interference and with large $M$, the elements of $\Delta x_m^{(i)}$ are sparse, and the elements of $e_m^{(i)}$ take small values, resulting in the simplifications [S1] [S2] and [S3] being reasonably accurate. Although the expression for $\Delta h_m^{(i)}$ (given in (77) in Appendix B) is explicitly dependent on $x_k$ and $\Delta x_k^{(i)}$, we neglect the correlation between these terms since $\Delta h_m^{(i)}$ 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 $e_m^{(i)}$ is assumed to be a zero-mean random variable, $\Delta x_k^{(i)}$ is also a zero-mean random variable, provided the constellation points in $\chi$ and their probability density
functions are symmetric about the origin. This is true since by definition, $Δx_k^{(i)}$ and $e_k^{(i)}$ are related to each other through the following equation

$$Δx_k^{(i)} = x_k - \eta (x_k - e_m^{(i)}) .$$  \hspace{1cm} (40)

From (40), an expression for the variance of the elements of $Δx_k^{(i)}$, i.e., $α_k^{(i)}$ can be found as

$$α_k^{(i)} \triangleq E \left\{ \left[ Δx_k^{(i)} \right]^2 \right\} = \int \int \left| x - η (x - e) \right|^2 p_{Δx_k^{(i)}} (x) dx \, dx$$

$$= \int \int \left| x - \eta (x - e) \right|^2 p_{e_k^{(i)}, x_k} (e, x) \, dx$$

$$= \int \int \left| x - \eta (x - e) \right|^2 p_{e_k^{(i)}} (e) p_{x_k} (x) \, dx$$  \hspace{1cm} (41)

where $p_{e_k^{(i)}} (\cdot)$, $p_{Δx_k^{(i)}} (\cdot)$, and $p_{x_k} (\cdot)$ are the probability density functions of the elements of $e_k^{(i)}$, $Δx_k^{(i)}$, and $x_k$, respectively, and $p_{e_k^{(i)}, x_k} (\cdot)$ is the joint density function of the random variables $e_k^{(i)}$ and $x_k$. The latter has been written as the product of their individual distributions in the final expression of (41), thanks to (31).

**Important example of $α_m^{(i)}$:** When the elements of $x_m$ are uniformly distributed and take values from a unit-power $P$-quaternary amplitude modulation (QAM) constellation, then under the assumption that the symbol errors in $Δx_k^{(i)}$ are dominated by the closest neighboring symbols, the expression for $α_m^{(i)}$ can be written as

$$α_m^{(i)} = \begin{cases} \frac{2^q}{\sqrt{p + 1}} Q \left( \sqrt{\frac{p - 1}{p + 1} I_m^n} \right), & i \geq 1 \\ 1, & i = 0 \end{cases}$$  \hspace{1cm} (42)

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

C. Choice of the Set of Users $U_m^{(i)}$

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

$$U_m^{(i)} = \arg \min_{U \in P(S)} \left\{ I_m^{(i)} (U) \right\} .$$  \hspace{1cm} (43)

Substituting (39) into (43) yields

$$U_m^{(i)} = \arg \min_{U \in P(S)} \left\{ \psi_m^{(i)} (U) \right\} .$$  \hspace{1cm} (44)

Now, $ψ_m^{(i)} (U)$ can be rewritten as

$$ψ_m^{(i)} (U) = c + \sum_{k=0}^{N-1} \left\{ ξ_k 1_{\{ k \notin U \}} + e_k^{(i)} (U) 1_{\{ k \in U, k < m \}} \right\}$$

$$+ e_k^{(i-1)} (U) 1_{\{ k \in U, k \geq m \}}$$  \hspace{1cm} (45)

where $c$, $ξ_k$, and $e_k^{(i)} (U)$ are defined as

$$c \triangleq Mσ^2 C_n^2 \left( \sum_{n=0}^{N-1} \beta_n \right)$$

$$ξ_k \triangleq M^2 \rho^2 \left( \sum_{n=0}^{N-1} \beta_n \beta_k \right)$$

$$e_k^{(i)} (U) \triangleq M^2 \rho^2 \left( \sum_{n=0}^{N-1} \beta_n \beta_k \right)$$

$$+ \left( 1 + α_k^{(i)} \right) ψ_k^{(i)} (U) .$$  \hspace{1cm} (48)

