On Detection and Mitigation of Reused Pilots in Massive MIMO Systems

Jitendra K. Tugnait, *Fellow, IEEE*

***Abstract*— In a time-division duplex multiple antenna system the channel state information (CSI) can be estimated using reverse training. In multi-cell multi-user massive MIMO systems, pilot contamination degrades CSI estimation performance and adversely affects massive MIMO system performance. In this paper we consider a subspace-based semi-blind approach where we have training data as well as information bearing data from various users (both in-cell and neighboring cells) at the base station (BS). Existing semi-blind approaches assume that the interfering users from neighboring cells are always at distinctly lower power levels at the BS compared with the in-cell users. This requires (perfect) power control. In this paper we do not make any such assumption. Unlike existing approaches, the BS estimates the channels of all users: in-cell and significant neighboring cell users, i.e., ones with comparable power levels at the BS. We exploit both subspace method using correlation as well as blind source separation using higher-order statistics. Finally, the estimated channels are used to detect information symbols, which, in turn, are used as pseudo-pilots to re-estimate the in-cell users' channels. The proposed approach is illustrated via simulation examples and compared with some existing semi-blind methods.**

***Index Terms*— Massive MIMO, pilot contamination, multi-user channel estimation, independent component analysis.**

I. Introduction

MOBILE data traffic continues to grow at an exponential rate. To meet this data challenge, massive MIMO (multiple-input multiple-output) system technology has been proposed where the base station employs a large number of antennas, allowing many single-antenna users to be served simultaneously [1]–[5]. It is regarded as one of the key enablers of future 5G wireless systems. Successful operation of massive MIMO depends critically on knowledge of the channel state information (CSI) between the base station (BS) and the end users. In a time-division duplex (TDD) system, the downlink (DL) and uplink (UL) channels can be assumed to be reciprocal Therefore, the BS can acquire the CSI in a TDD system using reverse training, where the users send individual pilot signals to the base station during the UL operation [5]. In a given cell, the pilots are selected to be orthogonal.

Manuscript received May 26, 2017; revised August 30, 2017; accepted September 27, 2017. Date of publication October 4, 2017; date of current version February 14, 2018. This work was supported by NSF Grant ECCS-1651133. The associate editor coordinating the review of this paper and approving it for publication was R. C. de Lamare.

The author is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: tugnajk@eng.auburn.edu).

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

Digital Object Identifier 10.1109/TCOMM.2017.2759111

In a multi-cell environment, since the same orthogonal pilots are re-used among the cells due to a large number of end users. Due to pilot reuse, the channel estimates obtained at a BS contain not only the desired CSI but also components (contamination) from neighboring cells. The effect of inter-cell interference does not vanish with increasing number of antennas at the BS. This phenomenon is called pilot contamination. It degrades CSI estimation performance and adversely affects massive MIMO system performance.

Several methods have been proposed to eliminate/mitigate the effects of pilot contamination [6]–[16]; a recent survey is in [5]. Time-shifted pilots have been proposed in [7], and also considered in [15], where when users from a given cell transmit pilots in UL, the BSs in the neighboring cells transmit data in DL. This avoids pilot contamination but this method is susceptible to UL/DL power disparity because of different transmit rates (UL rates are typically much lower than the DL rates). Therefore, DL data transmit power is significantly higher than the UL pilot power, and hence, this method may not yield accurate channel estimates. Multi-cell cooperative and coordinated methods have been proposed in [8]–[10]. These approaches require cooperation among BSs, as well as knowledge and exchange of channel covariance information. Such information is not always available, and the methods may not always be feasible with increasing number of antennas because the information exchange among BSs increases with number of antennas.

Blind/semi-blind methods have been proposed in [11]–[13]. In [11] and [12] signal subspace properties, based on SVD (singular value decomposition) of data matrix [12] or on EVD (eigenvalue decomposition) of data correlation matrix [11], are used to first separate in-cell signal subspace from interference-and-noise subspace, and then one uses pilots to resolve a much lower-dimension unitary matrix ambiguity. In [13] a semi-blind approach is proposed which requires neither inter-cell cooperation nor channel statistical information. It exploits data higher-order statistics (kurtosis). In [14] the approach of [12] is augmented with angle-of-arrival properties-based spatial filtering.

The method of [15] uses both DL and UL training where the training overhead increases with the number of cells, by a factor equal to the number of cells. In [16] an iterative (turbo) joint channel estimation and data detection method is investigated. This method requires knowledge of channel covariance information and of large-scale fading coefficients.

Thus, existing approaches differ in the underlying assumptions and availability of information: just training data, or training data plus information symbols-based data, or training data and statistical channel information about channel, and knowledge of large-scale fading coefficients, etc. They all assume that the interfering users from neighboring cells are always at distinctly lower power levels at the BS compared to the in-cell users. This requires (perfect) power control particularly if the interfering users are at the cell edge. In this paper we consider a subspace-based semi-blind approach where we have training data as well as information bearing data from various users (both in-cell and neighboring cells) at the BS. We augment the approach of [11], [12] by additional features. We allow the interfering users from neighboring cells to be be at higher power levels at the BS compared to the in-cell users, unlike [11], [12]. The same advantages also hold compared to the approach of [13]. Unlike existing approaches, the BS estimates the channels of all users: in-cell and significant neighboring cell users, i.e., ones with comparable power levels at the BS.

The rest of the paper is organized as follows. In Sec. II, we present our multi-user multi-cell system model, together with the EVD of the correlation matrices of the received signals in both training and data phases. In Sec. III, we present our proposed approach based on both pilots and data. Simulation results are presented in Sec. IV.

Notation: Superscripts $(.)^*$, $(.)^\top$ and $(.)^H$ represent complex conjugate, transpose and complex conjugate transpose (Hermitian) operation, respectively, on a vector/matrix. The notation $\mathbb{E}\{.\}$ denotes the expectation operation, $\mathbb{C}$ the set of complex numbers, $\mathbf{I}_M$ an $M \times M$ identity matrix, $\mathbf{1}_{\{A\}}$ is the indicator function, and $\delta_{i,j}$ denotes the Kronecker delta, i.e., $\delta_{i,j} = 1$ if $i = j$, $= 0$ otherwise. The notation $\mathbf{x} \sim \mathcal{N}_c(\mathbf{m}, \Sigma)$ denotes a random vector $\mathbf{x}$ that is circularly symmetric complex Gaussian with mean $\mathbf{m}$ and covariance Σ. The abbreviation w.p.1. stands for with probability one.

II. System Model

Consider a cellular wireless network composed of L cells with $K_\ell \leq \bar{K}$ single-antenna users in the ℓth cell, and one base station (BS) with N_r antennas. The system operates in a TDD mode. We focus on the uplink (UL) transmission phase. Let the $\ell = 1$ index the reference cell, with $\ell = 2, \cdots, L$ indexing the nearest neighbor cochannel cells. Consider a flat Rayleigh fading environment with the channel from the ith user in the ℓth cell to the reference-cell BS denoted as $\mathbf{h}_{\ell i} \in \mathbb{C}^{N_r}$, where $\mathbf{h}_{\ell i} \sim \mathcal{N}_c(0, \mathbf{I}_{N_r})$ represents small-scale fading. Let $p_{\ell i}$ denote the average transmitted power as well as the effects of large-scale fading, for the transmission of the ith user in the ℓth cell to the reference-cell BS. Then the received signal at reference-cell BS is given by

$$\mathbf{y}(n) = \sum_{\ell=1}^{L} \sum_{i_\ell=1}^{K_\ell} \sqrt{p_{\ell i_\ell}}\, \mathbf{h}_{\ell i_\ell} x_{\ell i_\ell}(n) + \mathbf{v}(n) \tag{1}$$

$$= \sum_{i=1}^{K_1} \sqrt{p_{1i}}\, \mathbf{h}_{1i} x_{1i}(n) + \underbrace{\sum_{\ell=2}^{L} \sum_{i_\ell=1}^{K_\ell} \sqrt{p_{\ell i_\ell}}\, \mathbf{h}_{\ell i_\ell} x_{\ell i_\ell}(n)}_{\text{inter-cell interference}} + \mathbf{v}(n) \tag{2}$$

where additive noise $\mathbf{v}(n) \sim \mathcal{N}_c(0, \sigma_v^2 \mathbf{I}_{N_r})$ and $x_{\ell i_\ell}(n)$ denotes the nth symbol transmitted by the i_ℓth user in the ℓth cell.

A. Training Phase

During the training phase, active users send training sequences as $x_{\ell i_\ell}(n)$. Suppose there are K_0 orthogonal training sequences $s_{ti}(n)$ of length P symbols, $i = 1, 2, \cdots, K_0$, $P \geq K_0$. In general, $K_0 \geq K_\ell$ for $\ell = 1, 2, \cdots, L$ but $K_0 \ll L\bar{K}$. The training sequences are assumed to be normalized to satisfy

$$P^{-1} \sum_{n=1}^{P} s_{ti}(n) s_{tj}^*(n) = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3}$$

All active users are assigned training sequences from the set of K_0 pilots by their respective BSs, which typically would lead to pilot reuse from cell-to-cell, but in a given cell, pilots are distinct and orthogonal. Suppose that the pilots are indexed (labeled) such that during the training phase, w.r.t. the reference-cell BS's choice of training sequences, we have

$$x_{1i}(n) = s_{ti}(n), \quad i = 1, 2, \cdots, K_1, \; n = 1, 2, \cdots, P. \tag{4}$$

Then, for $n = 1, 2, \cdots, P$, the received signal at reference-cell BS is given by

$$\mathbf{y}(n) = \sum_{i=1}^{K_1} \left(\sqrt{p_{1i}}\, \mathbf{h}_{1i} + \tilde{\mathbf{h}}_{1i} \right) s_{ti}(n) + \sum_{i=K_1+1}^{K_0} \tilde{\mathbf{h}}_{1i} s_{ti}(n) + \mathbf{v}(n) \tag{5}$$

where

$$\tilde{\mathbf{h}}_{1i} = \sum_{\ell=2}^{L} \sum_{i_\ell=1}^{K_\ell} \sqrt{p_{\ell i_\ell}}\, \mathbf{h}_{\ell i_\ell} \mathbf{1}_{\{x_{\ell i_\ell}(n) = s_{ti}(n),\; n=1,2,\cdots,P\}}. \tag{6}$$

Since a given pilot is assigned to no more than one user in a given cell, in (6), there are at most $L-1$ nonzero entries.

If there is no pilot reuse, then $\tilde{\mathbf{h}}_{1i} = 0$ for $i = 1, 2, \cdots, K_1$, and therefore, the BS would estimate $\sqrt{p_{1i}}\, \mathbf{h}_{1i}$ as the active in-cell ith user's channel using training $s_{ti}(n)$. In the case of reused pilots, based on (5), the BS would estimate $\sqrt{p_{1i}}\, \mathbf{h}_{1i} + \tilde{\mathbf{h}}_{1i}$ as the active in-cell ith user's channel.

B. Data Phase

During the data phase in uplink, active users transmit their information symbols as $x_{\ell i_\ell}(n)$. Using $x_{\ell i_\ell}(n)$ to denote these information sequences, the received signal at reference-cell BS is given by (2). These information sequences are assumed to be zero-mean i.i.d., mutually independent, and of known alphabet. We assume that $\mathbb{E}\left\{|x_{\ell i_\ell}(n)|^2\right\} = 1 \; \forall \ell, \, i_\ell$, with any non-unity constant absorbed in $\sqrt{p_{\ell i_\ell}}$. We assume that model (5) applies for $n = 1, 2, \cdots, P$ and model (2) applies for $n = P+1, P+2, \cdots, P+T_d$, with total $T = P+T_d$ available measurements. The BS knows K_1 and the pilot sequences of the in-cell active users, but does not know the number of reused pilots, and the data sequences of the various users.

