SUPERIMPOSED PILOTS: AN ALTERNATIVE PILOT STRUCTURE TO MITIGATE PILOT CONTAMINATION IN MASSIVE MIMO

Karthik Upadhya[⋆] *Sergiy A. Vorobyov*[⋆] *Mikko Vehkapera*[†]

[⋆] Dept. of Signal Processing and Acoustics , Aalto University, Espoo, Finland
[†]EEE Department, University of Sheffield, Sheffield, United Kingdom
Emails:karthik.upadhya@aalto.fi, svor@ieee.org, m.vehkapera@sheffield.ac.uk

ABSTRACT

Superimposed pilots are proposed as an alternative to time-multiplexed pilot and data symbols for mitigating pilot contamination in massive multiple-input multiple-output systems. Provided that the uplink duration is larger than the total number of users in the system, superimposed pilots enable each user to be assigned a unique pilot sequence, thereby allowing for a significant reduction in pilot contamination. Channel estimation performance in the uplink is further improved using an iterative data-aided algorithm. Based on approximate expressions for the uplink signal-to-interference-plus-noise ratio, it is shown that superimposed pilots provide a better performance when compared with methods that use time-multiplexed data and pilots. Numerical simulations are used to validate the approximations and the improved performance of the proposed method.

Index Terms— Massive MIMO, pilot decontamination, superimposed pilots

1. INTRODUCTION

Existing methods to mitigate pilot contamination for massive multiple-input multiple-output (MIMO) [1] systems are designed for the case where the pilots are time-multiplexed with the data, henceforth referred to as conventional time-multiplexed pilots. To mitigate interference, methods that employ conventional time-multiplexed pilots utilize the array gain offered by the large antenna array, asymptotic orthogonality between channel vectors and coordination between base stations (BS) [2–4].

On the other hand, superimposed pilots have been extensively studied for channel estimation in MIMO systems [5–8] in the context of minimizing the loss in data-transmission rate by foregoing the need for separate time slot for pilot-symbol transmission. However, in multi-cell massive MIMO systems, superimposed pilots allow each user in the system to be assigned a unique pilot sequence enabling the BS to estimate the channel vectors of both the desired and interfering users. Most recently, superimposed pilots have been used in the context of multi-cell multiuser MIMO systems [9] without realizing that it provides a superior solution to the pilot decontamination problem in massive MIMO systems.

In this paper, we show that superimposed pilots are superior to conventional time multiplexed pilots in the context of pilot contamination in massive MIMO systems. We propose a novel low-complexity iterative matched-filter based channel estimation scheme for superimposed pilots and demonstrate its effectiveness by deriving a closed-form approximation for the uplink (UL) signal-to-interference-plus-noise ratio (SINR) of the system. Although the use of superimposed pilots requires some coordination between the BSs in assigning pilot sequences to the users and estimating their path-loss coefficients, these are minor impediments compared to the performance improvements provided by the proposed scheme.

2. SYSTEM MODEL

For channel estimation, the massive MIMO uplink with L cells and K users per cell is considered. Each cell has a BS with $M \gg K$ antennas. Assuming that the channel is constant for C_u symbols in the uplink, the matrix of received values $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ at the j^{th} BS can be written as

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

where $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ is the channel vector from the k^{th} user in the ℓ^{th} cell to the j^{th} BS, $\mathbf{x}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ is the sequence of symbols transmitted by the k^{th} user in the ℓ^{th} cell, $(\cdot)^T$ denotes the transpose, $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise at the j^{th} BS with each column distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ and mutually independent of the other columns. The n^{th} element of $\mathbf{h}_{j,\ell,k}$ is given as $[\mathbf{h}_{j,\ell,k}]_n = g_{j,\ell,k,n} \sqrt{\beta_{j,\ell,k}}$, where $g_{j,\ell,k,n} \sim \mathcal{CN}(0, 1)$ and $\beta_{j,\ell,k}$ are the fast-fading and large-scale fading components, respectively. The elements $g_{j,\ell,k,n}$ are mutually independent of each other for all n and $\beta_{j,\ell,k}$ is assumed to be known at the BS. Therefore, the vectors $\mathbf{h}_{j,\ell,k}$ are complex normal distributed and are assumed to be mutually independent of each other $\forall j, \ell, k$.

