Pilot Decontamination Under Imperfect Power Control

Jitendra K. Tugnait
Dept. of Electrical & Computer Eng.
Auburn University, Auburn, AL 36849, USA

***Abstract*—In a time-division duplex (TDD) multiple antenna system the channel state information (CSI) can be estimated using reverse training. In multicell multiuser 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 subspace approaches assume that the interfering users from neighboring cells are always at distinctly lower power levels at the BS compared to the in-cell users. 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. The proposed approach is illustrated via simulation examples.**

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 [2], [3]. 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. In a given cell, the pilots are selected to be orthogonal.

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. The approaches include multi-cell cooperation [4], subspace-based methods using SVD (singular value decomposition) of data matrix [5] or EVD (eigenvalue decomposition) of data correlation matrix [6], semi-blind approaches [7] and others [8]. These 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, etc.

In this paper we consider a subspace-base 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 [5], [6] by additional features. 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.

This work was supported by the National Science Foundation under Grant ECCS-1651133.

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) per cell 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 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.

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

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

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

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

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 ($\mathbf{1}_A$ denotes an indicator function)

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

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.

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 $p_{\ell i_\ell}$. 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.

Correlation Matrices: Define the correlation matrices of measurements $\mathbf{R}_{yt} = P^{-1} \sum_{n=1}^{P} \mathbb{E}\left\{\mathbf{y}(n)\mathbf{y}^H(n)\right\}$, $\mathbf{R}_{yd} = T_d^{-1} \sum_{n=1+P}^{T} \mathbb{E}\left\{\mathbf{y}(n)\mathbf{y}^H(n)\right\}$, and the correlation matrices of users' signals as $\mathbf{R}_{st} = P^{-1} \sum_{n=1}^{P} \mathbb{E}\left\{[\mathbf{y}(n) - \mathbf{v}(n)][\mathbf{y}(n) - \mathbf{v}(n)]^H\right\}$, $\mathbf{R}_{sd} = T_d^{-1} \sum_{n=1+P}^{T} \mathbb{E}\left\{[\mathbf{y}(n) - \mathbf{v}(n)][\mathbf{y}(n) - \mathbf{v}(n)]^H\right\}$. Then we have

$$\mathbf{R}_{yt} = \mathbf{R}_{st} + \sigma_v^2 \mathbf{I}_{N_r}, \;\; \mathbf{R}_{yd} = \mathbf{R}_{sd} + \sigma_v^2 \mathbf{I}_{N_r}. \tag{7}$$

It follows from (3) and (5) that

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

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

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

$$\lim_{N_r \to \infty} N_r^{-1} \mathbf{h}_{\ell_1 i_{\ell_1}}^H \mathbf{h}_{\ell_2 i_{\ell_2}} = \delta_{\ell_1,\ell_2} \delta_{i_{\ell_1}, i_{\ell_2}} \;\; \text{w.p.1.} \tag{8}$$

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

$$\mathbf{h}_{Ci} = \begin{cases} \sqrt{p_{1i}}\, \mathbf{h}_{1i} + \tilde{\mathbf{h}}_{1i}, & 1 \le i \le K_1 \\ \tilde{\mathbf{h}}_{1i}, & K_1 + 1 \le i \le K_0 \end{cases} \tag{9}$$

This suggests that 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 $\mathbf{R}_{st}$, and they 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 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

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 (9), the channel corresponding to the ith pilot for $i = 1, 2, \cdots, K_1$, is estimated as

$$\hat{\mathbf{h}}_{Ci} = P^{-1} \sum_{n=1}^{P} \mathbf{y}(n) s_{ti}^*(n). \tag{10}$$

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 $\mathbf{H}^{(p)}$ and its estimate $\hat{\mathbf{H}}^{(p)}$ as

$$\mathbf{H}^{(p)} = [\mathbf{h}_{C1} \; \cdots \; \mathbf{h}_{CK_1}], \quad \hat{\mathbf{H}}^{(p)} = [\hat{\mathbf{h}}_{C1} \; \cdots \; \hat{\mathbf{h}}_{CK_1}]. \tag{11}$$