C. Correlation Matrix

Define the correlation matrices of measurements in training and data phases as

$$\mathbf{R}_{yt} = P^{-1}\sum_{n=1}^{P}\mathbb{E}\left\{\mathbf{y}(n)\mathbf{y}^H(n)\right\}, \tag{7}$$

$$\mathbf{R}_{yd} = T_d^{-1}\sum_{n=1+P}^{T}\mathbb{E}\left\{\mathbf{y}(n)\mathbf{y}^H(n)\right\}, \tag{8}$$

and the correlation matrices of users' signals as

$$\mathbf{R}_{st} = P^{-1}\sum_{n=1}^{P}\mathbb{E}\left\{[\mathbf{y}(n)-\mathbf{v}(n)][\mathbf{y}(n)-\mathbf{v}(n)]^H\right\}, \tag{9}$$

$$\mathbf{R}_{sd} = T_d^{-1}\sum_{n=1+P}^{T}\mathbb{E}\left\{[\mathbf{y}(n)-\mathbf{v}(n)][\mathbf{y}(n)-\mathbf{v}(n)]^H\right\}. \tag{10}$$

Then we have

$$\mathbf{R}_{yt} = \mathbf{R}_{st} + \sigma_v^2\mathbf{I}_{N_r}, \quad \mathbf{R}_{yd} = \mathbf{R}_{sd} + \sigma_v^2\mathbf{I}_{N_r}. \tag{11}$$

It follows from (3) and (5) that

$$\mathbf{R}_{st} = \sum_{i=1}^{K_1}\left(\sqrt{p_{1i}}\,\mathbf{h}_{1i}+\tilde{\mathbf{h}}_{1i}\right)\left(\sqrt{p_{1i}}\,\mathbf{h}_{1i}+\tilde{\mathbf{h}}_{1i}\right)^H + \sum_{i=K_1+1}^{K_0}\tilde{\mathbf{h}}_{1i}\tilde{\mathbf{h}}_{1i}^H. \tag{12}$$

Using (2), and independence of information sequences from different users, we have

$$\mathbf{R}_{sd} = \sum_{\ell=1}^{L}\sum_{i_\ell=1}^{K_\ell} p_{\ell i_\ell}\,\mathbf{h}_{\ell i_\ell}\mathbf{h}_{\ell i_\ell}^H. \tag{13}$$

By the asymptotic orthogonality of distinct channels in a massive MIMO system [1, (10)], we have

$$\lim_{N_r\to\infty} N_r^{-1}\mathbf{h}_{\ell_1 i_{\ell_1}}^H\mathbf{h}_{\ell_2 i_{\ell_2}} = \delta_{\ell_1,\ell_2}\delta_{i_{\ell_1},i_{\ell_2}} \quad \text{w.p.1.} \tag{14}$$

Using (6) and (14) we have

$$\lim_{N_r\to\infty}\frac{1}{N_r}\mathbf{h}_{Ci_1}^H\mathbf{h}_{Ci_2} = 0 \;\; \text{w.p.1.} \;\; \text{for } i_1 \neq i_2, \tag{15}$$

where

$$\mathbf{h}_{Ci} = \begin{cases}\sqrt{p_{1i}}\,\mathbf{h}_{1i}+\tilde{\mathbf{h}}_{1i}, & 1 \le i \le K_1\\ \tilde{\mathbf{h}}_{1i}, & K_1+1 \le i \le K_0\end{cases} \tag{16}$$

By (12), (15) and (16), it follows that for large N_r, $\mathbf{R}_{st}\mathbf{h}_{Cj} = (\sum_{i=1}^{K_0}\mathbf{h}_{Ci}\mathbf{h}_{Ci}^H)\mathbf{h}_{Cj} \approx \|\mathbf{h}_{Cj}\|^2\mathbf{h}_{Cj}$. That is, for large N_r, $\mathbf{h}_{Cj}$ is an eigenvector of $\mathbf{R}_{st}$ corresponding to eigenvalue $\|\mathbf{h}_{Cj}\|^2$, for $j = 1, 2, \cdots, K_0$. Therefore, by (15), for large N_r, the vectors $\mathbf{h}_{Ci}/\|\mathbf{h}_{Ci}\|$, $i = 1, 2, \cdots, K_0$, are a set of K_0 orthonormal eigenvectors of $\mathbf{R}_{st}$. Since $(\mathbf{R}_{st}+\sigma_v^2\mathbf{I}_{N_r})\mathbf{h}_{Cj} = \mathbf{R}_{st}\mathbf{h}_{Cj} + \sigma_v^2\mathbf{h}_{Cj} \approx (\|\mathbf{h}_{Cj}\|^2+\sigma_v^2)\mathbf{h}_{Cj}$, by (11), the vectors $\mathbf{h}_{Ci}/\|\mathbf{h}_{Ci}\|$, $i = 1, 2, \cdots, K_0$, are also orthonormal eigenvectors of $\mathbf{R}_{yt}$ corresponding to its largest K_0 eigenvalues $\|\mathbf{h}_{Ci}\|^2 + \sigma_v^2$. By similar arguments, for large N_r, the vectors $\mathbf{h}_{\ell i_\ell}/\|\mathbf{h}_{\ell i_\ell}\|$, $\ell = 1, 2, \cdots, L$, $i_\ell = 1, 2, \cdots, K_\ell$, are a set of $\sum_{\ell=1}^{L} K_\ell$ orthonormal eigenvectors of $\mathbf{R}_{sd}$, and by (11), they are also orthonormal eigenvectors of $\mathbf{R}_{yd}$ corresponding to its largest $\sum_{\ell=1}^{L} K_\ell$ eigenvalues $p_{\ell i_\ell}\|\mathbf{h}_{\ell i_\ell}\|^2 + \sigma_v^2$.

III. Reused Pilot Detection and Channel Estimation

We now present our proposed approach based on both pilots and information sequences (data). In Sec. III-A we consider pilot-based least-squares channel estimation that includes the effects of pilot contamination. In Sec. III-B, we present a matrix rank determination approach applied to the data correlation matrix to estimate the number of "significant" users in the system, which include in-cell users and interfering users with received power comparable to the weakest in-cell user. In Sec. III-C, using the eigenvectors (based on data correlation matrix) of significant users, we project the data onto the signal subspace, and then use higher-order statistics of the projected data to resolve a unitary matrix ambiguity in multiuser channel estimation for all significant users. Pilot-based results are then used in Sec. III-C1 to identify reused pilots. In Sec. III-D, we review a version of the approach of [11], [12] for the case when no interfering users are detected, and this approach is also used later in Sec. IV as representative of [11], [12] in carrying out performance comparisons with our proposed approach. In Sec. III-E, the estimated channels are used to detect information symbols, which, in turn, are used as pseudo-pilots to re-estimate the in-cell users' channels. Our complete solution is summarized as Algorithm 1 in Sec. III-F.

A. Pilot Based Channel Estimation in Training Phase

Here we use pilot-based least-squares procedure using training-phase measurements to estimate K_1 channels associated with K_1 pilots assigned to K_1 users in the reference cell. These channels have the (ill-)effect of pilot contamination. Using the least-squares approach, orthogonality of training, (5), and (16), the channel corresponding to the ith pilot for $i = 1, 2, \cdots, K_1$, is estimated as

$$\hat{\mathbf{h}}_{Ci} = P^{-1}\sum_{n=1}^{P}\mathbf{y}(n)s_{ti}^*(n). \tag{17}$$

It is easy to see that $\mathbb{E}\{\hat{\mathbf{h}}_{Ci}\} = \mathbf{h}_{Ci}$, $i = 1, 2, \cdots, K_1$, which shows that the channel estimate is biased for reused pilots. Define the contaminated-channel matrix

$$\mathbf{H}^{(p)} = [\mathbf{h}_{C1} \; \cdots \; \mathbf{h}_{CK_1}], \quad \hat{\mathbf{H}}^{(p)} = [\hat{\mathbf{h}}_{C1} \; \cdots \; \hat{\mathbf{h}}_{CK_1}]. \tag{18}$$

Using (14) and (15), it follows that for large N_r, taking expectation w.r.t. noise only,

$$\begin{aligned}\mathbb{E}&\left\{\|\hat{\mathbf{h}}_{Ci}\|^2\right\}\\ &= p_{1i}\|\mathbf{h}_{1i}\|^2 + \|\tilde{\mathbf{h}}_{1i}\|^2 + \sqrt{p_{1i}}\,\mathbf{h}_{1i}^H\tilde{\mathbf{h}}_{1i} + \sigma_v^2\, N_r/P \qquad (19)\\ &\approx p_{1i}\|\mathbf{h}_{1i}\|^2 + \sum_{\ell=2}^{L}\sum_{i_\ell=1}^{K_\ell} p_{\ell i_\ell}\,\|\mathbf{h}_{\ell i_\ell}\|^2\mathbf{1}_{\{x_{\ell i_\ell}(n)=s_{ti}(n),\, n=1,2,\cdots,P\}}\\ &\quad + \sigma_v^2 N_r/P. \qquad (20)\end{aligned}$$

For large N_r, invoking ergodicity,

$$\mathbb{E}\left\{\|\hat{\mathbf{h}}_{Ci}\|^2\right\} \approx \|\hat{\mathbf{h}}_{Ci}\|^2. \tag{21}$$

B. Matrix Rank Determination: How Many "Significant" Users in Data Phase?

Define the sample correlation matrices under training and data phases as

$$\hat{\mathbf{R}}_{yt} = P^{-1} \sum_{n=1}^{P} \mathbf{y}(n)\mathbf{y}^H(n), \tag{22}$$

$$\hat{\mathbf{R}}_{yd} = T_d^{-1} \sum_{n=1+P}^{T} \mathbf{y}(n)\mathbf{y}^H(n). \tag{23}$$

Let the ordered eigenvalues of $\hat{\mathbf{R}}_{yt}$ be denoted by $\ell_{t1} \geq \ell_{t2} \geq \cdots \geq \ell_{tN_r}$ in decreasing order of magnitude, and that of $\hat{\mathbf{R}}_{yd}$ be denoted by $\ell_{d1} \geq \ell_{d2} \geq \cdots \geq \ell_{dN_r}$.

