DISTRIBUTIVE ESTIMATION OF FREQUENCY SELECTIVE CHANNELS FOR MASSIVE MIMO SYSTEMS

Alam Zaib[*], *Mudassir Masood*[†], *Mounir Ghogho*[‡] *and Tareq Y. Al-Naffouri*[*†]

[*†] Electrical Engineering Department
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia

[†] Department of Electrical Engineering
King Abdullah University of Science & Technology
Thuwal 23955-6900, KSA

[‡] School of Electronics & Electrical Engineering
University of Leeds
LS2 9JT, UK

ABSTRACT

We consider frequency selective channel estimation in the uplink of massive MIMO-OFDM systems, where our major concern is complexity. A low complexity distributed LMMSE algorithm is proposed that attains near optimal channel impulse response (CIR) estimates from noisy observations at receive antenna array. In proposed method, every antenna estimates the CIRs of its neighborhood followed by recursive sharing of estimates with immediate neighbors. At each step, every antenna calculates the weighted average of shared estimates which converges to near optimal LMMSE solution. The simulation results validate the near optimal performance of proposed algorithm in terms of mean square error (MSE).

Index Terms— Channel estimation, massive MIMO, Least squares, LMMSE, distributed estimation

1. INTRODUCTION

Next generation wireless communication system require considerable data throughput and strong resilience against multipath fading. Massive MIMO systems by virtue of utilizing very large antenna arrays, typically of the order of few hundreds, at the base station (BS) can potentially provide huge gains in system throughput, energy efficiency, security and robustness of wireless systems [1–3]. Because of these vital advantages, massive MIMO has attracted a lot of research interests and also envisioned as an enabling technology for next generation (5G) broadband wireless communications [4].

One of the bottleneck in achieving full advantages of massive MIMO is the accurate estimation of CIR between each transmit-receive antenna pair. The LMMSE being an optimal estimator in the presence of additive white Gaussian noise (AWGN), has been extensively studied for MIMO-OFDM systems [5–7]. Unlike least squares (LS) or interpolation based techniques, LMMSE can exploit the additional statistical information to deliver improved performance. However, LMMSE is impractical for massive MIMO as it entails huge complexity due to matrix inversion of very large dimensionality. Some ways to reduce the complexity of LMMSE have been proposed e.g [8, 9]. Other methods showing increased interest in channel estimation for massive MIMO include [10, 11]. However, the existing methods mostly deal with single-carrier flat-fading channels or have an underlying assumption that the channel must be sparse.

Inspired by our previous work [12], where channels were assumed to be sparse, we propose a distributed channel estimation algorithm for massive MIMO under more realistic assumption of correlated Rayleigh fading CIRs. Although distributed estimation has been studied in various contexts (see [13, 14] and references there in), to the best of our knowledge, it is not yet investigated in massive MIMO.

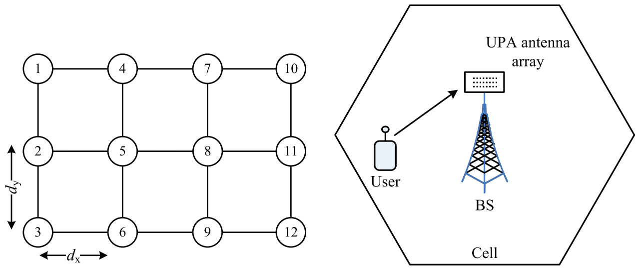
The remainder of the paper is organized as follows. Section 2 describes the system and channel model. Section 3 presents different MMSE based channel estimation and discuss their limitations for massive MIMO. The proposed algorithm is presented in Section 4. Simulation results are presented in Section 5 and we finally conclude in Section 6.

Notations: The lower case letters x and lower case boldface $\mathbf{x}$ represent the scaler and the (column) vector respectively while $x(i)$ denotes individual entries of vector $\mathbf{x}$. Matrices are denoted by upper boldface letters $\mathbf{X}$ whereas the calligraphic notation $\boldsymbol{\mathcal{X}}$ is reserved for vectors in the frequency domain. $(.)^{\mathrm{T}}$, $(.)^{*}$ and $(.)^{\mathrm{H}}$ represent transpose, conjugate and conjugate transpose (hermitian) operations respectively. $E\{.\}$ denotes the expectation and the weighted norm of a vector $\mathbf{x}$ is represented by $\|\mathbf{x}\|^2_{\mathbf{A}} \triangleq \mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x}$.