For large N_r, taking expectation w.r.t. noise only,

$$\begin{aligned} \mathbb{E}\left\{\|\hat{\mathbf{h}}_{Ci}\|^2\right\} &= p_{1i}\|\mathbf{h}_{1i}\|^2 + \|\tilde{\mathbf{h}}_{1i}\|^2 + \sqrt{p_{1i}}\, \mathbf{h}_{1i}^H \tilde{\mathbf{h}}_{1i} + \sigma_v^2 N_r/P \\ &\approx p_{1i}\|\mathbf{h}_{1i}\|^2 + \sigma_v^2 N_r/P \\ &+ \sum_{\ell=2}^{L} \sum_{i_\ell=1}^{K_\ell} p_{\ell i_\ell} \|\mathbf{h}_{\ell i_\ell}\|^2 \mathbf{1}_{\{x_{\ell i_\ell}(n) = s_{ti}(n),\; n=1,2,\cdots,P\}}. \end{aligned} \tag{12}$$

For large N_r, $\mathbb{E}\left\{\|\hat{\mathbf{h}}_{Ci}\|^2\right\} \approx \|\hat{\mathbf{h}}_{Ci}\|^2$.

Define the sample correlation matrices under training and data phases as $\hat{\mathbf{R}}_{yt} = P^{-1} \sum_{n=1}^{P} \mathbf{y}(n)\mathbf{y}^H(n)$, $\hat{\mathbf{R}}_{yd} = T_d^{-1} \sum_{n=1+P}^{T} \mathbf{y}(n)\mathbf{y}^H(n)$. Let the ordered eigenvalues of $\hat{\mathbf{R}}_{yt}$ be denoted by $\ell_{t1} \ge \ell_{t2} \ge \cdots \ge \ell_{tN_r}$ in decreasing order of magnitude, and that of $\hat{\mathbf{R}}_{yd}$ be denoted by $\ell_{d1} \ge \ell_{d2} \ge \cdots \ge$

ℓ_{dN_r}. First we wish to determine the significant number of 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 $\widehat{\sigma_v^2} = \frac{1}{N_r - K_0} \sum_{i=1+K_0}^{N_r} \ell_{ti}$. Then we have

$$p_{1i}\|\mathbf{h}_{1i}\|^2 + \sum_{\ell=2}^{L} \sum_{i_\ell=1}^{K_\ell} p_{\ell i_\ell} \|\mathbf{h}_{\ell i_\ell}\|^2 \mathbf{1}_{\{x_{\ell i_\ell}(n)=s_{ti}(n),\, n=1,2,\cdots,P\}} \approx \|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P. \tag{13}$$

Now consider the eigenvalues of data correlation matrix $\hat{\mathbf{R}}_{yd}$. 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 (13) is approximately less than $Lp_{1i}\|\mathbf{h}_{1i}\|^2$, so that $p_{1i}\|\mathbf{h}_{1i}\|^2 \geq (1/L)[\|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P]$, $i = 1, 2, \cdots, K_1$. Let

$$\alpha_1 = \min_{1 \leq i \leq K_1} (1/L)[\|\hat{\mathbf{h}}_{Ci}\|^2 - \widehat{\sigma_v^2} N_r/P]. \tag{14}$$

This discussion implies that the eigenvalues ℓ_{di} of the data correlation matrix corresponding to in-cell users will exceed $\alpha_1 + \widehat{\sigma_v^2}$, since the largest $\sum_\ell^L K_\ell$ eigenvalues ℓ_{di} of $\mathbf{R}_{yd}$ are of the form $p_{\ell i_\ell}\|\mathbf{h}_{\ell i_\ell}\|^2 + \sigma_v^2$. 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

$$\tau = \max\left(\alpha_1,\, (\alpha_2 N_r + 1)\widehat{\sigma_v^2}\right). \tag{15}$$

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. Let $\hat{K}_d$ denote the number of eigenvalues of $\hat{\mathbf{R}}_{yd}$ that exceed τ. (Note that $\hat{K}_d$ 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 $K_1 < K_0$.