First we wish to determine the number of significant user signals in the reference cell, given the measurements at the reference cell BS during both training and data phases. The BS knows K_1 and K_0, and therefore knows that the signal subspace rank of $\hat{\mathbf{R}}_{yt}$ is at least K_1, and no more than K_0. That is, the first K_0 eigenvalues of ordered eigenvalues $\ell_{t1} \geq \ell_{t2} \geq \cdots \geq \ell_{tN_r}$ are possibly the signal-plus-noise eigenvalues, whereas the remaining $N_r - K_0$ eigenvalues originate from σ_v^2. An estimate of σ_v^2 is, therefore, given by

$$\widehat{\sigma_v^2} = \frac{1}{N_r - K_0} \sum_{i=1+K_0}^{N_r} \ell_{ti}. \tag{24}$$

By (20), (21) and (24), we have

$$p_{1i}\|\mathbf{h}_{1i}\|^2 + \sum_{\ell=2}^{L}\sum_{i_\ell=1}^{K_\ell} p_{\ell i_\ell}\, \|\mathbf{h}_{\ell i_\ell}\|^2 \mathbf{1}_{\{x_{\ell i_\ell}(n)=s_{ti}(n),\; n=1,2,\cdots,P\}} \approx \|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P. \tag{25}$$

Now consider the eigenvalues of data correlation matrix $\hat{\mathbf{R}}_{yd}$. As discussed in Sec. II-C, the eigenvalues of $\mathbf{R}_{sd}$ corresponding to the reference cell users are $p_{1i}\,\|\mathbf{h}_{1i}\|^2$, $i = 1, 2, \cdots, K_1$. In the absence of perfect power control, signals from the reference cell users are not necessarily the strongest K_1 signals at the BS of the reference cell. If received power of signals from interfering users is not higher, on the average, than that from in-cell users, the left-side of (25) is approximately less than $Lp_{1i}\|\mathbf{h}_{1i}\|^2$, so that $p_{1i}\|\mathbf{h}_{1i}\|^2 \geq (1/L)[\|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P]$, $i = 1, 2, \cdots, K_1$. For some $0 < \mu \leq 1$, let

$$\alpha_1 = \min_{1\leq i\leq K_1} (\mu/L)[\|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P]. \tag{26}$$

This discussion implies that when $\mu = 1$, the eigenvalues ℓ_{di} of the data correlation matrix corresponding to in-cell users will likely exceed $\alpha_1 + \widehat{\sigma_v^2}$, since the largest $\sum_{\ell=1}^{L} K_\ell$ eigenvalues ℓ_{di} of $\mathbf{R}_{yd}$ are of the form $p_{\ell i_\ell}\,\|\mathbf{h}_{\ell i_\ell}\|^2 + \sigma_v^2$. Since this discussion is based on heuristic, approximate considerations, in order to ensure that all eigenvalues ℓ_{di} of $\hat{\mathbf{R}}_{yd}$ corresponding to in-cell users, do indeed exceed $\alpha_1 + \widehat{\sigma_v^2}$, we introduce the factor $\mu \in (0, 1]$ in (26). As discussed later, α_1 is used as a threshold on eigenvalues ℓ_{di} to determine a subspace of reduced rank which includes all in-cell users. If $\mu = 0$, then the subspace is of full rank that includes all users in the multi-cell system. If $\mu = 1$, one might occasionally exclude in-cell users from the selected reduced subspace. In simulations presented in Sec. IV, with $L = 7$, $\mu = 0.7$ was picked, leading to $\mu/L = 0.1$. We also show later via simulations in Fig. 9 where we picked μ =0.3, 0.7 or 1.0 for $L = 7$, that the results are not sensitive to the choice. Alternatively, suppose that BS knows that the SNR for any in-cell user at the BS is at least α_2. Then, since the SNR of the ith in-cell user equals $p_{1i}\|\mathbf{h}_{1i}\|^2/(N_r\sigma_v^2)$, the eigenvalues of $\mathbf{R}_{yd}$ corresponding to the in-cell users exceed $(\alpha_2\ N_r + 1)\widehat{\sigma_v^2}$.

The signal subspace of $\hat{\mathbf{R}}_{yd}$ is of rank $\sum_\ell^L K_\ell$. We need to pick a subspace of reduced rank from the signal subspace of $\hat{\mathbf{R}}_{yd}$ which includes all in-cell users, and additionally, interfering users whose received power is comparable to the weakest in-cell user. Consider a threshold τ for the ordered eigenvalues of $\hat{\mathbf{R}}_{yd}$, given by

$$\tau = \max\Big(\alpha_1,\ (\alpha_2 N_r + 1)\widehat{\sigma_v^2}\Big). \tag{27}$$

Then all $\ell_{di} \geq \tau$ are deemed to arise from signal subspace of the data-phase correlation matrix such that this reduced subspace includes all in-cell users as well as interfering users having power comparable to the weakest in-cell user. Recall that the eigenvalues $\{\ell_{di}\}_i$ of $\hat{\mathbf{R}}_{yd}$ are arranged in decreasing order. Our choice of the threshold τ is heuristic but reasonable. A principled approach exploiting source enumeration techniques such as minimum description length (MDL) can be employed but it also picks up the presence of very weak signals, well below thermal noise [17].

Let $\hat{K}_d$ denote the number of eigenvalues of $\hat{\mathbf{R}}_{yd}$ that exceed τ. (Note that, by the modeling assumptions, $\hat{K}_d$ cannot be less than K_1.) Then the significant number of extraneous (interfering) users are estimated as $K_r = \hat{K}_d - K_1$. Not all of these extraneous users necessarily have reused pilots if $K_1 < K_0$.

C. Blind Channel Estimation in Data Phase When $K_r > 0$

Here we only use data-phase measurements to estimate $\hat{K}_d = K_1 + K_r$ channels using both second and higher-order statistics, in two steps. First we rewrite (2) as

$$\mathbf{y}(n) = \sum_{i=1}^{K_1} \sqrt{p_{1i}}\,\mathbf{h}_{1i}x_{1i}(n) + \sum_{j=1}^{K_r} \sqrt{p_{rj}}\,\mathbf{h}_{rj}x_{rj}(n) + \tilde{\mathbf{v}}(n) \tag{28}$$

where p_{rj}, $\mathbf{h}_{rj}$, and x_{rj} are re-indexed entries from the sets ($\ell \geq 2$), $\{p_{\ell i_\ell}\}$, $\{\mathbf{h}_{\ell i_\ell}\}$, and $\{x_{\ell i_\ell}(n)\}$, respectively, that correspond to the extraneous users estimated in Sec. III-B on the basis of the eigenvalues of $\hat{\mathbf{R}}_{yd}$, and $\tilde{\mathbf{v}}(n)$ is the sum of $\mathbf{v}(n)$ and the remaining sources not included in the first two sums on the right-side of (28). Consider EVD of $\hat{\mathbf{R}}_{yd}$ to obtain

$$\hat{\mathbf{R}}_{yd} = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{U}}^H = [\hat{\mathbf{U}}_1\ \hat{\mathbf{U}}_2]\begin{bmatrix} \hat{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_2 \end{bmatrix}[\hat{\mathbf{U}}_1\ \hat{\mathbf{U}}_2]^H \tag{29}$$

where $\hat{\Sigma}$ is a $N_r \times N_r$ diagonal matrix with eigenvalues $\{\ell_{di}\}_i$ arranged in decreasing order of magnitude, columns of $\hat{\mathbf{U}}$ are the corresponding eigenvectors, and $\hat{\mathbf{U}}_1$ is $N_r \times (K_1 + K_r)$. Thus, $\hat{\mathbf{U}}_1$ determines the reduced signal subspace and $\hat{\mathbf{U}}_2$ determines the modified noise subspace (corresponding to $\tilde{\mathbf{v}}(n)$) of the estimated correlation matrix.

With reference to (28), define a channel matrix $\mathbf{H}_d \in \mathbb{C}^{N_r \times (K_1+K_r)}$ as

$$\mathbf{H}_d = [\sqrt{p_{11}}\mathbf{h}_{11} \cdots \sqrt{p_{1K_1}}\mathbf{h}_{1K_1} \sqrt{p_{r1}}\mathbf{h}_{r1} \cdots \sqrt{p_{rK_r}}\mathbf{h}_{rK_r}]. \tag{30}$$

Then we can rewrite (28) as

$$\mathbf{y}(n) = \mathbf{H}_d\mathbf{x}(n) + \tilde{\mathbf{v}}(n), \tag{31}$$

$$\mathbf{x}(n) = [x_{11}(n) \cdots x_{1K_1}(n)\ x_{r1}(n) \cdots x_{rK_r}(n)]^\top. \tag{32}$$

Since the data sequences $x_{1i}(n)$ and $x_{rj}(n)$ are zero-mean, unit variance, mutually independent and i.i.d., in the notation of (10)-(11), we have

$$\mathbf{R}_{yd} = \mathbf{U}\Sigma\mathbf{U}^H = [\mathbf{U}_1\ \mathbf{U}_2]\begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix}[\mathbf{U}_1\ \mathbf{U}_2]^H \tag{33}$$

where $\mathbf{U}$, Σ, etc. in (33) are the true counterparts of the estimated $\hat{\mathbf{U}}$, $\hat{\Sigma}$, etc. in (29).

The channels $\mathbf{h}_{1i}$ and $\mathbf{h}_{rj}$ lie in the subspace spanned by the columns of $\mathbf{U}_1$. Consider, for $n = P+1, \cdots, P+T_d = T$,

$$\tilde{\mathbf{y}}(n) = \hat{\mathbf{U}}_1^H\mathbf{y}(n) \in \mathbb{C}^{K_1+K_r}, \tag{34}$$

and its true counterpart, obtained after replacing $\hat{\mathbf{U}}_1$ with $\mathbf{U}_1$,

$$\check{\mathbf{y}}(n) = \mathbf{U}_1^H\mathbf{y}(n) \in \mathbb{C}^{K_1+K_r}. \tag{35}$$

Then we have

$$\check{\mathbf{y}}(n) = \mathbf{U}_1^H(\mathbf{H}_d\mathbf{x}(n) + \mathbf{v}(n)) = \tilde{\mathbf{H}}_d\mathbf{x}(n) + \check{\mathbf{v}}(n) \tag{36}$$

where $\tilde{\mathbf{H}}_d \in \mathbb{C}^{(K_1+K_r)\times(K_1+K_r)}$, $\tilde{\mathbf{H}}_d = \mathbf{U}_1^H\mathbf{H}_d$, and $\check{\mathbf{v}}(n) \in \mathbb{C}^{K_1+K_r}$. For large N_r, by orthogonality of distinct channels from distinct users (see (14)), we have $\mathbb{E}\{\check{\mathbf{v}}(n)\check{\mathbf{v}}^H(n)\} \approx \sigma_v^2\mathbf{I}_{K_1+K_r}$ since $\mathbf{U}_1^H\mathbf{U} = \mathbf{I}_{K_1+K_r}$, and we have neglected contributions from the source terms not included in the first two sums on the right-side of (28) by appealing to (14). In practice, we approximate $\check{\mathbf{y}}(n)$ with $\tilde{\mathbf{y}}(n)$.

Since data sequences are independent non-Gaussian, one can apply higher-order statistics-based approaches to estimate $\tilde{\mathbf{H}}_d$ [18]–[21]. We will use the RobustICA algorithm of [20] that uses kurtosis of "unmixed" measurements. It provides an estimate $\hat{\tilde{\mathbf{H}}}_d$ of $\tilde{\mathbf{H}}_d$ using $\tilde{\mathbf{y}}(n)$; a modified version of this algorithm has been used in [13], but we do not modify it in any way. One obtains, for some $\theta_i s$,

$$\hat{\tilde{\mathbf{H}}}_d \approx \tilde{\mathbf{H}}_d\mathcal{P}\Gamma_\theta, \quad \Gamma_\theta = \mathrm{diag}\{e^{j\theta_i},\ i = 1, 2, \cdots, K_1+K_r\} \tag{37}$$

where $\mathcal{P}$ is a permutation matrix – the order of "extracted" sources, hence, the order of extracted columns of $\tilde{\mathbf{H}}_d$ cannot be determined by RobustICA (indeed, by any blind source separation method for instantaneous mixtures [21]), and one can only recover channels up to a constant of modulus one when using kurtosis and related criteria for unmixing.