It can be seen from (45) that the optimization problem (44) is separable over the user indices, implying that the decision to include user $k$ in $U_m^{(i)}$ is independent of the other $N - 1$ users. Therefore, the channel and data estimates of user $k$ are used in the $i$'th iteration if the following condition is satisfied

$$k \in U_m^{(i)} \quad \text{if} \quad \psi_m^{(i)} |_{k \notin U_m^{(i)}} < \psi_m^{(i)} |_{k \in U_m^{(i)}} .$$  \hspace{1cm} (49)

From (45) and (49), the set $U_m^{(i)}$ is obtained as

$$U_m^{(i)} = \left\{ k \in \mathbb{N} \left| e_k^{(i)} (U) < ξ_k \text{ when } k < m \right. \right\}$$

$$\text{and} \quad e_k^{(i-1)} (U) < ξ_k \text{ when } k \geq m \} .$$  \hspace{1cm} (50)

Equivalently, using (47) and (48), the above expression simplifies to

$$U_m^{(i)} = \left\{ k \in \mathbb{N} \left| α_k^{(i)} < γ_k^{(i)} \text{ when } k < m \right. \right\}$$

$$\text{and} \quad α_k^{(i-1)} < γ_k^{(i-1)} \text{ when } k \geq m \}$$  \hspace{1cm} (51)

where

$$γ_k^{(i)} \triangleq \begin{cases} \frac{β_0}{M} + \sum_{n=0}^{N-1} β_k β_n - \frac{ψ_k^{(i)} |_{k \notin U_m^{(i)}}}{M} \\frac{β_0}{M} + \sum_{n=0}^{N-1} β_k β_n + \frac{ψ_k^{(i)} |_{k \in U_m^{(i)}}}{M} \end{cases} \hspace{1cm} (52)$$

\[ \]
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, \( \tau \) 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 \( M \rightarrow \infty \), 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 \( U_{\text{TP}} \) and \( U_{\text{SP}} \). The users in the set \( U_{\text{TP}} \) employ time-multiplexed pilots, with parameters selected using approaches such as in [18]. The users in the set \( U_{\text{SP}} \) maintain radio silence during the pilot training phase of the users in \( U_{\text{TP}} \), i.e., for \( \tau \) 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 \( \tau \) symbols, they do not affect the quality of the channel estimates of the users in \( U_{\text{TP}} \). As a result, under the assumption of asymptotic orthogonality of the channels, the per-cell sum-rate in the UL for the users in \( U_{\text{TP}} \) remains unchanged and can be found from (12) to be

\[
R_j^{\text{UL}} (U_{\text{TP}}) = \left( \frac{C_u - \tau}{C} \right) \sum_{k=0}^{K-1} \log_2 \left( 1 + \frac{\beta^2}{\sum_{\ell \in L_j, k} \beta_{j,m}^2} \right),
\]

(56)

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

\[
R_j^{\text{UL}} (U_{\text{SP}}) = \left( \frac{C_u - \tau}{C} \right) \sum_{m \in U_{\text{SP}}} \log_2 (1 + \text{SINR}_m (U_{\text{SP}}))
\]

(57)

\[
\text{SINR}_m (U_{\text{SP}}) \triangleq \frac{\beta^2}{\sum_{k \in U_{\text{SP}}} \sum_{\ell \in L_j, k} \beta_{j,m}^2}.
\]

(58)

In obtaining the above expression, it has been assumed that the transmit power \( p_u \) of the users in \( U_{\text{TP}} \) is small enough such that the interference to the users in \( U_{\text{SP}} \) can be neglected. Therefore, from (56) and (58), the combined rate \( R_j^{\text{UL}} (U_{\text{TP}}) + R_j^{\text{UL}} (U_{\text{SP}}) \) is strictly greater than \( R_j^{\text{UL}} (U_{\text{TP}}) \). 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 \( U_{\text{SP}} \) can contain a maximum of \( C_u - \tau \) users. This concludes the proof.

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 sum-rate\(^7\) let \( K \) be the

\(^7\)Such as the scheme described in [18].
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 4-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 $\rho$ and $\lambda$ are computed from (29) and (30), respectively, and $\omega$ is set to 1, where $\omega$ is the design parameter in the statistics-aware power control scheme. The signal-to-noise ratio (SNR), i.e., $\omega/\sigma^2$ is set to 10dB. The methods based on time-multiplexed pilots have been simulated with $r = 1$ and $p_u = 1$. In addition, these 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 shows the variation of the 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 a 16-QAM constellation, is plotted for the same scenario. It has to be noted that we resort to computing the approximate rate using a 16-QAM constellation since expressions for calculating $U_{m}^{(i)}$ are not available when Gaussian signaling is employed. The SINR, and hence the throughput, 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 could be eliminated.