2. SYSTEM MODEL

Consider a multi-user massive MIMO wireless system where the base station (BS) in each cell is equipped with a uniform

This work was funded by a CRG3 grant ORS#2221 from the Office of Competitive Research (OCRF) at King Abdullah University of Science and Technology (KAUST), Saudi Arabia.



(a) (b)

Fig. 1: (a) UPA structure for $M_y = 3$ and $N_x = 4$ with antenna indexing (b) Single-cell system layout

planar array (UPA) consisting of large number of antennas, while each user terminal has a single antenna. The BS antennas are distributed across M_y rows and N_x columns of $N_p = M_y \times N_x$ elements with horizontal and vertical spacing of d_x and d_y respectively. We define the (m, n)th antenna as an element in mth row and nth column of the array with index $r = m + M_y(n-1)$ where $1 \leq m \leq M_y, 1 \leq n \leq N_x$ and $1 \leq r \leq N_p$. Fig. 1(a) shows an example of 3×4 UPA structure with antenna indexing.

For uplink channel estimation, we assume that all users in a particular cell are assigned orthogonal OFDM tones so that there is no intra-cell interference. Moreover, we do not address pilot contamination that may result from reusing of pilots in the neighboring cells [1]. Hence for the sake of channel estimation, the discussion that follows assumes a single cell with just one user and a BS as shown in Fig. 1(b).

A. Channel Model

The frequency selective channel between a user and receive antenna r is modeled by gaussian L-tap channel impulse response (CIR) vector $\mathbf{h}_r \triangleq [h_r(0), h_r(1), \cdots, h_r(L-1)]^{\mathrm{T}}$. We collect the lth tap of all transmit-receive pairs to form an N_p dimensional tap vector $\mathbf{h}^{(l)} \triangleq \left[h_1(l), h_2(l), \cdots, h_{N_p}(l)\right]^{\mathrm{T}}$ and let the N_pL dimensional composite channel vector be defined as $\mathbf{h} \triangleq \left[\mathbf{h}_1^{\mathrm{T}}, \mathbf{h}_2^{\mathrm{T}}, \cdots, \mathbf{h}_{Np}^{\mathrm{T}}\right]^{\mathrm{T}}$. Then, the $N_pL \times N_pL$ dimensional (composite) channel correlation matrix is,

$$\mathbf{R_h} \triangleq E\{\mathbf{h}\mathbf{h}^{\mathrm{H}}\} = \mathbf{R}_s \otimes \mathbf{R}_{tap} \tag{1}$$

where, $\mathbf{R}_s = E\{\mathbf{h}^{(l)}\mathbf{h}^{(l)\mathrm{H}}\}, \forall l$ represents an $N_p \times N_p$ dimensional antenna spatial correlation matrix, $\mathbf{R}_{tap} = E\{\mathbf{h}_r\mathbf{h}_r^{\mathrm{H}}\}, \forall r$ represents an $L \times L$ dimensional channel tap correlation matrix that depends on channel power delay profile (PDP), assumed to be identical for all channels and $\otimes$ represents the Kronecker product. For the spatial correlation matrix, we shall adopt the ray-based 3D channel model from [15] which is well suited for rectangular arrays.

B. Signal model

We assume that there are N OFDM sub-carriers and let $\boldsymbol{\mathcal{X}}$ represent the N-dimensional information symbols of the user. The equivalent time-domain symbols are obtained by taking inverse Fourier transform i.e. $\mathbf{x} = \mathbf{F}^{\mathrm{H}}\boldsymbol{\mathcal{X}}$, where $\mathbf{F}$ is an $N \times N$ unitary DFT matrix whose (l, k) entry is $f_{l,k} = \frac{1}{\sqrt{N}}e^{-j2\pi lk/N}$. Hence, the N dimensional frequency domain OFDM symbol vector received at rth BS is given by,

$$\boldsymbol{\mathcal{Y}}_r = \mathrm{diag}(\boldsymbol{\mathcal{X}})\underline{\mathbf{F}}\mathbf{h}_r + \boldsymbol{\mathcal{W}}_r = \mathbf{A}\mathbf{h}_r + \boldsymbol{\mathcal{W}}_r \tag{2}$$