3. CONVENTIONAL TIME-MULTIPLEXED PILOTS AND THE PILOT CONTAMINATION PROBLEM

With conventional time-multiplexed pilots, each user in a cell transmits a $\tau \geq K$ length orthogonal pilot sequence for channel estimation and the same pilot sequences are transmitted by the users in the $L-1$ interfering cells. The received signal at the j^{th} BS during pilot training $\mathbf{Y}_j^{(p)} \in \mathbb{C}^{M\times\tau}$ can be written as

$$\mathbf{Y}_j^{(p)} = \sum_{\ell=0}^{L-1} \mathbf{H}_{j,\ell}\mathbf{\Phi}^T + \mathbf{W}_j \tag{2}$$

where $\mathbf{H}_{j,\ell} \triangleq [\mathbf{h}_{j,\ell,0}, \ldots, \mathbf{h}_{j,\ell,K-1}]$, $\mathbf{\Phi} \in \mathbb{C}^{\tau\times K}$ is the matrix of pilot sequences with mutually orthogonal columns such that $\mathbf{\Phi}^T\mathbf{\Phi}^* = \tau\mathbf{I}_K$, $(\cdot)^*$ denotes conjugation, and $\mathbf{I}_N$ represents the $N\times N$ identity matrix. The least-squares (LS) estimate of the channel of the m^{th} user in the j^{th} BS can be written as

$$\widehat{\mathbf{h}}_{j,j,m}^{\text{CP}} \triangleq \frac{1}{\tau}\mathbf{Y}_j^{(p)}\mathbf{\Phi}_m^* = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell=0\\ \ell\neq j}}^{L-1} \mathbf{h}_{j,\ell,m} + \frac{1}{\tau}\mathbf{W}_j\mathbf{\Phi}_m^* \tag{3}$$

where $\mathbf{\Phi}_m$ is the m^{th} column of $\mathbf{\Phi}$ and it is the pilot sequence transmitted by the m^{th} user in each cell. It can be seen that the estimates of the channel vectors of the users in the j^{th} cell are contaminated by the channel vectors of the users in the remaining $L-1$ cells. The MSE of the channel estimate of the m^{th} user at the j^{th} BS can be shown to be

$$\text{MSE}_{j,j,m}^{\text{CP}} \triangleq \frac{1}{M}\mathbb{E}\left[\|\mathbf{h}_{j,j,m} - \widehat{\mathbf{h}}_{j,j,m}\|^2\right] = \sum_{\substack{\ell=0\\ \ell\neq j}}^{L-1} \beta_{j,\ell,m} + \frac{\sigma^2}{\tau}\ . \tag{4}$$

4. SUPERIMPOSED PILOTS

When using superimposed pilots, the pilot symbols are transmitted at a reduced power alongside the data symbols for the entire duration of the uplink data slot C_u. If the total number of users in the system is less than the uplink pilot duration, i.e., $KL \leq C_u$, each user can be assigned a unique orthogonal pilot sequence $\mathbf{p}_{j,m} \in \mathbb{C}^{C_u\times 1}$ taken from the columns of an orthogonal matrix $\mathbf{P} \in \mathbb{C}^{C_u\times C_u}$. The received signal at the j^{th} BS $\mathbf{Y}_j \in \mathbb{C}^{M\times C_u}$, when using the superimposed pilot scheme, can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k}\left(\rho_d\mathbf{x}_{\ell,k} + \rho_p\mathbf{p}_{\ell,k}\right)^T + \mathbf{W}_j \tag{5}$$

where ρ_d^2 and ρ_p^2 are the fractions of the transmit powers reserved for the pilot and data symbols, respectively. The total transmitted power p_u is given as $p_u = \rho_p^2 + \rho_d^2$.

4.1. Non-Iterative Channel Estimation

The LS estimate of the channel of the m^{th} user at the j^{th} BS can be written as

$$\widehat{\mathbf{h}}_{j,j,m} \triangleq \frac{1}{C_u\rho_p}\mathbf{Y}_j\mathbf{p}_{j,m}^* = \mathbf{h}_{j,j,m} + \frac{\rho_d}{C_u\rho_p}\mathbf{W}_j\mathbf{p}_{j,m}^* + \frac{\rho_d}{C_u\rho_p}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\mathbf{h}_{j,\ell,k}\mathbf{x}_{\ell,k}^T\mathbf{p}_{j,m}^*\ . \tag{6}$$