B. Blind Channel Estimation in Data Phase

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

$$\mathbf{y}(n) = \sum_{i=1}^{K_1} \sqrt{p_{1i}}\, \mathbf{h}_{1i} x_{1i}(n) + \sum_{j=1}^{K_r} \sqrt{p_{rj}}\, \mathbf{h}_{rj} x_{rj}(n) + \tilde{\mathbf{v}}(n) \tag{16}$$

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 earlier 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 (16). Consider EVD of $\hat{\mathbf{R}}_{yd}$ to obtain

$$\hat{\mathbf{R}}_{yd} = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{U}}^H = [\hat{\mathbf{U}}_1\ \hat{\mathbf{U}}_2] \begin{bmatrix} \hat{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_2 \end{bmatrix} [\hat{\mathbf{U}}_1\ \hat{\mathbf{U}}_2]^H \tag{17}$$

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

$$\mathbf{H}_d = [\sqrt{p_{11}}\, \mathbf{h}_{11}\ \cdots\ \sqrt{p_{1K_1}}\, \mathbf{h}_{1K_1}\ \sqrt{p_{r1}}\, \mathbf{h}_{r1}\ \cdots\ \sqrt{p_{rK_r}}\, \mathbf{h}_{rK_r}]. \tag{18}$$

Then we can rewrite (16) as

$$\mathbf{y}(n) = \mathbf{H}_d \mathbf{x}(n) + \tilde{\mathbf{v}}(n), \tag{19}$$

$$\mathbf{x}(n) = [x_{11}(n)\ \cdots\ x_{1K_1}(n)\ x_{r1}(n)\ \cdots\ x_{rK_r}(n)]^T. \tag{20}$$

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 (7), we have

$$\mathbf{R}_{yd} = \mathbf{U}\Sigma\mathbf{U}^H = [\mathbf{U}_1\ \mathbf{U}_2] \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} [\mathbf{U}_1\ \mathbf{U}_2]^H \tag{21}$$

where $\mathbf{U}$, Σ, etc. in (21) are the true counterparts of the estimated $\hat{\mathbf{U}}$, $\hat{\Sigma}$, etc. in (17).

The channels $\mathbf{h}_{1i}$ and $\mathbf{h}_{rj}$ lie in the subspace spanned by the columns of $\mathbf{U}_1$. Consider, for $n = P + 1, \cdots, P + T_d = T$, $\tilde{\mathbf{y}}(n) = \mathbf{U}_1^H \mathbf{y}(n) \in \mathbb{C}^{K_1+K_r}$. Then we have $\tilde{\mathbf{y}}(n) = \mathbf{U}_1^H (\mathbf{H}_d \mathbf{x}(n) + \mathbf{v}(n)) = \tilde{\mathbf{H}}_d \mathbf{x}(n) + \check{\mathbf{v}}(n)$ where $\tilde{\mathbf{H}}_d \in \mathbb{C}^{(K_1+K_r)\times(K_1+K_r)}$ 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 (8)), we have $\mathbb{E}\{\tilde{\mathbf{v}}(n)\tilde{\mathbf{v}}^H(n)\} \approx \sigma_v^2 \mathbf{I}_{K_1+K_r}$ 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 (16) by appealing to (8).