where $\mathbf{A} \triangleq \mathrm{diag}(\boldsymbol{\mathcal{X}})\underline{\mathbf{F}}$, $\boldsymbol{\mathcal{W}}_r$ is complex AWGN noise vector with *pdf:* $\mathcal{CN}\left(\mathbf{0}, \sigma_w^2\mathbf{I}_N\right)$ assumed uncorrelated with the channel and $\underline{\mathbf{F}}$ is truncated Fourier matrix formed by selecting first L columns of $\mathbf{F}$. Let K sub-carriers are reserved for pilots and the remaining $N-K$ for the data transmission, then for a set of pilot indices $\mathcal{P}$, equation (2) reduces to,

$$\boldsymbol{\mathcal{Y}}_r(\mathcal{P}) = \mathbf{A}(\mathcal{P})\mathbf{h}_r + \boldsymbol{\mathcal{W}}_r(\mathcal{P}) \tag{3}$$

where $r = 1, \cdots, N_p$. These N_p systems of equations can be combined by stacking all the observations to get,

$$\boldsymbol{\mathcal{Y}}(\mathcal{P}) = [\mathbf{I}_{N_p} \otimes \mathbf{A}(\mathcal{P})]\mathbf{h} + \boldsymbol{\mathcal{W}}(\mathcal{P}) \tag{4}$$

where, $\boldsymbol{\mathcal{Y}}(\mathcal{P}) = \left[\boldsymbol{\mathcal{Y}}_1(\mathcal{P})^{\mathrm{T}}, \cdots, \boldsymbol{\mathcal{Y}}_{N_p}(\mathcal{P})^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{\mathcal{W}}(\mathcal{P}) = \left[\boldsymbol{\mathcal{W}}_1(\mathcal{P})^{\mathrm{T}}, \cdots, \boldsymbol{\mathcal{W}}_{N_p}(\mathcal{P})^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\mathbf{I}_N$ represents an $N \times N$ identity matrix. Note that the number of unknown channel coefficients in (4) are N_pL whereas the total number of equations are N_pK. In the following we assume that the necessary condition $L \leq K \ll N$, to solve (3) and (4) is satisfied.

3. LMMSE BASED CHANNEL ESTIMATION

We pursue different LMMSE techniques that can be adopted for estimation of CIRs between user and each BS antenna.

A. The Localized LMMSE (L-LMMSE) estimation

In this approach all CIRs are estimated independently based on the observations received at each antenna element. Thus dropping the index vector $\mathcal{P}$ for convenience, the classical LMMSE estimate of $\mathbf{h}_r$ from (3), is obtained by [16],

$$\hat{\mathbf{h}}_r^{(local)} = \left[\mathbf{R}_{tap}^{-1} + \frac{1}{\sigma_w^2}\mathbf{A}^{\mathrm{H}}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{R}_w^{-1}\boldsymbol{\mathcal{Y}}_r \tag{5}$$

The error covariance matrix of rth channel vector is given by,

$$\mathbf{C}_e^r = \left[\mathbf{R}_{tap}^{-1} + \frac{1}{\sigma_w^2}\mathbf{A}^{\mathrm{H}}\mathbf{A}\right]^{-1} \tag{6}$$

The MSE at rth antenna is trace of (6) and the global MSE can be obtained by $\mathrm{MSE}^{(local)} = \sum_{r=1}^{N_p} \mathrm{trace}\left(\mathbf{C}_e^r\right)$. Obviously, the performance of L-LMMSE is not optimal in the sense of minimizing the global MSE but has low complexity of the order $O\left(N_pL^3\right)$, which increases linearly with number of BS antennas.

For massive MIMO with extremely large number of antennas located in close proximity at the BS, the channels are highly likely to be correlated. The L-LMMSE is unable to exploit these spatial correlations among array elements.