The estimated channel can be used to obtain the data vector using matched filtering (MF) as

$$(\widehat{\mathbf{x}}_{j,m})^T = \eta\left\{\frac{1}{M\rho_d}\widehat{\mathbf{h}}_{j,j,m}^H\left(\mathbf{Y}_j - \rho_p\widehat{\mathbf{h}}_{j,j,m}\mathbf{p}_{j,m}^T\right)\right\} \tag{7}$$

where $\eta\{\cdot\}$ is an hard-slicing function that replaces each element of the input vector with the constellation point that is closest in Euclidean distance to that element and $(\cdot)^H$ denotes the Hermitian transpose. The mean-squared error (MSE) of the channel estimate in (6) can be obtained as

$$\text{MSE}_{j,j,m}^{\text{SP}} = \frac{\rho_d^2}{C_u\rho_p^2}\left(\sum_{k=0}^{K-1}\beta_{j,j,k} + \sum_{\substack{\ell=0\\ \ell\neq j}}^{L-1}\sum_{k=0}^{K-1}\beta_{j,\ell,k}\right) + \frac{\sigma^2}{C_u\rho_p^2}\ . \tag{8}$$

The details of all derivations are given in [10, 11] which are our associated journal papers. Large coherence times, and in turn large values of C_u, reduce the MSE in (8) by a significant amount. Moreover, by comparing (4) and (8), it can be seen that in scenarios with little or no pilot contamination, i.e., ($\beta_{j,\ell,k} \approx 0,\ \forall\, \ell \neq j$), the first term in (8), which is due to the overlapping data symbols, dominates and leads to a higher MSE than for the case of time-multiplexed pilots. In addition, since this scheme allows for estimating the channels of the interfering users, iterative data-aided methods can be used to remove interference from both the desired and interfering users.

4.2. Iterative Data-Aided Channel Estimation

For the sake of clarity and without loss of generality, we replace the two indices k, ℓ with a single index k that is used to index the users in all the L cells. Assuming that the path-loss coefficients of all the users are available at the j^{th} BS, they are arranged in the decreasing order, i.e., $\beta_{j,0} \geq \beta_{j,1} \geq \ldots \geq \beta_{j,N-1}$, where $N \triangleq KL$. It is assumed that the BSs have access to the exact values of the path-loss coefficients $\beta_{j,k}$ and that there is no false-ordering. This assumption is reasonable since for large M, the path-loss coefficients can be computed

$$\mathcal{I}^{\mathrm{ul}}_{j,m}(i) \approx \frac{p_u}{M\rho_d^2}\sum_{\substack{k=0\\k\neq m}}^{N-1}\beta_{j,k}\beta_{j,m} + \frac{p_u}{C_u\rho_p^2}\left\{\sum_{\substack{k=0\\k\in\mathcal{U}_j}}^{m-1}\beta_{j,k}^2\alpha_{j,k}^{(i)} + \sum_{\substack{k=m\\k\in\mathcal{U}_j}}^{N-1}\beta_{j,k}^2\alpha_{j,k}^{(i-1)} + \sum_{\substack{k=0\\k\notin\mathcal{U}_j}}^{N-1}\beta_{j,k}^2\right\} + \frac{\sigma^2\beta_{j,m}}{M\rho_d^2}$$

$$+ \frac{p_u}{MC_u\rho_p^2}\left\{\sum_{\substack{k=0\\k\in\mathcal{U}_j}}^{m-1}\sum_{\substack{n=0\\n\neq k}}^{N-1}\beta_{j,n}\beta_{j,k}\alpha_{j,k}^{(i)} + \sum_{\substack{k=m\\k\in\mathcal{U}_j}}^{N-1}\sum_{\substack{n=0\\n\neq k}}^{N-1}\beta_{j,n}\beta_{j,k}\alpha_{j,k}^{(i-1)} + \sum_{\substack{k=0\\k\notin\mathcal{U}_j}}^{N-1}\sum_{\substack{n=0\\n\neq k}}^{N-1}\beta_{j,n}\beta_{j,k} + \sum_{k=0}^{N-1}\beta_{j,k}\sigma^2\right\} \quad (13)$$

---

at the BS with negligible error by averaging the power of the channel coefficients over the entire array.