In Figs. 4 and 5 the BER is plotted for different values of $M$ in Scenarios 1 and 2, respectively. The BER of the proposed scheme is significantly lower than the BER of the LS and EVD-based methods that use time-multiplexed pilots.

In Fig. 6 the cumulative distribution of the SINR in dBs for users in Scenario 1 with $M = 300$ antennas. The purple line indicates SINRs with probability $\geq 0.95$.

In Fig. 7 the BER is plotted in Scenario 2 against the user radius for $M = 300$ antennas. It can be seen that channel estimation methods based on superimposed pilots 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 $\kappa$, 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 than methods based on time-multiplexed pilots. Since it is clear from Fig. 7 that superimposed pilots and time-multiplexed pilots are better than each other in different scenarios, a hybrid system constructed by assigning time-multiplexed pilots to users that are close to the BS and superimposed pilots to users that are at the cell edge would result in a win-win situation for both sets of users. We describe this partitioning in further detail in Part II [26] of this paper.

VIII. CONCLUSION

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 only utilizes time-multiplexed data and pilots. In addition, a system that only uses superimposed pilots benefits from a higher value of $C_u$, with its performance superseding that of
a system that only employs time-multiplexed pilots, when $C_u$ exceeds a threshold that is dependent on the channel scenario.

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.

APPENDIX A

Uplink SINR of the Non-Iterative Channel Estimation Method

Using the notation described in Section II.A, (58) can be rewritten as

$$Y = \sum_{m=0}^{N-1} h_m (\rho_m x_m + \lambda_m p_m)^T + W$$  \hspace{1cm} (59)

From (10), the estimation error of the channel estimate can be obtained as

$$\Delta h_m \triangleq h_m - \hat{h}_m = -\frac{1}{C_u\lambda_m} \left( \sum_{k=0}^{N-1} \rho_k x_k^T + W \right) p_m^T .$$ \hspace{1cm} (60)

From (11) and (60), the estimate of the received data after MF with the estimated channel can be written as

$$\tilde{x}_m^T = \frac{1}{M\rho_m^2\beta_m} \left( h_m^H - \Delta h_m^H \right) \times \left( \sum_{k=0}^{N-1} \rho_k x_k + \lambda_k p_k \right)^T + W$$

$$-\lambda_m (h_m - \Delta h_m) p_m^T = g^T + i^T$$ \hspace{1cm} (61)

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

$$g \triangleq \frac{\|h_m\|^2}{M\beta_m} x_m$$ \hspace{1cm} (62)

$$i \triangleq \sum_{n=1}^{N-1} i_n$$ \hspace{1cm} (63)

$$i_1 \triangleq \sum_{n=0}^{N-1} \frac{h_m^H h_n}{M\rho_m^2\beta_m} (\lambda_n p_n + \rho_n x_n) + \frac{(h_m^H W)^T}{M\rho_m^2\beta_m}$$ \hspace{1cm} (64)

$$i_2 \triangleq \frac{\lambda_m}{M\rho_m^2\beta_m} h_m^H \Delta h_m p_m$$ \hspace{1cm} (65)

$$i_3 \triangleq -\frac{1}{M\rho_m^2\beta_m} \Delta h_m^H h_m x_m$$ \hspace{1cm} (66)

$$i_4 \triangleq -\frac{1}{M\rho_m^2\beta_m} \sum_{n=0}^{N-1} \Delta h_m^H h_n (\lambda_n p_n + \rho_n x_n) - \frac{(\Delta h_m^H W)^T}{M\rho_m^2\beta_m}$$ \hspace{1cm} (67)

where the approximation errors in (70) – (73) are proportional to either $N/M$, $N/C_u$, or $C_u/M$. In addition, the remaining terms of the form $i_n^T i_p, \forall n \neq p$ in the expansion of (69) 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 (70) – (73) into the expansion of (69), the interference power is obtained as

$$\mathbb{E}\{\|i\|^2\} \approx \frac{C_u}{M\rho_m^2\beta_m} \sum_{n=0}^{N-1} \sum_{\nu \neq m} \beta_n \mu_n$$ \hspace{1cm} (74)

Using (74), the SINR can be obtained as

$$\text{SINR}_{m}^{\text{SP-ul}} \triangleq \frac{\mathbb{E}\{\|g\|^2\}}{\mathbb{E}\{\|i\|^2\}} \frac{C_u}{\sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{M\lambda_n^2 p_n^2 \beta_m^2}} + \sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{M\lambda_n^2 p_n^2 \beta_m^2} \frac{C_u}{M\rho_m^2\beta_m}$$ \hspace{1cm} (75)