B. The Optimal LMMSE (O-LMMSE) Solution

The optimal estimation strategy is to minimize the overall or the global MSE i.e. $E\{\|\mathbf{h}-\hat{\mathbf{h}}\|^2\}$ based on all observations $\boldsymbol{\mathcal{Y}}$ as given in (4). The solution is obtained by

$$\hat{\mathbf{h}}^{(opt)} = \left[\mathbf{R}_{\mathbf{h}}^{-1} + \frac{1}{\sigma_w^2}\acute{\mathbf{A}}^{\mathrm{H}}\acute{\mathbf{A}}\right]^{-1}\acute{\mathbf{A}}^{\mathrm{H}}\mathbf{R}_w^{-1}\boldsymbol{\mathcal{Y}} \tag{7}$$

where, $\acute{\mathbf{A}} = \mathbf{I}_{N_p} \otimes \mathbf{A}$, $\mathbf{R}_{\mathbf{h}}$ is given in (1) and for convenience we continue to omit the index vector $\mathcal{P}$. The corresponding error covariance matrix in this case is

$$\mathbf{C}_e^{(opt)} = E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathrm{H}}\} = \left[\mathbf{R}_{\mathbf{h}}^{-1} + \frac{1}{\sigma_w^2}\acute{\mathbf{A}}^{\mathrm{H}}\acute{\mathbf{A}}\right]^{-1} \tag{8}$$

where $\tilde{\mathbf{h}}$ represents the estimation error vector. The (global) MSE is $\mathrm{MSE}^{(opt)} = \mathrm{trace}\left(\mathbf{C}_e^{opt}\right)$, where it is obvious that $\mathrm{MSE}^{(opt)} \leq \mathrm{MSE}^{(local)}$, i.e. the optimal solution yields better MSE performance than the localized one by utilizing the spatially correlated observations of each antenna. However, it has two major drawbacks: (i) Realization of optimal strategy requires the global sharing of information between each antenna element and a central processor which results in communication overhead, (ii) The computational complexity of optimal LMMSE from (7) is of the order $O\left(N_p^3 L^3\right)$. In massive MIMO scenario, where N_p is of order of few hundreds, both of these operations are challenging and impractical. The shortcomings of localized and centralized solutions motivate us to propose a low complexity distributed LMMSE estimation as described next.

4. THE DISTRIBUTED LMMSE ESTIMATION

We aim to solve the optimal LMMSE problem in a distributed manner where the information processing takes place locally at each antenna node and at the same time the nodes can collaborate with each other. The proposed distributed LMMSE (D-LMMSE) estimator is composed of two steps; the estimation step and the sharing step as described below.

A. Estimation Step

In this step, each antenna acting as a center node r_C, estimates not only its own CIR but also the CIRs of its neighborhood. The neighborhood of r_C consists of 4-direct neighbors (see Fig. 2) represented by the set $\mathcal{N} = \{r_L, r_R, r_U, r_D\}$. For elements lying at the edges of array the number of neighbors are $2 \leq |\mathcal{N}| \leq 4$. The set $r_C \cup \mathcal{N}$ is denoted by $\mathcal{N}^+$. Also, let the corresponding channel vectors be represented by $\mathbf{h}_C$, $\mathbf{h}_L$, $\mathbf{h}_R$, $\mathbf{h}_U$ and $\mathbf{h}_D$ respectively and $\mathbf{h}^c$ be the $|\mathcal{N}^+|L \times 1$ dimensional composite channel vector of the central node

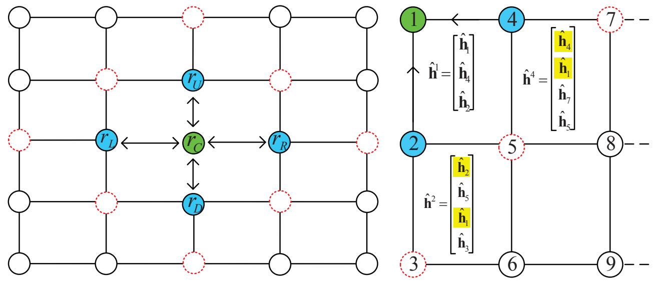


Fig. 2: (Left) Information diffusion process : During first iteration r_C (green) receives information from its 4-direct neighbors (blue). In the second iteration, the information from next nearest neighbors (red) also comes in and so on. (Right) Sharing details in a 3×4 antenna array where the neighboring elements (indices 4 and 2) share only the partial information (highlighted in yellow) with the central node (index 1).