Remark 1: Consider the (restricted) independent component analysis (ICA) problem

$$\mathbf{z}(n) = \mathbf{A}\tilde{\mathbf{s}}(n) \tag{38}$$

where $\mathbf{z}(n), \tilde{\mathbf{s}}(n) \in \mathbb{C}^p$, $\mathbf{A} \in \mathbb{C}^{p\times p}$, the sequence $\{\tilde{\mathbf{s}}(n)\}$ is zero-mean, i.i.d., non-Gaussian (finite alphabet), and the objective is to recover $\tilde{\mathbf{s}}(n)$ and estimate $\mathbf{A}$. Such problems have been addressed in [18]–[21], among others. We will consider only square $\mathbf{A}$, hence the term restricted ICA; this is sufficient for our purposes. Ignoring noise $\check{\mathbf{v}}(n)$ in (36), we see that (36) corresponds to (38) with $K_1 + K_r = p$, $\tilde{\mathbf{H}}_d = \mathbf{A}$, $\mathbf{x}(n) = \tilde{\mathbf{s}}(n)$ and $\check{\mathbf{y}}(n) = \mathbf{z}(n)$. For some $\mathbf{w} \in \mathbb{C}^p$, let $e(n) = \mathbf{w}^H\mathbf{z}(n)$. In the approach of [20], $\mathbf{w}$ is picked to maximize $|\gamma_4|$, where the kurtosis (normalized 4th cumulant) γ_4 of $e(n)$ is given by

$$\gamma_4 = \frac{\mathbb{E}\{|e(n)|^4\} - 2\mathbb{E}\{|e(n)|^2\} - \mathbb{E}\{e^2(n)\}}{\left(\mathbb{E}\{|e(n)|^2\}\right)^2}. \tag{39}$$

When $|\gamma_4|$ is maximized for $\mathbf{w} = \bar{\mathbf{w}}$, one has $e(n) = \bar{\mathbf{w}}^H\mathbf{z}(n) = \tilde{s}_m(n)$ for some $1 \le m \le p$, where $\tilde{s}_m(n)$ is the mth component of $\tilde{\mathbf{s}}(n)$. Thus, one can obtain $\tilde{s}_m(n)$, and using $\tilde{s}_m(n)$, (38) and least-squares, estimate $A_{\ell m}$ for $1 \le \ell \le p$, where $A_{\ell m}$ is (ℓ, m)th element of $\mathbf{A}$. After estimating the mth column of $\mathbf{A}$, contribution of $\tilde{s}_m(n)$ to $\mathbf{z}(n)$ is subtracted (deflated) from $\mathbf{z}(n)$, and the entire process is repeated till all sources (components of $\tilde{\mathbf{s}}(n)$) are extracted, and $\mathbf{A}$ is estimated. Such a procedure for more general class of systems may also be found in [18]. The RobustICA algorithm of [20] provides an optimal step-size in the iterative maximization of $|\gamma_4|$ via a gradient-descent method, leading to fast convergence. Note that this method can also be applied to (31), but this would require one to process a much larger dimensional $\tilde{\mathbf{H}}_d$ in (31), compared to smaller $\tilde{\mathbf{H}}_d$ in (36). More importantly, projecting as in (34), further suppresses weaker interfering signals by virtue of (14). □

Thus, an estimate of $\mathbf{H}_d = \mathbf{U}_1\tilde{\mathbf{H}}_d$ is given by

$$\hat{\mathbf{H}}_d = \hat{\mathbf{U}}_1\hat{\tilde{\mathbf{H}}}_d \approx \mathbf{H}_d\mathcal{P}\Gamma_\theta. \tag{40}$$

1) Using Pilot-Based Channel Estimates to Identify Reused Pilots and Interfering Users: Consider $\mathbf{H}^{(p)} \in \mathbb{C}^{N_r\times K_1}$ defined in (18), and $\mathbf{H}_d\mathcal{P} \in \mathbb{C}^{N_r\times(K_1+K_r)}$ where $\mathcal{P} \in \mathbb{C}^{(K_1+K_r)\times(K_1+K_r)}$ is a permutation matrix, and Γ_θ is as in (37). The pilot-based channel estimates (17) yield $\hat{\mathbf{H}}^{(p)}$ while (40) yields $\hat{\mathbf{H}}_d$. Observe that if the ith pilot is not reused, then the ith column of $\mathbf{H}^{(p)}$ equals a scaled version of some column of $\mathbf{H}_d\mathcal{P}\Gamma_\theta$. If the ith pilot is reused, then the ith column of $\mathbf{H}^{(p)}$ equals a weighted sum of two or more columns of $\mathbf{H}_d\mathcal{P}\Gamma_\theta$. Therefore, there exists a matrix $\mathbf{G} \in \mathbb{C}^{(K_1+K_r)\times K_1}$ such that

$$(\mathbf{H}_d\mathcal{P}\Gamma_\theta)\,\mathbf{G} = \mathbf{H}^{(p)} \Rightarrow \hat{\mathbf{H}}_d\mathbf{G} \approx \hat{\mathbf{H}}^{(p)}. \tag{41}$$

Hence an estimate of $\mathbf{G}$ is given by

$$\hat{\mathbf{G}} = \left(\hat{\mathbf{H}}_d^H\hat{\mathbf{H}}_d\right)^{-1}\hat{\mathbf{H}}_d^H\hat{\mathbf{H}}^{(p)}. \tag{42}$$

The number of nonzero entries in kth column of $\mathbf{G}$ signify that the kth column of $\hat{\mathbf{H}}^{(p)}$ is a weighted sum of the columns of $\hat{\mathbf{H}}_d$ that correspond to the rows of the kth column of $\mathbf{G}$ with nonzero entries. Suppose that the third column of $\mathbf{G}$ has one nonzero entry (in the fourth row). This means that the third column $\hat{\mathbf{H}}^{(p)}$ equals a scaled version of the fourth column of $\hat{\mathbf{H}}_d$ and there is no pilot reuse. On the other hand, suppose that the third column of $\mathbf{G}$ has two nonzero entries (in the second and fourth rows). This means that the third column $\hat{\mathbf{H}}^{(p)}$ equals a weighted sum of the second and fourth columns of $\hat{\mathbf{H}}_d$, and there is pilot reuse with the third pilot being used by two users. Suppose that some column of $\hat{\mathbf{H}}_d$ corresponds to an interfering user that does not reuse any pilot in the reference cell. Then the row of $\mathbf{G}$ corresponding to this out-of-cell user with non-reused pilot, will have all zero entries.

In practice, we only have noisy $\hat{\mathbf{G}}$. In order for $\hat{\mathbf{G}}$ to "represent" ideal $\mathbf{G}$, we adopt the following procedure to replace $\hat{\mathbf{G}}$ with $\bar{\mathbf{G}}$, which then plays the role of the ideal $\mathbf{G}$.

1) Replace the ith column $\hat{\mathbf{G}}_i$ of $\hat{\mathbf{G}}$ with $[|\hat{\mathbf{G}}_{i1}| \cdots |\hat{\mathbf{G}}_{i(K_1+K_r)}|]^\top / \|\hat{\mathbf{G}}_i\|$, i.e., each column is first normalized to unit norm, and then each normalized entry is replaced with its absolute value. Denote the resulting matrix by $\bar{\mathbf{G}}$.
2) If $\bar{\mathbf{G}}_{ij} < \tau_1$, set $\bar{\mathbf{G}}_{ij} = 0$, where in our simulations we set $\tau_1 = 0.15$. Since the BS knows the number K_1 of active reference cell users, one expects at least K_1 nonzero entries in thresholded $\bar{\mathbf{G}}_{ij}$; if this number is less than K_1, we lower τ_1. Otherwise, $\tau_1 > 0$ is picked to ignore weak dependence of columns of $\hat{\mathbf{H}}^{(p)}$ on columns of $\hat{\mathbf{H}}_d$, and in simulations we used $\tau_1 = 0.15$
3) *Channel Resolution*: Consider the ith column $\hat{\mathbf{H}}_i^{(p)}$ of $\hat{\mathbf{H}}^{(p)}$, $i = 1, 2, \cdots, K_1$.
   a) If the ith column $\bar{\mathbf{G}}_i$ of $\bar{\mathbf{G}}$ has only one nonzero element in its jth row, then we pick
   $$\hat{\mathbf{h}}_{Ci} = \hat{\mathbf{H}}_i^{(p)} = P^{-1} \sum_{n=1}^{P} \mathbf{y}(n) s_{ti}^*(n).$$
   That is, the ith pilot $s_{ti}(n)$ is not reused and $\hat{\mathbf{h}}_{Ci}$ is the least-squares estimate of the ith user's channel $\mathbf{h}_{Ci} = \sqrt{p_{1i}}\,\mathbf{h}_{1i}$ based on training data. We relabel it as $\hat{\mathbf{h}}_{1i}$.
   b) Suppose the ith column $\bar{\mathbf{G}}_i$ of $\bar{\mathbf{G}}$ has $q > 1$ nonzero elements in rows j_ℓ, $1 \le \ell \le q$. Then we have
   $$\sum_{\ell=1}^{q} c_\ell \hat{\mathbf{H}}_{dj_\ell} \approx \hat{\mathbf{H}}_i^{(p)}$$
   where we wish to determine complex c_ℓs instead of using thresholded, scaled $\bar{\mathbf{G}}_{ij_\ell}$. Define
   $$\bar{\mathbf{H}} = [\hat{\mathbf{H}}_{dj_1} \ \cdots \ \hat{\mathbf{H}}_{dj_q}] \in \mathbb{C}^{N_r \times q}, \quad \mathbf{c} = [c_1 \ \cdots \ c_q]^\top.$$
   We estimate $\mathbf{c}$ as $\hat{\mathbf{c}} = (\bar{\mathbf{H}}^H \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}^H \hat{\mathbf{H}}_i^{(p)}$. Then we have q channels associated with the ith pilot: $\hat{c}_\ell \hat{\mathbf{H}}_{dj_\ell}$, $1 \le \ell \le q$. One of these is from a reference cell user and the remaining $q-1$ are from neighboring cells. Without any additional information we cannot determine the true origin of these q channels. We assume that the corresponding data phase measurements have some information embedded in them regarding user identification and one can extract this from decoded data, decoded using, for instance, matched filter beamforming based on the estimated channel. Denote the total number of interfering users with pilot reuse by K_{ri}, and order these estimated channels as $\hat{\mathbf{h}}_{rj}$, $j = 1, 2, \cdots, K_{ri}$.
   c) If the mth row of $\bar{\mathbf{G}}$ has all zero entries, then the mth column of $\hat{\mathbf{H}}_d$ corresponds to an interfering user that does not reuse any pilot in the reference cell. In this case, we take the mth column of $\hat{\mathbf{H}}_d$ as an estimate of $\sqrt{p_{rj}}\,\mathbf{h}_{rj}$ for some $1 \le j \le K_r$. Denote the total number of interfering users without pilot reuse by K_{ro}, and order these estimated channels as $\hat{\mathbf{h}}_{rj}$, $j = K_{ri}+1, 2, \cdots, K_{ri}+K_{ro}$, where $K_{ri}+K_{ro} = K_r$.
   d) As a result of steps a), b) and c) above, we have the channel estimates $\hat{\mathbf{h}}_{1i}$ and $\hat{\mathbf{h}}_{rj}$ of $\sqrt{p_{1i}}\,\mathbf{h}_{1i}$ and $\sqrt{p_{rj}}\,\mathbf{h}_{rj}$, respectively, for $i = 1, 2, \cdots, K_1$ and $j = 1, 2, \cdots, K_r$. The interfering user obtained in step c) above corresponds to a user that does not reuse any pilot in the reference cell, and that obtained in step b) corresponds to a user that does reuse a pilot. Since pilots are used to obtain $\hat{\mathbf{H}}^{(p)}$ (see (17) and (18)), which, in turn, is used to obtain $\hat{\mathbf{c}}$ (see step b) above), there is no scaling ambiguity (such as $e^{j\theta_i}$ in (37)) in the estimates $\hat{\mathbf{h}}_{rj}$, $j = 1, 2, \cdots, K_{ri}$, of channels of interfering users with pilot reuse. On the other hand, the estimates $\hat{\mathbf{h}}_{rj}$, $j = K_{ri}+1, \cdots, K_{ri}+K_{ro}$, of channels of interfering users without pilot reuse will have such ambiguities, as discussed earlier in the context of (37).