(67) It completes the derivation of (13).
APPENDIX B

Interference Power of the Iterative Method

To derive the SINR, using the definition of $\Delta x^{(i)} = x_m - \hat{x}^{(i)}$, the channel estimate in (36) can be simplified as

$$\hat{h}^{(i)}_m = h_m + \frac{1}{C_u \lambda_m} \left( \sum_{k=0}^{N-1} \rho_k h_k x_k - \sum_{k=0}^{m-1} \rho_k \hat{h}^{(i)}_k (\hat{x}^{(i)}_k)^T \right)$$

$$- \sum_{k=m}^{N-1} \rho_k \hat{h}^{(i-1)}_k (\hat{x}^{(i-1)}_k)^T + W_j \right) p_m^* \right)$$

(76)

where

$$\Delta h^{(i)}_m = - \frac{1}{C_u \lambda_m} \left( \sum_{k=0}^{m-1} \rho_k \left\{ h_k (\Delta x^{(i)}_k)^T + \Delta h^{(i)}_k x_k \right\}$$

$$- \Delta h^{(i)}_k (\Delta x^{(i)}_k)^T \right\} + \sum_{k=m}^{N-1} \rho_k \left\{ h_k (\Delta x^{(i-1)}_k)^T + \Delta h^{(i-1)}_k x_k \right\}$$

$$- \Delta h^{(i-1)}_k (\Delta x^{(i-1)}_k)^T \right\} + \sum_{k=0}^{m-1} \rho_k h_k x_k + W \right) p_m^* .$$

(77)

The received symbols after MF in (37) are then given as

$$\hat{x}_m = \frac{1}{M \rho_m} \left( h_m^H \left( \Delta h^{(i)}_m \right)^H \right) \sum_{k=0}^{N-1} h_k (\rho_k x_k + \lambda_k p_k)^T$$

$$+ W - \lambda_m \left( h_m - \Delta h^{(i)}_m \right) p_m$$

$$= \frac{1}{M} \| h_m \|^2 x_m + \sum_{k=1}^{7} a_k$$

(78)

where

$$a_1 \triangleq \frac{1}{M \rho_m} \sum_{k=0}^{N-1} h_m^H h_k (\rho_k x_k + \lambda_k p_k)$$

(79)

$$a_2 \triangleq \frac{1}{M \rho_m} (h_m^H W)^T$$

(80)

$$a_3 \triangleq \frac{\lambda_m}{M \rho_m} h_m^H \Delta h^{(i)}_m p_m$$

(81)

$$a_4 \triangleq - \frac{1}{M} \left( \Delta h^{(i)}_m \right)^H h_m x_m$$

(82)

$$a_5 \triangleq - \frac{1}{M \rho_m} \sum_{k=0}^{N-1} \left( \Delta h^{(i)}_m \right)^H h_k (\rho_k x_k + \lambda_k p_k)$$

(83)

$$a_6 \triangleq - \frac{1}{M \rho_m} \left( \left( \Delta h^{(i)}_m \right)^H W \right)^T$$

(84)

$$a_7 \triangleq - \frac{\lambda_m}{M \rho_m} \left\| \Delta h^{(i)}_m \right\|^2 p_m .$$

(85)

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

$$I_m^{(i)} = \frac{1}{C_u} \mathbb{E} \left\{ \left\| \sum_{k=1}^{7} a_k \right\|^2 \right\} \approx \frac{1}{C_u} \mathbb{E} \left\{ \sum_{k=1}^{5} \| a_k \|^2 \right\}$$

(86)

where the terms $a_6$, $a_7$, and $a_8^H a_9$, $\forall p, q$ have been dropped. Further, it can be shown straightforwardly that

$$\mathbb{E} \left\{ ||a_1||^2 \right\} = \frac{C_u \lambda_m^2}{M^2 \rho_m^2} \sum_{k=0}^{N-1} \beta_k \beta_m$$

(87)

$$\mathbb{E} \left\{ ||a_2||^2 \right\} = \frac{C_u \sigma^2 \beta_m}{M \rho_m^2} .$$

(88)