The channel vector for the m^{th} user can be estimated similar to (6) and can be written as

$$\widehat{\mathbf{h}}_{j,m}^{(1)} = \mathbf{h}_{j,m} + \frac{\rho_d}{C_u\rho_p}\sum_{k=0}^{N-1}\mathbf{h}_{j,k}\mathbf{x}_k^T\mathbf{p}_m^* + \frac{1}{C_u\rho_p}\mathbf{W}_j\mathbf{p}_m^* \quad (9)$$

where the superscript (1) is the iteration index and it denotes the first iteration. The estimate for the m^{th} user is contaminated by the product of the channel and data vectors of all the users projected onto the m^{th} orthogonal pilot. Therefore, to improve the quality of the channel estimate, it is necessary to estimate the channel and data vectors of the remaining $N-1$ users and subtract them from the channel estimate. The decision-aided estimation is started with the 0^{th} user who incidentally has the highest SINR, owing to the ordering of the users. The hard-sliced data from the 0^{th} user is then used to correct the channel estimate of the user with the next highest SINR, which is the 1^{st} user. This can be written as

$$\left(\widetilde{\mathbf{x}}_{j,0}^{(1)}\right)^T = \frac{1}{M\rho_d}\left(\widehat{\mathbf{h}}_{j,0}^{(1)}\right)^H\left(\mathbf{Y}_j - \rho_p\widehat{\mathbf{h}}_{j,0}^{(1)}\mathbf{p}_0^T\right) \quad (10)$$

$$\widehat{\mathbf{x}}_{j,0}^{(1)} = \eta\left(\widetilde{\mathbf{x}}_{j,0}^{(1)}\right) \quad (11)$$

$$\widehat{\mathbf{h}}_{j,1}^{(1)} = \frac{1}{C_u\rho_p}\left(\mathbf{Y}_j - \rho_d\widehat{\mathbf{h}}_{j,0}^{(1)}\left(\widehat{\mathbf{x}}_{j,0}^{(1)}\right)^T\right)\mathbf{p}_1^* \quad (12)$$

where the subscript j in $\widehat{\mathbf{x}}_{j,0}$ and $\widetilde{\mathbf{x}}_{j,0}^{(1)}$ indicates that these estimates are computed at the j^{th} BS. To prevent error propagation, it is necessary to exclude users with poor SINR from the feedback loop. This exclusion can be accomplished by analyzing the impact of symbol errors on the MSE.

To derive the MSE, let $\mathcal{U}_j$ be the set of users whose estimated data is fed back to improve the channel estimate. It is assumed that the estimation errors of both the channel and data vectors in the feedback loop are uncorrelated in every iteration and that the estimation error in each element of $\widetilde{\mathbf{x}}_{j,m}$ is zero-mean circular complex-Gaussian with variance equal to the uplink interference power $\mathcal{I}^{\mathrm{ul}}_{j,m}(i)$. The expression for $\mathcal{I}^{\mathrm{ul}}_{j,m}(i)$, whose derivation is omitted here and will appear in [10], is given in (13) at the top of the page. The parameter $\alpha_{j,m}^{(i)}$ in (13) is the variance of each element of the symbol error vector $\Delta\mathbf{x}_{j,m}^{(i)} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_{j,m}^{(i)}$. For example, when $\mathbf{x}_m$ takes values from a quadrature phase-shift keying (QPSK) constellation, $\alpha_{j,m}^{(i)}$ is given as

$$\alpha_{j,m}^{(i)} = \begin{cases}\mathbb{E}\left\{\left|\left[\Delta\mathbf{x}_{j,m}^{(i)}\right]_n\right|^2\right\} = 4Q\left(\frac{\beta_{j,m}}{\sqrt{\mathcal{I}^{\mathrm{ul}}_{j,m}(i)}}\right), & i \geq 1\\ 1, & i = 0\end{cases} \quad (14)$$

where $Q(x)$ is the Q-function. Using (13) and (14), the MSE (derived in [11]) can be written as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP-iter}}(i) \approx \frac{1}{C_u\rho_p^2}\left[\rho_d^2\sum_{\substack{k=0\\k\in\mathcal{U}_j}}^{m-1}\beta_{j,k}\alpha_{j,k}^{(i)} + \rho_d^2\sum_{\substack{k=m\\k\in\mathcal{U}_j}}^{N-1}\beta_{j,k}\alpha_{j,k}^{(i-1)} + \rho_d^2\sum_{\substack{k=0\\k\notin\mathcal{U}_j}}^{N-1}\beta_{j,k} + \sigma^2\right]. \quad (15)$$