and its $|\mathcal{N}|$ direct neighbors. In estimation step, the centre node obtains the estimate of $\mathbf{h}^c$ by solving the weighted least squares (WLS) problem,

$$\hat{\mathbf{h}}^c = \underbrace{\min}_{\mathbf{h}^c} \|\boldsymbol{\mathcal{Y}}_C(\mathcal{P}) - \mathbf{A}(\mathcal{P})\mathbf{h}_C\|^2_{\mathbf{R}_w^{-1}} + \|\mathbf{h}^c\|^2_{\mathbf{R}_{\mathbf{h}^c}^{-1}} \tag{9}$$

where $\mathbf{R}_{\mathbf{h}^c} \triangleq E\{\mathbf{h}^c(\mathbf{h}^c)^{\mathrm{H}}\}$. The solution to above minimization problem is,

$$\hat{\mathbf{h}}^c = \left[\mathbf{R}_{\mathbf{h}^c}^{-1} + \frac{1}{\sigma_w^2}\bar{\mathbf{A}}^{\mathrm{H}}\bar{\mathbf{A}}\right]^{-1}\bar{\mathbf{A}}^{\mathrm{H}}\mathbf{R}_w^{-1}\boldsymbol{\mathcal{Y}}_C(\mathcal{P}) \tag{10}$$

where, $\bar{\mathbf{A}} = \begin{bmatrix}\mathbf{A}(\mathcal{P}) & \mathbf{0}_{K\times|\mathcal{N}|L}\end{bmatrix}$ as each antenna exploits only its own observations. Likewise, the center node also computes the error covariance matrix,

$$\mathbf{C}_e^c = \left[\mathbf{R}_{\mathbf{h}^c}^{-1} + \frac{1}{\sigma_w^2}\bar{\mathbf{A}}^{\mathrm{H}}\bar{\mathbf{A}}\right]^{-1} \tag{11}$$

which can be computed off-line. Observe that the complexity for computing above estimates and error covariance matrix is still of the order $O\left(N_p L^3\right)$ for all antennas. It is assumed that the center node r_C has available correlation information of its neighborhood to construct $\mathbf{R}_{\mathbf{h}^c}$. From (11) this also implies that the center node has available error covariance matrices of its neighbors. Having found $|\mathcal{N}^+|$ estimates, each antenna is ready to initiate the sharing step.

B. Sharing Step

The sharing step is the key to our distributed algorithm where the antennas collaborate locally to share their estimates with the neighbors. The information sharing is done only partially such that each antenna transmits selected components i.e., the estimate of his own and that of the neighboring channel. An example of how this sharing takes place is detailed in Fig. 2 for a 3×4 array with central element $r_C = 1$ having two neighbors; $\mathcal{N} = \{r_R = 4, r_D = 2\}$. The composite vector of the central node and those received (shown underlined) from its neighbors are explicitly given by,

$$\hat{\mathbf{h}}^1 = \begin{bmatrix}\hat{\mathbf{h}}_1\\ \hat{\mathbf{h}}_4\\ \hat{\mathbf{h}}_2\end{bmatrix}, \ \underline{\hat{\mathbf{h}}}^4 = \begin{bmatrix}\hat{\mathbf{h}}_1\\ \hat{\mathbf{h}}_4\\ \mathbf{0}\end{bmatrix} \leftarrow \hat{\mathbf{h}}^4, \ \underline{\hat{\mathbf{h}}}^2 = \begin{bmatrix}\hat{\mathbf{h}}_1\\ \mathbf{0}\\ \hat{\mathbf{h}}_2\end{bmatrix} \leftarrow \hat{\mathbf{h}}^2 \quad (12)$$

where we refer to the component $\hat{\mathbf{h}}_j$ of $\hat{\mathbf{h}}^k$ as the estimate of node j computed by node k and the null entries correspond to the components which are not shared by the neighbors.