Since data sequences are independent non-Gaussian, one can apply higher-order statistics-based approaches to estimate $\tilde{\mathbf{H}}_d$. We will use the RobustICA algorithm of [10] that uses kurtosis of "unmixed" measurements. It provides an estimate $\hat{\tilde{\mathbf{H}}}_d$ of $\tilde{\mathbf{H}}_d$ using $\tilde{\mathbf{y}}(n)$. For some θ_is, one obtains

$$\hat{\tilde{\mathbf{H}}}_d \approx \tilde{\mathbf{H}}_d \mathcal{P} \Gamma_\theta, \quad \Gamma_\theta = \mathrm{diag}\{e^{j\theta_i},\ i = 1, \cdots, K_1 + K_r\} \tag{22}$$

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 [11]), and one can only recover channels up to a constant of modulus one when using kurtosis and related criteria for unmixing. Thus, an estimate of $\mathbf{H}_d = \mathbf{U}_1 \tilde{\mathbf{H}}_d$ is given by

$$\hat{\mathbf{H}}_d = \hat{\mathbf{U}}_1 \hat{\tilde{\mathbf{H}}}_d \approx \mathbf{H}_d \mathcal{P} \Gamma_\theta \,. \tag{23}$$

C. 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 (11), 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 (22). The pilot-based channel estimates (10) yield $\hat{\mathbf{H}}^{(p)}$ while (23) 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 $(\mathbf{H}_d \mathcal{P} \Gamma_\theta)\,\mathbf{G} = \mathbf{H}^{(p)} \;\Rightarrow\; \hat{\mathbf{H}}_d \mathbf{G} \approx \hat{\mathbf{H}}^{(p)}$. Hence an estimate of $\mathbf{G}$ is given by $\hat{\mathbf{G}} = \left(\hat{\mathbf{H}}_d^H \hat{\mathbf{H}}_d\right)^{-1} \hat{\mathbf{H}}_d^H \hat{\mathbf{H}}^{(p)}$.

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 of $\hat{\mathbf{H}}^{(p)}$ 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 $\bar{\mathbf{G}}$ corresponding to this out-of-cell user with non-reused pilot, will have 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.

(1) Replace the ith column $\hat{\mathbf{G}}_i$ of $\hat{\mathbf{G}}$ with $[|\hat{\mathbf{G}}_{i1}| \;\cdots\; |\hat{\mathbf{G}}_{i(K_1+K_r)}|]^T / \|\hat{\mathbf{G}}_i\|$, 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 $s_{t,i}(n)$ 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.

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

$$\bar{\mathbf{H}} = [\hat{\mathbf{H}}_{dj_1} \;\cdots\; \hat{\mathbf{H}}_{dj_q}] \in \mathbb{C}^{N_r \times q}, \quad \mathbf{c} = [c_1 \;\cdots\; c_q]^T .$$

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 estimated channel.

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}/σ_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.

Fig. 1 shows the normalized mean-square error (MSE) in multiuser channel estimation, which for estimated multi-user channel $\hat{\mathbf{H}}_{tr}^{m}$ and true channel $\mathbf{H}_{tr}^{m}$ (defined as in (11) with $\mathbf{h}_{Ci}$ replaced with $\mathbf{h}_{1i}$) in the mth Monte Carlo run, is defined as

$$\text{NMSE} = \frac{1}{M}\sum_{m=1}^{M} \frac{\|\hat{\mathbf{H}}_{tr}^{m} - \mathbf{H}_{tr}^{m}\|_F^2}{\|\mathbf{H}_{tr}^{m}\|_F^2}\,, \tag{24}$$

where $\|\mathbf{H}\|_F$ denotes the Frobenius norm, and there are $M = 10000$ runs. We also show the results of the approach of [5], labeled "semi-blind." 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 semi-blind approach yields poorer results compared to the proposed approach. Figs. 2 and 3 show the bit-error rate for QPSK information sequences (Fig. 3 corresponds to the results of Fig. 1), when we employ linear MMSE multi-user deocder/equalizer/beamformer using the estimated channels via either the proposed approach or the semi-blind approach. Again, at higher power levels of interfering out-of-cell users, the performance is much poorer for the semi-blind method, compared to the proposed approach.