Let us now assume that the decoded data vector of only the m^{th} user is used in the feedback loop, so that the MSE after the first iteration is obtained from (15) by setting $\mathcal{U}_j = \{m\}$ as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP-iter}}(2) = \frac{\rho_d^2}{C_u\rho_p^2}\left(\beta_{j,m}\alpha_{j,m}^{(1)} + \sum_{\substack{k=0\\k\neq m}}^{N-1}\beta_{j,k}\right) + \frac{\sigma^2}{C_u\rho_p^2}. \quad (16)$$

On the other hand, the MSE for the non-iterative scheme, given by equation (8), can be rewritten as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP}} = \frac{\rho_d^2}{C_u\rho_p^2}\left(\sum_{k=0}^{N-1}\beta_{j,k}\right) + \frac{\sigma^2}{C_u\rho_p^2}. \quad (17)$$

Then, a reasonable approach for the inclusion of the m^{th} user in the feedback loop would be to include the user only if it does not increase the MSE. The impact of including the m^{th} user in the feedback loop is clarified by comparing (16) and (17). Therefore, with $\mathcal{U}_j = \{m\}$, the estimated data of the m^{th} user is included in the feedback loop if

$$4Q\left(\frac{\beta_{j,m}}{\sqrt{\mathcal{I}^{(\mathrm{ul})}_{j,m}(1)}}\right) < 1. \quad (18)$$

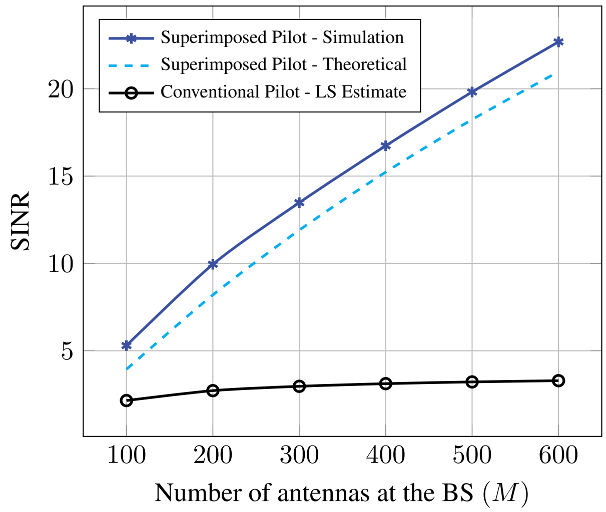

Fig. 1. Comparison of the SINR of a user in the reference BS for the LS-based and the proposed schemes at 10dB SNR.

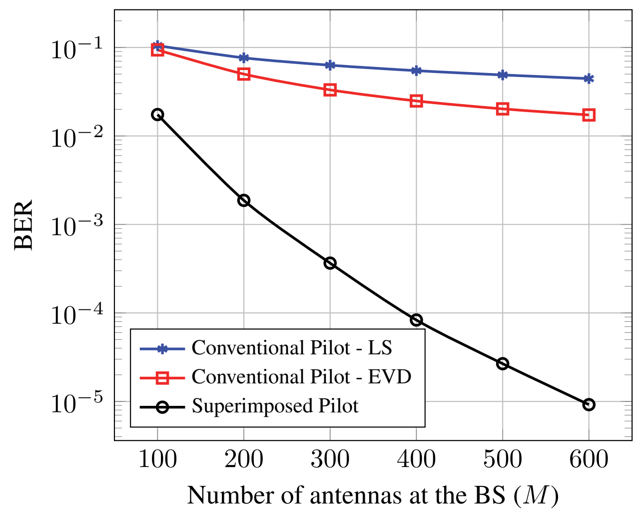

Fig. 2. Comparison of the BER for the EVD-based, the LS-based, and the proposed scheme at 10dB SNR.