Update: As a result of sharing, the central node receives information from its $|\mathcal{N}|$-direct neighbors and consequently updates its estimate and the error covariance matrix by optimally combining them as follows [16] ,

$$\mathbf{C}_e^{-1c(i+1)}\hat{\mathbf{h}}^{c(i+1)} = \mathbf{C}_e^{-1c(i)}\hat{\mathbf{h}}^{c(i)} + \sum_{r\in\mathcal{N}} \underline{\mathbf{C}}_e^{-1r(i)}\underline{\hat{\mathbf{h}}}^{r(i)} \quad (13)$$

$$\mathbf{C}_e^{-1c(i+1)} = \mathbf{C}_e^{-1c(i)} + \sum_{r\in\mathcal{N}} \left(\underline{\mathbf{C}}_e^{-1r(i)} - \underline{\mathbf{R}}_{\mathbf{h}^r}^{-1}\right) \quad (14)$$

where, i is the iteration index. Since the center node has available correlation and error covariances, it can construct the required (i.e. underlined) matrices corresponding to the shared estimates for each of its neighbour. Observe from (13) that the reliable estimates are weighted more than the unreliable ones, hence the null block entries of resulting error covariance and correlation matrices should be assigned as $\beta\mathbf{I}$ and $\mathbf{I}$ respectively, where β is an arbitrarily large number. The complete procedure is summarized in Algorithm 1.

Algorithm 1 distributed LMMSE (D-LMMSE) algorithm

For each antenna acting as a central node r_C repeat

1. Compute $\hat{\mathbf{h}}^c$ and $\mathbf{C}_e^c$ by using (10) and (11).
2. Share estimates with $|\mathcal{N}|$ neighbors as explained above.
3. Using steps 1&2, construct $\{\underline{\mathbf{R}}_{\mathbf{h}^r}\}_{r=1}^{|\mathcal{N}|}$ and $\{\underline{\mathbf{C}}_e^r\}_{r=1}^{|\mathcal{N}|}$.
4. Update $\hat{\mathbf{h}}^c$ and $\mathbf{C}_e^c$ using (13) -(14).
5. Repeat steps 2-4, D times where D is the maximum number of iterations and then output $\hat{\mathbf{h}}_C = \hat{\mathbf{h}}^c(1:L)$.

Remarks: The recursive sharing enables each node in the array to utilize the correlated observations from distant nodes. A simple loose upper bound on maximum number of iterations is given by $D \leq \sqrt{N_p/2 - 1/4} - 1/2$, which ensures that each antenna receives information from every other antenna in the array. The actual value of D also depends on the degree of correlations among antennas and is far less than given by the bound (see Fig. 3, where convergence is achieved in 2-3 iterations). Moreover, the proposed algorithm has the advantage of low communication and low computation at each node as compared to the optimal solution.

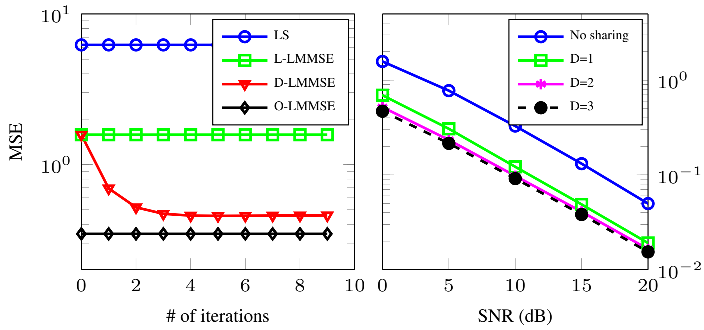


Fig. 3: Convergence of distributed algorithm.

5. SIMULATION RESULTS

We adopt the channel model in (1) with symmetric spatial correlation matrix of [15] whose parameters are $\phi = \pi/3$, $\theta = 3\pi/8$, $\sigma = \pi/12$ and $\xi = \pi/36$ which represent mean horizontal angle of departure (AoD), mean vertical AoD, standard deviation of horizontal and vertical AoD respectively. For the channel taps, we use an exponentially decaying PDP $E\{|h_r(\tau)|^2\} = e^{-\tau}$. Other parameters are described in the Table 1, where λ represents the carrier frequency wave length.

Parameter	**Value**
Array dimensions ($M_y \times N_x$)	5×5
Array element spacing d_x, d_y	0.3λ, 0.5λ
Number of OFDM sub-carriers (N)	256
Number of pilots (K)	32
Signal constellation modulation	4-QAM
Channel length (L)	8

Table 1: Parameters for simulation

The performance of different algorithms namely, the localized, the optimal, distributed LMMSE and LS is assessed by computing the MSE performance criteria: $\text{MSE} = \frac{1}{\nu}\sum_{i=1}^{\nu} \|\mathbf{h}^i - \hat{\mathbf{h}}^i\|^2$ where, $\mathbf{h}^i$ and $\hat{\mathbf{h}}^i$ are true and estimated CIR vectors of size $N_pL \times 1$ at i^{th} trial and ν represent total number of trials. We used ν =100 in our simulations.