D. Semi-Blind Channel Estimation When $K_r = 0$

If one determines $K_r = 0$, then we proceed in a manner similar to [11], [12]. Set $K_r = 0$ in Sec. III-C and estimate $\mathbf{U}_1$ as $\hat{\mathbf{U}}_1$, as in (29)–(33). This is as in [12] except that [12] uses SVD of a data matrix instead of EVD of $\hat{\mathbf{R}}_{yd}$, as we do following [11]. Then, instead of using (34) with data phase measurements, we use training phase measurements (as in one of the options of [12]):

$$\mathbf{y}'(n) = \mathbf{U}_1^H \mathbf{y}(n) \in \mathbb{C}^{K_1}, \quad n = 1, 2, \cdots, P \tag{43}$$
$$= \tilde{\mathbf{H}}_{d1} \mathbf{s}(n) + \mathbf{v}'(n) \tag{44}$$

where $\tilde{\mathbf{H}}_{d1} = \mathbf{U}_1^H \mathbf{H}_{d1} \in \mathbb{C}^{K_1 \times K_1}$,

$$\mathbf{s}(n) = [s_{t1}(n) \ \cdots \ s_{tK_1}(n)]^\top, \tag{45}$$
$$\mathbf{H}_{d1} = [\sqrt{p_{11}}\,\mathbf{h}_{11} \ \cdots \ \sqrt{p_{1K_1}}\,\mathbf{h}_{1K_1}]. \tag{46}$$

Finally, $\tilde{\mathbf{H}}_{d1}$ is estimated as $\hat{\tilde{\mathbf{H}}}_{d1}$ using pilots, (44) and the method of least-squares, to obtain the final estimate

$$\hat{\mathbf{H}}_{d1} = \hat{\mathbf{U}}_1 \hat{\tilde{\mathbf{H}}}_{d1} \approx \mathbf{H}_{d1} \tag{47}$$

where

$$\hat{\tilde{\mathbf{H}}}_{d1} = P^{-1} \sum_{n=1}^{P} \mathbf{y}'(n) \mathbf{s}^H(n). \tag{48}$$

Since we use pilots, there is no permutation or scaling ambiguity, unlike (40). Notice that this approach (also of [11], [12]) does not acknowledge the presence of any pilot contamination, under the assumption that contaminating pilot strength is (much) weaker owing to perfect power control and distance to the reference cell BS.

E. Data Detection and Channel Re-Estimation

It is noted in [13] (see also [11], [12]) that data-aided channel estimation where one also uses detected information symbols in addition to the pilot symbols, can achieve better estimation performance than the pilot-only channel estimation. Following this observation, we now use the estimated channels, of both reference cell users $\hat{\mathbf{h}}_{1i}$ and strong interfering users from neighboring cells $\hat{\mathbf{h}}_{rj}$, to design linear minimum mean-square error (MMSE) detector to estimate the information symbols of the reference cell users.

With $\mathbf{H}_{d1}$ as in (46), define

$$\mathbf{H}_{dr} = [\sqrt{p_{r1}}\,\mathbf{h}_{r1}\ \cdots\ \sqrt{p_{rK_r}}\,\mathbf{h}_{rK_r}], \tag{49}$$

$$\hat{\mathbf{H}}_{d1} = [\hat{\mathbf{h}}_{11}\ \cdots\ \hat{\mathbf{h}}_{1K_1}], \tag{50}$$

$$\hat{\mathbf{H}}_{dr} = [\hat{\mathbf{h}}_{r1}\ \cdots\ \hat{\mathbf{h}}_{rK_r}]. \tag{51}$$

Then $\hat{\mathbf{H}}_{d1}$ and $\hat{\mathbf{H}}_{dr}$ are the estimates of $\mathbf{H}_{d1}$ and $\mathbf{H}_{dr}$, respectively. We can express (28) as

$$\mathbf{y}(n) = \mathbf{H}_{d1}\mathbf{x}_1(n) + \mathbf{H}_{dr}\mathbf{x}_r(n) + \tilde{\mathbf{v}}(n) \tag{52}$$

where

$$\mathbf{x}_1(n) = [x_{11}(n)\ \cdots\ x_{1K_1}(n)]^\top, \tag{53}$$

$$\mathbf{x}_r(n) = [x_{r1}(n)\ \cdots\ x_{rK_r}(n)]^\top, \tag{54}$$

and ignoring other (low-power) interfering users, we take the covariance matrix of $\tilde{\mathbf{v}}(n)$ to be the same as that of $\mathbf{v}(n)$, i.e., it equals $\sigma_v^2\mathbf{I}_{N_r}$. Based on (52), the linear MMSE estimator of $\mathbf{x}_1(n)$ is given by

$$\hat{\mathbf{x}}_1(n) = \mathbf{H}_e\mathbf{y}(n) \tag{55}$$

where

$$\mathbf{H}_e = \begin{cases} \mathbf{H}_{d1}^H\left[\mathbf{H}_{d1}\mathbf{H}_{d1}^H + \mathbf{H}_{dr}\mathbf{H}_{dr}^H + \sigma_v^2\mathbf{I}_{N_r}\right]^{-1} & \text{if } K_r > 0 \\ \mathbf{H}_{d1}^H\left[\mathbf{H}_{d1}\mathbf{H}_{d1}^H + \sigma_v^2\mathbf{I}_{N_r}\right]^{-1} & \text{if } K_r = 0. \end{cases} \tag{56}$$

In practice, we replace $\mathbf{H}_{d1}$, $\mathbf{H}_{dr}$, and σ_v^2 with their estimates $\hat{\mathbf{H}}_{d1}$, $\hat{\mathbf{H}}_{dr}$, and $\widehat{\sigma_v^2}$, respectively. Note that the scaling ambiguity referred to in step d) of item 3), *Channel Resolution*, of Sec. III-C1, has no influence on (56) since we would use $\hat{\mathbf{H}}_{dr}\hat{\mathbf{H}}_{dr}^H$, which is not dependent upon Γ_θ or $e^{j\theta_i}$ because $\Gamma_\theta\Gamma_\theta^H = \mathbf{I}$.

Let $\mathbf{x}_{1q}(n)$ denote the quantized $\hat{\mathbf{x}}_1(n)$, obtained by exploiting the known symbol constellation in the data phase. Define

$$\tilde{\mathbf{x}}_1(n) = \begin{cases} \mathbf{s}(n) & \text{for } n = 1, 2, \cdots, P, \\ \mathbf{x}_{1q}(n) & \text{for } n = P+1, P+2, \cdots, P+T_d. \end{cases} \tag{57}$$

Now use $\tilde{\mathbf{x}}_1(n)$ above to re-estimate the multi-user channel $\mathbf{H}_{d1}$ (see (46)), via least-squares as

$$\hat{\mathbf{H}}_{d1} = T^{-1}\sum_{n=1}^{T}\mathbf{y}(n)\tilde{\mathbf{x}}_1^H(n). \tag{58}$$

The improved estimates may be used in (56) to re-compute the linear MMSE estimator, which, in turn, is then used to recompute quantized $\mathbf{x}_{1q}(n)$.

F. Summary of the Solution

The overall approach is summarized in Algorithm 1.

We conclude with the following three remarks.

Remark 2 (Interference Cancellation): If $K_{ri} > 0$, one may also choose to performance interference cancellation (IC) at the end of step 10 in Algorithms 1. Using the notation of item 3), *Channel Resolution*, of Sec. III-C1, define

$$\mathbf{H}_{dri} = [\sqrt{p_{r1}}\,\mathbf{h}_{r1}\ \cdots\ \sqrt{p_{rK_{ri}}}\,\mathbf{h}_{rK_{ri}}], \tag{59}$$

$$\mathbf{H}_{dro} = [\sqrt{p_{r(K_{ri}+1)}}\,\mathbf{h}_{r(K_{ri}+1)}\ \cdots\ \sqrt{p_{r(K_{ri}+K_{ro})}}\,\mathbf{h}_{r(K_{ri}+K_{ro})}], \tag{60}$$

$$\hat{\mathbf{H}}_{dri} = [\hat{\mathbf{h}}_{r1}\ \cdots\ \hat{\mathbf{h}}_{rK_{ri}}], \tag{61}$$

$$\hat{\mathbf{H}}_{dr0} = [\hat{\mathbf{h}}_{r(K_{ri}+1)}\ \cdots\ \hat{\mathbf{h}}_{r(K_{ri}+K_{ro})}], \tag{62}$$

where $\hat{\mathbf{H}}_{dri}$ and $\hat{\mathbf{H}}_{dro}$ are estimates of $\mathbf{H}_{dri}$ and $\mathbf{H}_{dro}$, respectively. The channel matrix $\mathbf{H}_{dri}$ represents the channels of the interfering users reusing pilots of reference cell users, whereas $\mathbf{H}_{dro}$ represents the channels of the interfering users not reusing any pilot of reference cell users. As discussed in item 3), *Channel Resolution*, of Sec. III-C1, there is no scaling ambiguity in $\mathbf{H}_{dri}$, unlike $\mathbf{H}_{dro}$. Therefore, we can exploit $\hat{\mathbf{H}}_{dri}$ to estimate the information symbols of the associated interfering users, cancel their contribution from the received data, and then estimate the channels of the reference cell users using the interference-canceled data.

Based on (52) and the decomposition $\mathbf{H}_{dr} = [\mathbf{H}_{dri}\ \ \mathbf{H}_{dro}]$, the linear MMSE estimator of $\mathbf{x}_c(n)$ is given by

$$\hat{\mathbf{x}}_c(n) = \mathbf{H}_{ec}\mathbf{y}(n) \tag{63}$$

where

$$\mathbf{x}_c(n) = [\underbrace{x_{11}(n)\ \cdots\ x_{1K_1}(n)}_{=\mathbf{x}_1^\top(n)}\ \underbrace{x_{r1}(n)\ \cdots\ x_{rK_{ri}}(n)}_{=\mathbf{x}_{ri}^\top(n)}]^\top, \tag{64}$$

$$\mathbf{H}_{ec} = \begin{bmatrix} \mathbf{H}_{d1}^H \\ \mathbf{H}_{dri}^H \end{bmatrix}\left[\mathbf{H}_{d1}\mathbf{H}_{d1}^H + \mathbf{H}_{dr}\mathbf{H}_{dr}^H + \sigma_v^2\mathbf{I}_{N_r}\right]^{-1} \tag{65}$$

$$= \begin{bmatrix} \mathbf{H}_e \\ \mathbf{H}_{ei} \end{bmatrix}, \tag{66}$$

and $\mathbf{H}_e$ in (66) is the same as that in (56) for $K_r > 0$. In practice, we replace $\mathbf{H}_{d1}$, $\mathbf{H}_{dr}$, and σ_v^2 with their estimates $\hat{\mathbf{H}}_{d1}$, $\hat{\mathbf{H}}_{dr}$, and $\widehat{\sigma_v^2}$, respectively. Let $\mathbf{x}_{1q}(n)$ and $\mathbf{x}_{riq}(n)$ denote quantized $\hat{\mathbf{x}}_{1q}(n) = \mathbf{H}_e\mathbf{y}(n)$ and $\hat{\mathbf{x}}_{riq}(n)\mathbf{H}_{ei}\mathbf{y}(n)$, respectively, obtained via hard thresholding based on know signal constellation. We cancel contribution of $\mathbf{x}_{riq}(n)$, and of pilot sequences, to obtain

$$\mathbf{y}_c(n) = \begin{cases} \mathbf{y}(n) - \hat{\mathbf{H}}_{dri}\mathbf{s}_{ri}(n) & \text{for } n = 1, 2, \cdots, P, \\ \mathbf{y}(n) - \hat{\mathbf{H}}_{dri}\mathbf{x}_{riq}(n) & \text{for } n = P+1, \cdots, P+T_d \end{cases} \tag{67}$$

where the ith component of K_{ri}-column $\mathbf{s}_{ri}(n)$ equals the corresponding re-used pilot sequence. Since $\hat{\mathbf{H}}_{dr0}$ is not free of scaling ambiguities, it is not suitable for IC. We now

Algorithm 1 Detection and Mitigation of Reused Pilots

Input: Received signal $\mathbf{y}(n)$, for $n = 1, 2, \cdots, P$, during training phase, and for $n = P+1, P+2, \cdots, P+T_d$, during data phase; orthogonal training sequences of K_1 users in the reference cell; total number of training sequences K_0.