If γ_j is a threshold on $\beta_{j,m}$ obtained from (18) and if $\mathcal{U}_j \triangleq \{m \,|\, \beta_{j,m} > \gamma_j\}$, then the estimate of the channel and data vectors for the m^{th} user at the i^{th} iteration can be written as

$$\widehat{\mathbf{h}}_{j,m}^{(i)} = \frac{1}{C_u \rho_p} \Bigg[\mathbf{Y}_j - \rho_d \sum_{\substack{k=0 \\ k \in \mathcal{U}_j}}^{m-1} \widehat{\mathbf{h}}_{j,k}^{(i)} \left(\widehat{\mathbf{x}}_k^{(i)} \right)^T - \rho_d \sum_{\substack{k=m \\ k \in \mathcal{U}_j}}^{N-1} \widehat{\mathbf{h}}_{j,k}^{(i-1)} \left(\widehat{\mathbf{x}}_k^{(i-1)} \right)^T \Bigg] \mathbf{p}_{j,m}^* \quad (19)$$

$$\left(\widetilde{\mathbf{x}}_{j,m}^{(i)} \right)^T = \frac{1}{M \rho_d} \left(\widehat{\mathbf{h}}_{j,m}^{(i)} \right)^H \left(\mathbf{Y}_j - \rho_p \widehat{\mathbf{h}}_{j,m}^{(i)} \mathbf{p}_{j,m}^T \right) \quad (20)$$

$$\widehat{\mathbf{x}}_{j,m}^{(i)} = \eta \left(\widetilde{\mathbf{x}}_{j,m}^{(i)} \right) \quad (21)$$

where $\widehat{\mathbf{h}}_{j,m}^{(0)} = \mathbf{0}, \widehat{\mathbf{x}}_{j,m}^{(0)} = \mathbf{0} \;\forall\, m = 0, \ldots, N-1$. It has to be noted that the threshold is computed in the first iteration and is fixed for the subsequent iterations. Ideally, the threshold should decrease at each iteration since the improvement in the channel estimate would allow users with lower SINR to be included in $\mathcal{U}_j$. However, this is not feasible since the expression for $\mathcal{I}_{j,m}^{\text{ul}}(i)$ is only an approximation and a conservative threshold is used instead. The SINR for the iterative scheme, given the above assumptions, is written then as

$$\text{SINR}_{j,m}^{\text{SP}-\text{ul}}(i) = \frac{\beta_{j,m}^2}{\mathcal{I}_{j,m}^{\text{ul}}(i)} \quad (22)$$

5. SIMULATION RESULTS

The bit-error rate (BER) and SINR of the proposed superimposed pilot scheme is compared with that of the LS-based channel estimate [1] and the eigenvalue decomposition (EVD)-based method [3] that use time-multiplexed pilots. The methods are tested with $L = 7$ cells and $K = 5$ users per cell. The users are assumed to be spaced at equal intervals on a circle of radius 800m from the BS in hexagonal cells with radius 1km. The path-loss coefficient is assumed to be 3 and C_u is assumed to be 100 symbols. The power of the transmitted symbols p_u is set to 1 and for simplicity, $\rho_p = \rho_d = 1/\sqrt{2}$. The simulations for the proposed superimposed pilot-based method are performed for 4 iterations.

It can be seen from the SINR plot presented in Fig. 1 that as M increases, the SINR saturates for the conventional-pilot scheme due to pilot contamination. However, the SINR increases almost linearly in M for the proposed superimposed pilots scheme and this trajectory can be potentially maintained using techniques such as adaptive modulation and coding, thereby implying that the pilot contamination effect can be eliminated. From Fig. 2, it can be seen that the proposed pilot structure and the iterative channel estimation algorithm offer a significant improvement in BER over both the LS-based and EVD-based schemes.

6. CONCLUSION

We have proposed using superimposed pilots as a superior alternative to time-multiplexed data and pilots for uplink channel estimation in massive MIMO systems. In addition, an iterative data-aided channel estimation scheme is developed. This scheme uses data symbols from both the desired and interfering users in the feedback loop, provided the SINR of these users exceeds a threshold. Theoretical expressions and numerical simulations show that the proposed method significantly alleviates the pilot contamination problem.

7. REFERENCES

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

[2] 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. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Mar. 2012, pp. 3249–3252.

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

[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.

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

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

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

[8] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommun. Conf.*, St. Louis, Mo, Dec. 2005, pp. 2229–2233.

[9] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol., to be published*.

[10] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO systems part I: Theory and channel estimation," *IEEE Trans. Signal Process.*, Submitted, Nov. 2015.

[11] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO systems part II: Performance analysis and optimization," *IEEE Trans. Signal Process.*, Submitted, Nov. 2015.