Moreover, $\mathbb{E} \left\{ ||a_3||^2 \right\}$, $\mathbb{E} \left\{ ||a_4||^2 \right\}$, and $\mathbb{E} \left\{ ||a_5||^2 \right\}$ can be written as

$$\mathbb{E} \left\{ ||a_3||^2 \right\} = \frac{\lambda_m^2}{M^2 \rho_m^2} \sum_{k=0}^{N-1} \sum_{k=0}^{N-1} \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H h_m h_k^H \Delta h^{(i)}_m \right\}$$

$$= \frac{C_u \lambda_m^2}{M^2 \rho_m^2} \sum_{k=0}^{N-1} \sum_{k=0}^{N-1} \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H h_m h_k^H \Delta h^{(i)}_m \right\}$$

(89)

$$\mathbb{E} \left\{ ||a_4||^2 \right\} = \frac{1}{M^2} \left\{ \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H h_m x_m^H x_m h_m^H \Delta h^{(i)}_m \right\} \right\}$$

$$= \frac{1}{M^2} \left\{ \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H h_m x_m^H x_m h_m^H \Delta h^{(i)}_m \right\} \right\}$$

(90)

and

$$\mathbb{E} \left\{ ||a_5||^2 \right\} = \frac{1}{M^2 \rho_m^2} \sum_{k=0}^{N-1} \sum_{k=0}^{N-1} \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H \sum_{k=0}^{N-1} h_k h_k^H \Delta h^{(i)}_m \right\}$$

$$= \frac{C_u}{M^2 \rho_m^2} \left\{ \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H \sum_{k=0}^{N-1} h_k h_k^H \Delta h^{(i)}_m \right\} \right\} .$$

(91)

Summing up (89), (90), and (91), we obtain

$$\mathbb{E} \left\{ \sum_{k=3}^{5} \| a_k \|^2 \right\} = \frac{C_u}{M^2 \rho_m^2} \left\{ \mathbb{E} \left\{ \sum_{k=3}^{5} \| a_k \|^2 \right\} \right\} .$$

(92)

Now, let $\psi_m^{(i)}$ be defined as the second term in (92), i.e.,

$$\psi_m^{(i)} \triangleq \mathbb{E} \left\{ \left( \Delta h^{(i)}_m \right)^H \left( \sum_{k=0}^{N-1} h_k h_k^H \right) \Delta h^{(i)}_m \right\} .$$

(93)

Using (77) and the simplifications (S1) to (S4), (93) can be simplified to obtain (47). Substituting (87), (88), (92), and (47) into (86), $I_m^{(i)}$ can be obtained as

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

(94)

It completes the derivation of (39).
APPENDIX C

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

For P-QAM constellation and $i \geq 1$, the integral over $x_m$ in (41) reduces to a summation, which can be written as

$$\alpha_m^{(i)} = \sum_{x \in \chi} \int |x - \eta(x - e)|^2 p_{x_m}(e) p_{x_m}(x) \, dx \cdot$$ (95)

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

$$\alpha_m^{(i)} = \frac{1}{P} \sum_{x \in \chi} d_x \kappa_x Q \left( \frac{d_x}{\sqrt{2}} \right)$$ (96)

where $d_x$ is the distance between the symbol $x$ and its closest neighbor and $\kappa_x$ 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 $\alpha_m^{(i)}$. For a unit-power P-QAM constellation, $d_x = \sqrt{\eta/P - 1}, \forall x$ [27]. In addition, it can be easily verified that $\kappa_x = 2$ for the 4 corner symbols, $\kappa_x = 3$ for the $(\sqrt{P} - 2)4$ symbols on the outer edges, and $\kappa_x = 4$ for the remaining $P - 4\sqrt{P} + 4$ symbols. Substituting these values into (96) yields

$$\alpha_m^{(i)} \bigg|_{i \geq 1} = \frac{24}{\sqrt{P}} \left( \frac{d_x}{\sqrt{2}} \right) Q \left( \frac{d_x^{(i)}}{\sqrt{2}} \right)$$ (97)

Moreover, since $\Delta_{x_m}^{(0)} = x_{m_i}$, the value of $\alpha_m^{(0)}$ is 1. It completes the derivation of (42).

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. Björnson, J. Hoydis, M. Kontouris, 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. Müller, L. Cottatellucci, and M. Vehkaperä, “Blind pilot decontamination,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 773–786, Oct. 2014.
[15] 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. Upadhyaya 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.
[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] 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.
[25] K. Upadhyaya, S. A. Vorobyov, and M. Vehkaperä, “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), Mar. 2016, pp. 3366–3370.
[26] ——, “Superimposed pilots are superior for mitigating pilot contamination in massive MIMO part II: Performance analysis and optimization,” IEEE Trans. Signal Process., Submitted, Jun. 2016.
[27] J. Proakis and M. Salehi, Digital Communications. McGraw-Hill, 2008.