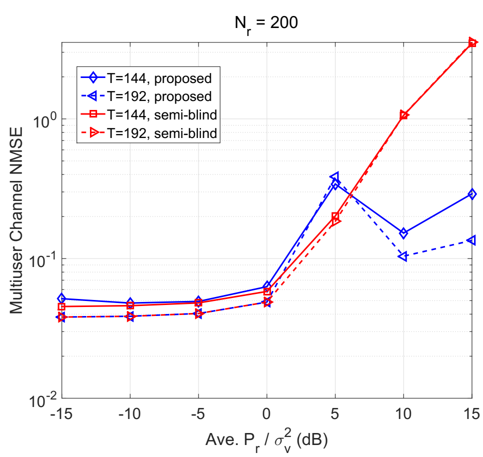

Fig. 1. *Normalized MSE (24) of channel estimation error for reference cell users vs average p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. $N_r = 200$. Based on 10,000 runs. The approach labeled "semi-blind" is based on [5], [6].*

References

[1] T.L. Marzetta, "Noncooperative cellular wireless with unlimited number of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590-3600, Nov. 2010.

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

[3] E. Bjornson, E.G. Larsson and T.L. Marzetta, "Massive MIMO: Ten myths and one critical question," *arXiv:1503.06854v2*, Aug. 2015.

[4] 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.

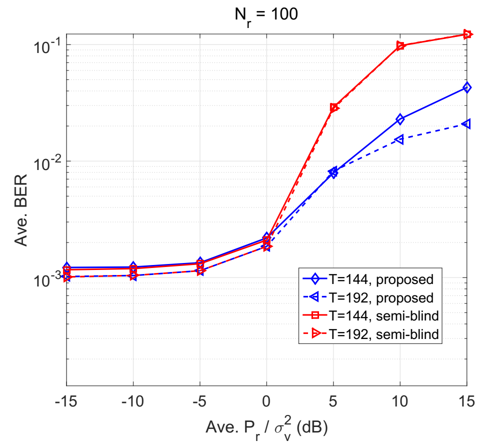

Fig. 2. *Ave. BER for reference cell users vs p_r/σ_v^2 with $p_1/\sigma_v^2 = 10$dB. Based on 10,000 runs. The approach labeled "semi-blind" is based on [5], [6].*

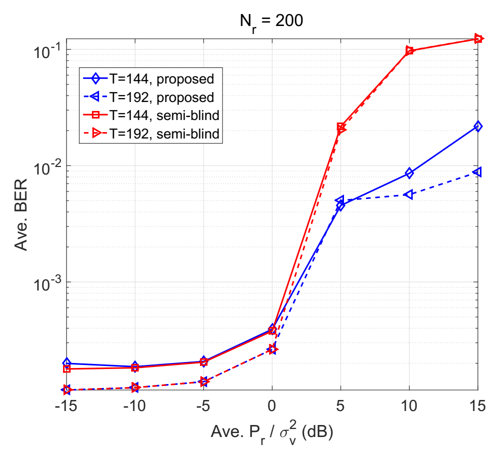

Fig. 3. *As for Fig. 2 except that $N_r = 200$.*

[5] R.R. Müller, L. Cottatellucci and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Proc.*, vol. 8, pp. 773-786, Oct. 2014.

[6] H.Q. Ngo and E.G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. ICASSP 2012*, pp. 3249-3252, Kyoto, Japan, March 2012.

[7] D. Hu, L. He and X. Wang, "Semi-blind pilot decontamination for massive MIMO system," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 525-536, Jan. 2016.

[8] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Proc.*, vol. 62, pp. 3111-3124, Jume 2014.

[9] J.K. Tugnait, "Self-contamination for detection of pilot contamination attack in multiple antenna systems," *IEEE Wireless Communications Letters*, vol. 4, No. 5, pp. 525-528, Oct. 2015.

[10] 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, pp. 248-261, Feb. 2010.

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