Output: Estimated channels $\hat{\mathbf{h}}_{1i}$, $i = 1, 2, \cdots, K_1$, of K_1 users in the reference cell.

1: Compute sample correlation matrix $\hat{\mathbf{R}}_{yt}$ as in (22) using training phase data. Carry out EVD of $\hat{\mathbf{R}}_{yt}$ to obtain ordered eigenvalues $\ell_{t1} \geq \ell_{t2} \geq \cdots \geq \ell_{tN_r}$. Estimate noise variance $\widehat{\sigma_v^2}$ as in (24).

2: Using training data and training sequences of K_1 users in the reference cell, compute the estimate $\hat{\mathbf{h}}_{Ci}$ of $\mathbf{h}_{Ci}$ as defined in (16), via (17), for $i = 1, 2, \cdots, K_1$. Using $\widehat{\sigma_v^2}$ and $\mathbf{h}_{Ci}$s, compute α_1 as in (26) for some chosen $0 < \mu \leq 1$. With $\alpha_2 =$ minimum received SNR for in-cell users, compute threshold τ as in (27).

3: Compute sample correlation matrix $\hat{\mathbf{R}}_{yd}$ as in (23) using data phase received signal. Carry out EVD of $\hat{\mathbf{R}}_{yd}$ to obtain ordered eigenvalues $\ell_{d1} \geq \ell_{d2} \geq \cdots \geq \ell_{dN_r}$. The number of significant user signals in the reference cell is estimated as

$$\hat{K}_d = \max\left(K_1, \sum_{i=1}^{N_r} \mathbf{1}_{\{\ell_{di} \geq \tau\}}\right),$$

and the number of significant extraneous user signals is estimated as

$$K_r = \hat{K}_d - K_1.$$

4: **if** $K_r > 0$ **then**

5: With EVD of $\hat{\mathbf{R}}_{yd}$ as expressed in (29), compute $\tilde{\mathbf{y}}(n) = \hat{\mathbf{U}}_1^H \mathbf{y}(n)$ for $n = P+1, \cdots, P+T_d = T$. Apply RobustICA algorithm of [20] to $\tilde{\mathbf{y}}(n)$ to estimate $\tilde{\mathbf{H}}_d$ specified in (36), as $\hat{\tilde{\mathbf{H}}}_d$. Next estimate $\mathbf{H}_d$ specified in (30) and (31) as $\hat{\mathbf{H}}_d = \hat{\mathbf{U}}_1 \hat{\tilde{\mathbf{H}}}_d$.

6: Using the estimates $\hat{\mathbf{h}}_{Ci}$, $i = 1, 2, \cdots, K_1$, compute $\hat{\mathbf{H}}^{(p)}$ as in (18). Using $\hat{\mathbf{H}}_d$ and $\hat{\mathbf{H}}^{(p)}$, compute $\hat{\mathbf{G}}$ as specified in (42). Replace the ith column $\hat{\mathbf{G}}_i$ of $\hat{\mathbf{G}}$ with $[|\hat{\mathbf{G}}_{i1}| \cdots |\hat{\mathbf{G}}_{i(K_1+K_r)}|]^\top / \|\hat{\mathbf{G}}_i\|$, and denote the resulting matrix by $\bar{\mathbf{G}}$. If $\bar{\mathbf{G}}_{ij} < \tau_1$, set $\bar{\mathbf{G}}_{ij} = 0$, where in our simulations we set $\tau_1 = 0.15$. The thresholded $\bar{\mathbf{G}}$ is still denoted by $\bar{\mathbf{G}}$.

7: Follow the procedure given in item 3), *Channel Resolution*, of Sec. III-C1, to obtain the estimated channels $\hat{\mathbf{h}}_{1i}$, $i = 1, 2, \cdots, K_1$, where $\hat{\mathbf{h}}_{1i}$ estimates $\sqrt{p_{1i}}\,\mathbf{h}_{1i}$ and knowledge of p_{1i} is not assumed. We also obtain estimates $\hat{\mathbf{h}}_{rj}$ of interfering users' channels $\sqrt{p_{rj}}\,\mathbf{h}_{rj}$, $j = 1, 2, \cdots, K_r$.

8: **else**

9: Carry out EVD of $\hat{\mathbf{R}}_{yd}$ as expressed in (29) with $K_r = 0$. Compute $\mathbf{y}'(n) = \mathbf{U}_1^H \mathbf{y}(n)$ for $n = 1, 2, \cdots, P$, and then compute $\hat{\tilde{\mathbf{H}}}_{d1}$ via (48). Finally, the estimated channels $\hat{\mathbf{h}}_{1i}$, $i = 1, 2, \cdots, K_1$, are the K_1 columns of $\hat{\mathbf{H}}_{d1}$ specified in (47).

10: **end if**

11: Perform data detection and channel re-estimation using the entire dataset, as described in Sec. III-E.

modify (58) using interference-canceled $\mathbf{y}_c(n)$, to re-estimate the multi-user channel $\mathbf{H}_{d1}$ via least-squares as

$$\hat{\mathbf{H}}_{d1} = T^{-1} \sum_{n=1}^{T} \mathbf{y}_c(n) \tilde{\mathbf{x}}_1^H(n). \tag{68}$$

The above estimates may be used to design a linear MMSE detector to operate on $\mathbf{y}_c(n)$. □

Remark 3 (Approaches of [11], [12]): The semi-blind method discussed in Sec. III-D (for $K_r = 0$), coupled with data detection and channel re-estimation discussed in Sec. III-E, follows [11], [12] closely, but not exactly. The matrix $\mathbf{U}_1$ in Sec. III-D is estimated exactly as in [11] based on the EVD of $\hat{\mathbf{R}}_{yd}$, instead of using the SVD of a data matrix as in [12]. We have not found any discernible difference in the two approaches. The primary difference from [11], [12] lies in how $\tilde{\mathbf{H}}_{d1}$ in (44) is estimated, and how data is detected based on the estimated channel. In [11], for large N_r, $\tilde{\mathbf{H}}_{d1}$ is shown to be a diagonal matrix, and only its diagonal elements are estimated using pilots, with off-diagonal elements all set to zero. In this paper, we estimate the entire matrix $\tilde{\mathbf{H}}_{d1}$ using pilots, without setting off-diagonal elements to zero, an assumption that holds only as $N_r \to \infty$. Turning to detection of data discussed in Sec. III-E, we use a linear MMSE detector (56) based on model (52). In [11], a zero-forcing receiver is used, which in our notation is $\mathbf{H}_e = \left(\mathbf{H}_{d1}^H \mathbf{H}_{d1}\right)^{-1} \mathbf{H}_{d1}^H$, the pseudo-inverse of $\mathbf{H}_{d1}$. In [12], the data is detected not using the full channel matrix, but the "subspace channel matrix" which is $\tilde{\mathbf{H}}_{d1}$ in our notation. This is also the method simulated in [13] to compare their approach with that of [12]. We use the full channel matrix in our detector, as well as in the detector used to simulate [12]. Also, although not explicitly noted therein, [12] seems to use the data detector scheme of [1], which would be $\mathbf{H}_e = \tilde{\mathbf{H}}_{d1}^H$, the conjugate-transpose of the subspace channel matrix. Note that in implementation, all true channels are replaced with their estimates. Finally, we note again that the approaches of [11], [12] assume that the strongest K_1 users at the BS are the in-cell users, hence, they do not have counterparts to our Secs. III-B and III-C. □

Remark 4 (Distributed Antenna Systems): The proposed approach applies to massive MIMO cellular wireless networks with any type of cells (macro or small cells, such as femto, pico or micro cells) so long as the base stations are centralized, since our mathematical system model (1)-(2) applies to all of them. How about a distributed antenna system (DAS) where instead of a single centralized antenna system in each cell, one has several distributed remote access units (RAUs) with multiple antennas, geographically separated over the cell [22], [23]. Note that processing in a DAS is still centralized [22], [23]. We claim that, in principle, our approach applies to DAS also. We now briefly outline modifications needed in our model to accommodate DAS. In (1)-(2), $p_{\ell i}$ also models the effects of large scale fading. For $\mathbf{y}(n) \in \mathbb{C}^{N_r}$, different sets of components of $\mathbf{y}(n)$ belong to different RAUs in DAS. Suppose that we have 2 RAUs per cell, each with $\bar{N}$ antennas. Then $N_r = 2\bar{N}$, and let $\mathbf{y}(n) = [\mathbf{y}_1^\top(n) \;\; \mathbf{y}_2^\top(n)]^\top$ where $\mathbf{y}_k(n)$ is data collected at RAU k, $k = 1, 2$. Then in (1)-(2), we let $\mathbf{h}_{\ell i_\ell} = [\mathbf{h}_{1\ell i_\ell}^\top(n) \;\; \mathbf{h}_{2\ell i_\ell}^\top(n)]^\top$, where $\mathbf{h}_{k\ell i_\ell}(n)$ is the channel from the ith user in the ℓth cell to RAU k of the reference-cell. Unlike $\mathbf{h}_{\ell i} \sim \mathcal{N}_c(0, \mathbf{I}_{N_r})$ representing small-scale fading in Sec. II, now we let $\mathbf{h}_{k\ell i_\ell}(n)$ represent the effects of both large- and small-scale fading, so that $\mathbf{h}_{k\ell i_\ell}(n) \sim \mathcal{N}_c(0, \beta_k \mathbf{I}_{\bar{N}})$ for some $\beta_k > 0$, and $p_{\ell i}$ now represents just the average transmitted power of the ith user in the ℓth cell. So long as the asymptotic orthogonality of distinct channels holds as in (14), all our results will hold true. Of course, details such as selection of threshold τ in step 3 of Algorithm 1, need to be worked out. □

IV. Simulation Examples

Consider a 7-cell network, with $K_\ell = 5$ users/cell, $\ell = 1, 2, \cdots, 7$, total 35 users, and $K_0 = 8$ orthogonal pilots of length $P = 8$ symbols. In the 6 nearest-neighbor cells, among total 30 users, 20 users re-use some of the reference cell pilots, and 10 users employ others pilots that are not in use in the reference cell. The nominal average SNR for reference cell ($\ell = 1$) users at the reference cell BS is 10dB ($=p_{1i}/\sigma_v^2$, $i = 1, 2, \cdots, 5$). There is a lack of perfect power control. In order to reflect this, actual average SNR for cell $\ell = 1$ was taken as uniformly distributed over 10 ± 3dB. Of the 20 interfering users that reuse pilots, average SNR at the reference cell BS is uniform over $(p_{rj}/\sigma_v^2) \pm 3dB$ for 5 users, and it is uniform over $(p_{rj}/\sigma_v^2) - 9 \pm 3dB$ for 15 users, and p_{rj} is such that p_{rj}/σ_v^2 varies from -20dB through 20dB, and it is the same for all indexes j. The stronger 5 users may be thought of being located at cell edges when p_{rj} is comparable to p_{1i}, while other 15 interfering users are farther off from BS. Of the 10 interfering users that do not reuse any reference cell pilots, average SNR at the reference cell BS is uniform over $(p_{rj}/\sigma_v^2) \pm 3dB$ for 2 users, and it is uniform over $(p_{rj}/\sigma_v^2) - 9 \pm 3dB$ for 8 users.