To show convergence of distributed algorithm, the MSE is plotted against (i) the number of iterations D at fixed SNR of 0 dB and (ii) both SNR and D in Fig. 3. As can be seen, the MSE of proposed algorithm decreases exponentially during each step of iteration because each node utilizes correlated observations from distant nodes to improve its estimate. Moreover, the algorithm converges to near optimal solution in just 2-3 iterations.

The MSE performance of different algorithms is shown in Fig. 4 over the SNR range of 0-20 dB. It is clear that O-LMMSE is better than L-LMMSE as the later doesn't make

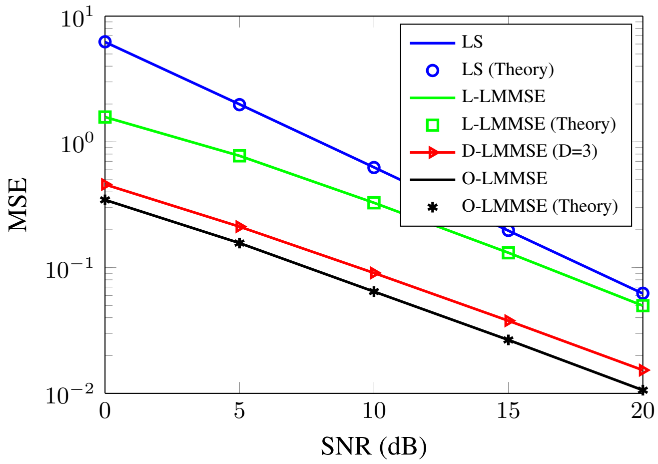


Fig. 4: MSE performance of different algorithms.

use of channel spatial correlations. However, the proposed algorithm is able to achieve near optimal performance.

6. CONCLUSIONS

Channel estimation is very crucial in massive MIMO systems where conventional techniques of MIMO systems cannot be employed due to several limitations. We proposed a distributed LMMSE algorithm for uplink channel estimation. By relying on coordination among neighboring antennas, the proposed algorithm is very much tractable and attains near optimal solution at significantly low complexity.

REFERENCES

[1] T.L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[2] F. Rusek, D. Persson, Buon Kiong Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 160–171, February 2013.

[4] F Boccardi, R.W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, February 2014.

[5] J.-J. van de Beek, O. Edfors, M. Sandell, S.K. Wilson, and P. Ola Borjesson, "On channel estimation in OFDM systems," in *Vehicular Technology Conference, 1995 IEEE 45th*, Jul 1995, vol. 2, pp. 815–819 vol.2.

[6] O. Edfors, M. Sandell, J.-J. van de Beek, S.K. Wilson, and P.O. Borjesson, "OFDM channel estimation by singular value decomposition," in *Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th*, Apr 1996, vol. 2, pp. 923–927 vol.2.

[7] M.K. Ozdemir, H. Arslan, and E. Arvas, "Toward real-time adaptive low-rank LMMSE channel estimation of MIMO-OFDM systems," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 10, pp. 2675–2678, Oct 2006.

[8] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debbah, "Low-complexity channel estimation in large-scale MIMO using polynomial expansion," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, Sept 2013, pp. 1157–1162.

[9] P. Xu, J. Wang, J. Wang, and F. Qi, "Analysis and design of channel estimation in multicell multiuser mimo ofdm systems," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–11, 2014.

[10] Hien Quoc Ngo and E.G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3249–3252.

[11] Sinh Le Hong Nguyen and Ali Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 2890–2895.

[12] M. Masood, L.H. Afify, and T.Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive mimo," *Signal Processing, IEEE Transactions on*, vol. 63, no. 1, pp. 104–118, Jan 2015.

[13] I.D. Schizas and G.B. Giannakis, "Consensus-based distributed estimation of random signals with wireless sensor networks," in *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, Oct 2006, pp. 530–534.

[14] Ali H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[15] Dawei Ying, F.W. Vook, T.A. Thomas, D.J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3d channel model," in *Communications (ICC), 2014 IEEE International Conference on*, June 2014, pp. 5865–5870.

[16] A.H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.