At the reference-cell BS we have $N_r = 100$ or 200 antennas. Orthogonal (binary) Hadamard sequences of length $P = 2^3 = 8$ are selected as training sequences, and the information sequences $\{x_{\ell i_\ell}(n)\}$ were i.i.d. QPSK. We have $P = 8$ (training bits), and pick T_d =136 or 184 (data symbols), leading to T =144 or 192. All simulation results are based on 10,000 Monte Carlo runs.

In applying our proposed approach, we picked τ used in step 3 of Algorithm 1 as specified in (27), where α_1 is given by (26) with $\mu = 0.7$, and $\alpha_2 = 5$ (least SNR at the BS for any in-cell user). For performance comparisons, we also simulated the semi-blind subspace-based approach of [11], [12], as discussed in Sec. III-D and as clarified in Remark 3, labeled "SB-SVD" in figures, and the semi-blind approach of [13], labeled "SB-kurt" in figures, after changing synchronous pilots (all pilots from all users in all cells occupy the same time-slot) to asynchronous pilots, as detailed in [13]. The other details such as length of pilot sequences and number of available pilots, were kept the same for all approaches.

Fig. 1 shows the results of our matrix rank determination method of Sec. III-B for determining the number of extraneous users, including reused pilots. With reference to (28), it is seen that for higher values of $p_{rj}/\sigma_v^2 = p_r/\sigma_v^2$ ($\forall j$), the number of detected interfering users increase with average p_r/σ_v^2, since

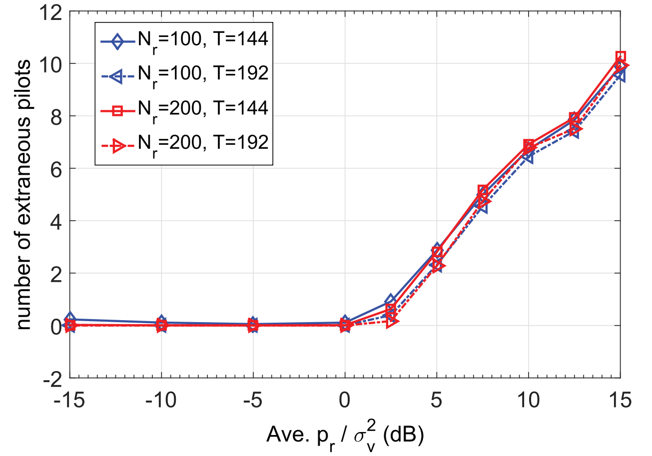

Fig. 1. Total number of detected extraneous (interfering) users K_r as a function of average p_r ($=p_{rj}$ $\forall j$) relative to noise power σ_v^2 when in-cell user's average power is fixed at average $p_{1i}/\sigma_v^2 = 10$dB $\forall i$: K_1=5= number of in-cell users, total 35 users in 7 cells. The method of Sec. III-B is used for extraneous user detection. The notation is in reference to (28).

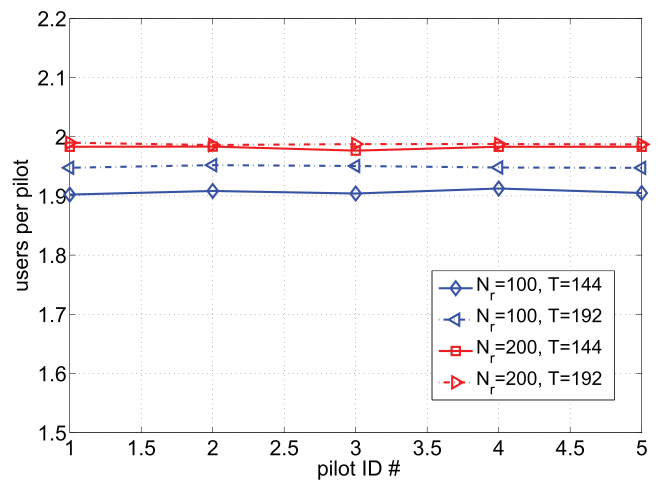

Fig. 2. Number of detected users per pilot. Ave. $p_r/\sigma_v^2 = 10dB$, $p_1/\sigma_v^2 = 10$dB. The pilots are numbered 1 through 8 and pilots 1 through 5 are reused. Based on 10,000 runs. The method of Sec. III-C1 is used for pilot identification.

we have five strong interfering users that reuse reference cell pilots, and five other strong interfering users that do not reuse reference cell pilots. The method is designed to ignore weak signals, and as seen in Fig. 1, for lower values of p_r/σ_v^2 (relative to average $p_{1i}/\sigma_v^2 = 10$dB $\forall i$), the proposed method ignores reused pilots and extraneous users.

Figs. 2-4 show the performance of our method of Sec. III-C1 regarding determination of reused/contaminated pilots, for average p_r/σ_v^2=15, 10 or 5dB, respectively. It is seen that reused pilots are detected accurately. Since in all of Figs. 2-4, average $p_{1i}/\sigma_v^2 = 10$dB $\forall i$, when average p_r/σ_v^2=5dB, the number of interfering users detected per pilot decrease significantly in Fig. 4, as they are all, on the average, weak relative to in-cell users. The opposite is true in Fig. 2 where average p_r/σ_v^2=15dB.

Figs. 5 and 6 show the normalized mean-square error (MSE) in multi-user channel estimation, which for estimated

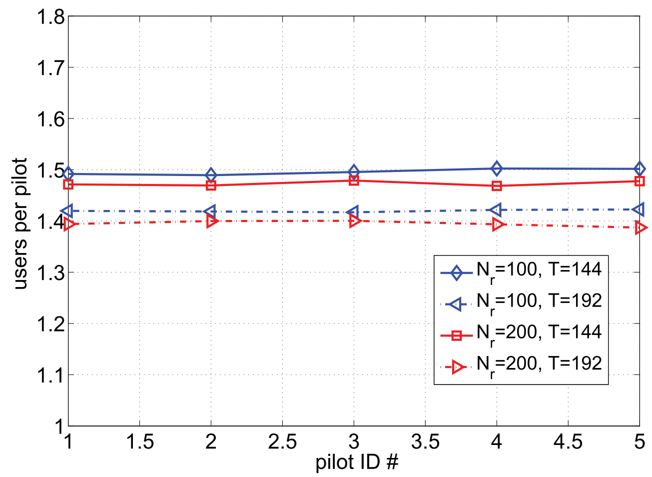

Fig. 3. Number of detected users per pilot when ave. $p_r/\sigma_v^2 = 5dB$; the rest as in Fig. 2.

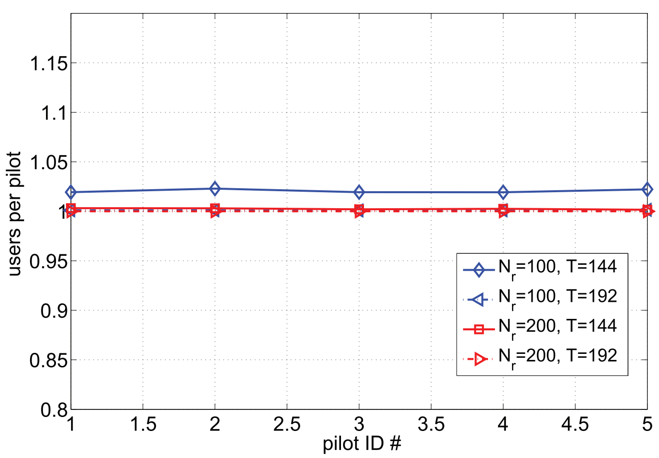

Fig. 4. Number of detected users per pilot when ave. $p_r/\sigma_v^2 = 0dB$; the rest as in Fig. 2.

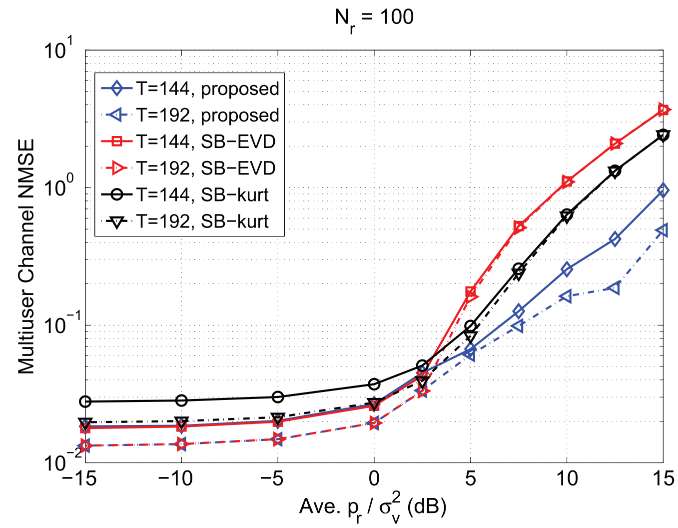

Fig. 5. Normalized MSE (69) of channel estimation error for reference cell users vs average p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. $N_r = 100$. Based on 10,000 runs. The approach labeled "SB-SVD" is based on [11], [12]; see Sec. III-D and Remark 3 in Sec. III-F. The approach labeled "SB-kurt" is that of [13].

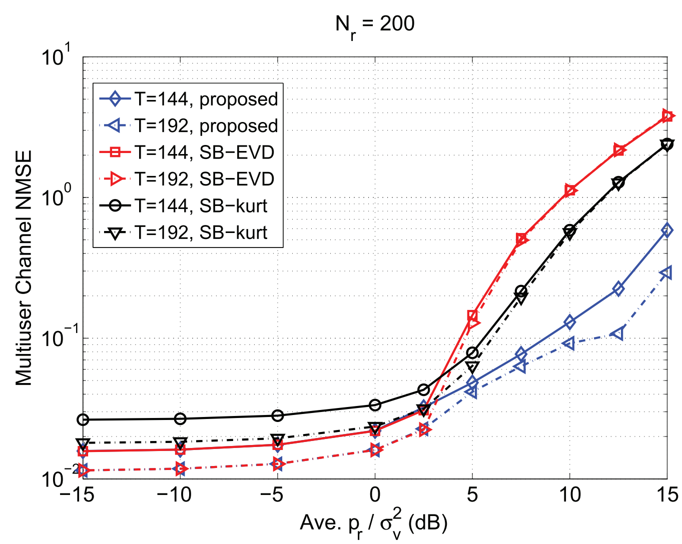

Fig. 6. As for Fig. 5 except that $N_r = 200$.

multi-user channel $\hat{\mathbf{H}}_{tr}^m$ and true channel $\mathbf{H}_{tr}^m$ (defined as in (46), with power levels p_{1i}s included) in the mth Monte Carlo run, is defined as

$$\text{NMSE} = \frac{1}{M}\sum_{m=1}^{M}\frac{\|\hat{\mathbf{H}}_{tr}^m - \mathbf{H}_{tr}^m\|_F^2}{\|\mathbf{H}_{tr}^m\|_F^2}, \tag{69}$$

where $\|\mathbf{H}\|_F$ denotes the Frobenius norm, there are $M = 10000$ runs and $\hat{\mathbf{H}}_{tr}^m$ follows from the procedure of Sec. III-C. We also show the results of the semi-blind subspace approach of [11], [12], as discussed in Sec. III-D and as clarified in Remark 3, labeled "SB-SVD," and of the approach of [13], labeled "SB-kurt." Comparing our approach with that of [11], [12], it is seen that when reused pilots are at a power significantly lower than in-cell users, there is little ill-effect. But as the out-of-cell users with reused pilots become relatively stronger, the approach of [11], [12] yields poorer results compared to the proposed approach, since [11], [12] consider K_1 strongest users as in-cell users, which is not always true. Comparing our proposed approach with that of [13], it is seen from Figs. 5 and 6 that our approach outperforms [13]. While [13] exclusively uses higher-order statistics (kurtosis), we use both second-order statistics (correlation matrix) and kurtosis. In general, for a given data length, the estimates of second-order statistics have lower variances than the estimate of kurtosis. In [13], for each in-cell user, the number of complex-valued unknowns (equivalent to $\mathbf{w}$ in Remark 1) needed to be estimated using kurtosis equal N_r (=100 or 200), whereas in our approach, the number of complex-valued unknowns per significant user (in-cell as well as interfering users) is $\hat{K}_d$, which in our simulations varies from 5 to 15 (see Fig. 1). That is, the approach of [13] has much more high variance parameter estimators than our proposed approach. Also, the presence of higher-power interfering users' signals, not explicitly modeled in [13] unlike in our approach, leads to higher variance estimate of the kurtosis of the desired in-cell signals.

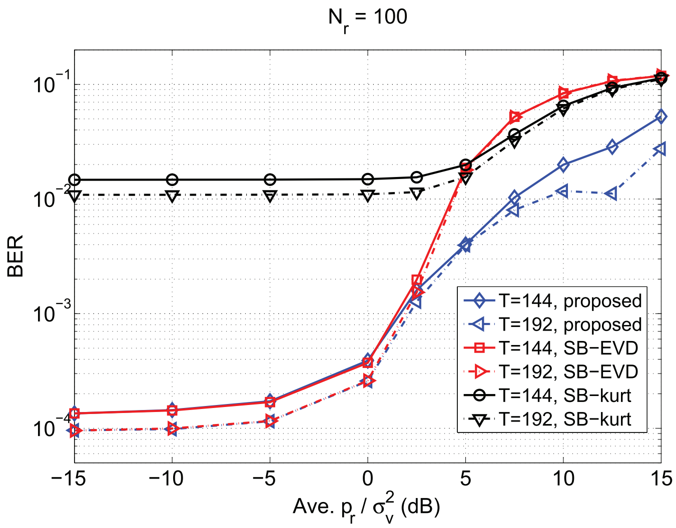

Fig. 7. Ave. BER for reference cell users vs p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. Based on 10,000 runs. The approach labeled "SB-SVD" is based on [11], [12]; see Sec. III-D and Remark 3 in Sec. III-F. The approach labeled "SB-kurt" is that of [13].

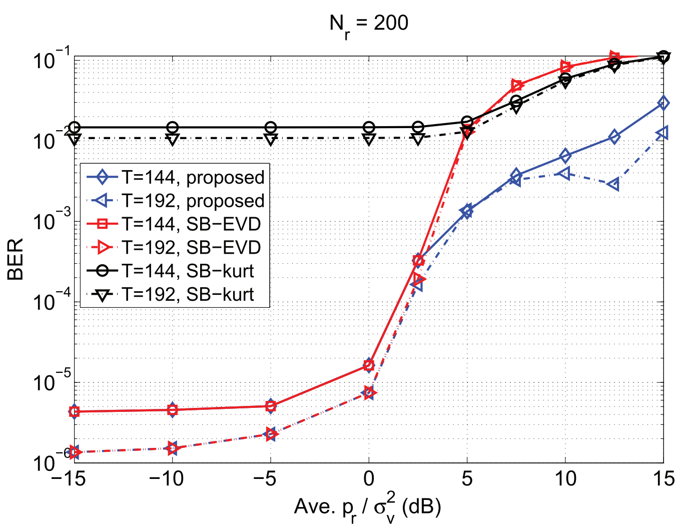

Fig. 8. As for Fig. 7 except that $N_r = 200$.

Figs. 7 and 8 show the bit-error rate for QPSK information sequences corresponding to the results of Figs. 5 and 6, respectively. Again, at higher power levels of interfering out-of-cell users, the performance of the semi-blind approach of [11], [12] is much poorer compared to the proposed approach. The BER results of the approach of [13] are also much worse than that for our proposed approach, and are similar to what has been reported in [13] for a different example.

We now turn to the effects of the choice of μ in (26) and step 2 of Algorithm 1. As noted in the discussion following (26), $\mu \in (0, 1]$ impacts the threshold used to determine a subspace of reduced rank. For our proposed approach, in Fig. 9, we redo the results of Fig. 6 for μ =0.3, 0.7 and 1. It is seen that there is no difference in channel MSE for different values of μ, except for average p_r/σ_v^2=15dB, i.e., the results are not sensitive to the choice.

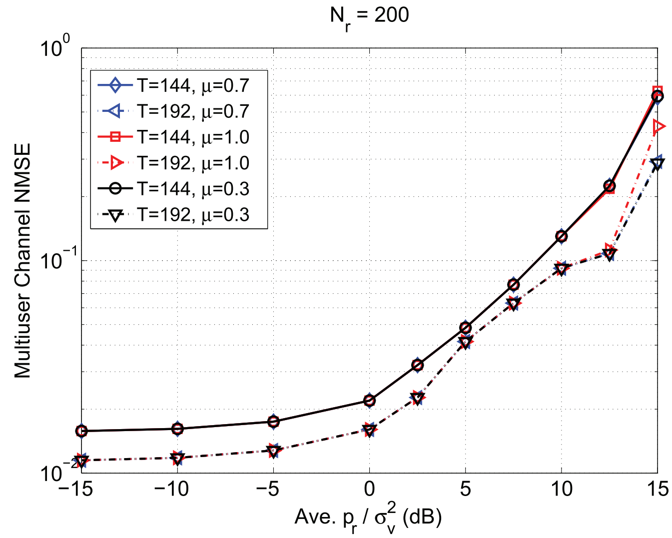

Fig. 9. Normalized MSE (69) of channel estimation error for varying μ (see (26) and step 2 of Algorithm 1), using the proposed approach, for reference cell users vs average p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. μ is 0.3, 0.7 or 1, $N_r = 200$. Based on 10,000 runs.

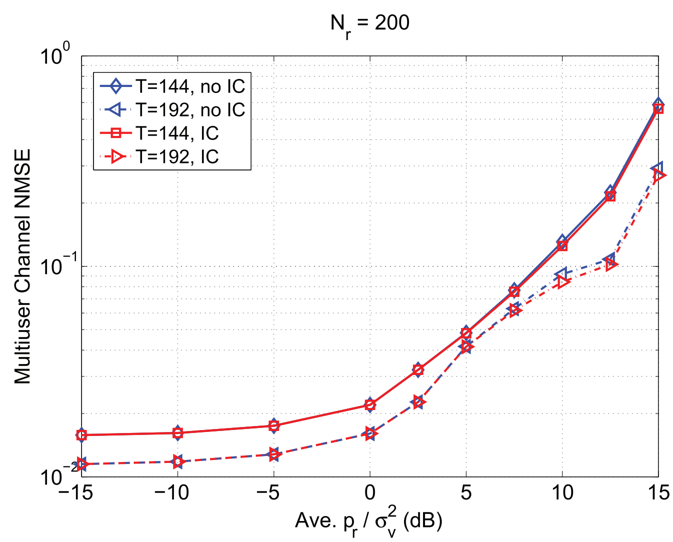

Fig. 10. Normalized MSE (69) of channel estimation error, using the proposed approach, with and without interference cancellation (IC) as discussed in Remark 2 in Sec. III-F, for reference cell users vs average p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. Based on 10,000 runs.

In Remark 2 of Sec. III-F, an IC approach was suggested where contributions of interfering users with estimated channels having no scaling ambiguities, could be removed from data $\mathbf{y}(n)$ to reduce interference, and then one could estimate in-cell users' channels and information symbols by operating on interference-canceled data $\mathbf{y}_c(n)$. For our proposed approach, in Fig. 10, we redo the results of Fig. 6 with and without IC. One can see small improvements in channel MSE when IC is used compared to when IC is not employed (as in Figs. 5-9). However, the improvement does not appear to be significant enough to warrant use of the IC approach.

V. Conclusion

We proposed a novel subspace-based semi-blind approach where we have training data as well as information bearing data from various users (both in-cell and neighboring cells) at the base station. Unlike existing semi-blind methods, we allow the interfering users from neighboring cells to be at higher power levels at the BS compared to the in-cell users. Unlike existing approaches, the BS estimates the channels of all users: in-cell and significant neighboring cell users, i.e., ones with comparable power levels at the BS. We exploited both subspace method using pilot and data correlation matrices, as well as blind source separation (ICA) using higher-order statistics. The proposed approach was illustrated via simulation examples and was shown to outperform [11]–[13].

References

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

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

[3] L. Lu, G. Y. Li, A. L. 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.

[4] 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.

[5] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2nd Quart., 2016.

[6] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[7] S. Jin, X. Wang, Z. Li, K.-K. Wong, Y. Huang, and X. Tang, "On massive MIMO zero-forcing transceiver using time-shifted pilots," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 59–74, Jan. 2016.

[8] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[9] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

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

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

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

[13] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 525–536, Jan. 2016.

[14] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.

[15] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[16] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[17] J. K. Tugnait, "Self-contamination for detection of pilot contamination attack in multiple antenna systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 5, pp. 525–528, Oct. 2015.

[18] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 658–672, Mar. 1997.

[19] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[20] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 248–261, Feb. 2010.

[21] P. Comon and C. Jutten, *Handbook of Blind Source Separation*. New York, NY, USA: Academic, 2010.

[22] X.-H. You, D.-M. Wang, B. Sheng, X.-Q. Gao, X.-S. Zhao, and M. Chen, "Cooperative distributed antenna systems for mobile communications [coordinated and distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 35–43, Jun. 2010.

[23] H. Ren, N. Liu, C. Pan, and C. He, "Energy efficiency optimization for MIMO distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2276–2288, Mar. 2017.

Jitendra K. Tugnait (M'79–SM'93–F'94) received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and E.E. degrees from Syracuse University, Syracuse, NY, USA, and the Ph.D. degree from the University of Illinois at Urbana–Champaign in 1973, 1974, and 1978, respectively, all in electrical engineering.

From 1978 to 1982, he was an Assistant Professor of electrical and computer engineering with the University of Iowa, Iowa City, IA, USA. From 1982 to 1989, he was with the Long Range Research Division of the Exxon Production Research Company, Houston, TX, USA. He joined as a Professor with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA, in 1989. He currently holds the title of James B. Davis Professor. His current research interests are in statistical signal processing, wireless communications, and multiple target tracking.

Dr. Tugnait is a past Associate Editor of the IEEE Transactions on Automatic Control, the IEEE Transactions on Signal Processing, the IEEE Signal Processing Letters, the IEEE Transactions on Wireless Communications, and past Senior Area Editor of the IEEE Transactions on Signal Processing. He is currently a Senior Editor of the IEEE Wireless Communications Letters..