Enhanced Sparse Bayesian Learning-based Channel Estimation for Massive MIMO-OFDM Systems

Hayder AL-Salihi*, Mohammad Reza Nakhai* and Tuan Anh Le+
*Centre for Telecommunications Research, Department of Informatics, King's College London, London, UK
+Department of Design Engineering and Mathematics, Middlesex University, London, UK
Email: ana.al-salihi@kcl.ac.uk

***Abstract*—Pilot contamination limits the potential benefits of massive multiple input multiple output (MIMO) systems. To mitigate pilot contamination, in this paper, an efficient channel estimation approach is proposed for massive MIMO systems, using sparse Bayesian learning (SBL) namely coupled hierarchical Gaussian framework where the sparsity of each coefficient is controlled by its own hyperparameter and the hyperparameters of its immediate neighbours. The simulation results show that the proposed method can reconstruct original channel coefficients more effectively compared to the conventional channel estimators in terms of channel estimation accuracy in the presence of pilot contamination.**

***Index Terms*—Sparse Bayesian learning; massive MIMO; channel estimation.**

I. Introduction

MASSIVE multiple input multiple output (MIMO) is a promising technique for achieving fifth generation targets of peak data rates up to 10 Gbit/s. The channel estimation process in such systems is performed via usage of uplink orthogonal pilot signals. However, these pilot signals must be reused in neighbouring cells as a results of the shortage of the orthogonal signals that cause pilot contamination. However, pilot contamination can be mitigated by reducing the number of pilots. Hence, the development of efficient channel estimation techniques for massive MIMO that are computationally low complexity and require a small number of pilots is a challenge that should be thoroughly addressed [1]-[3].

Compressed sensing (CS) techniques have received much attention since they can recover unknown signals from just a small number of measurements, thereby using significantly fewer samples than is possible via the conventional Nyquist rate, which is the signal recovery scheme developed for CS to exploit the sparse nature of signals (that is, only a small number of components in a signal vector are non-zero). CS allows for accurate system parameter estimation with less training, thereby addressing the pilot contamination problem and improving bandwidth efficiency as a consequence [4], [5]. However, classical CS algorithms require prior knowledge of channel sparsity, which is usually unknown in practical scenarios [4]-[6]. In addition, to applying CS algorithms, the sampling matrix must satisfy the restricted isometry property (RIP) to guarantee reliable estimation. Such a condition cannot be easily verified because it is computationally demanding [6], [7].

Recently, several methods has been proposed to tame the scarcity of CS-based channel estimation, however, these works assume dependency between antenna elements, as in realistic environments MIMO channel is generally correlated and statistically dependent, as the antennas are not sufficiently separated and the propagation environment does not provide a sufficient amount of rich scattering [8].

Considering the impact of antenna correlation, in this paper, we proposed an improved channel estimation technique based on sparse Bayesian learning (SBL) scheme, namely, a pattern-coupled hierarchical Gaussian framework [9]-[10], whereby, a priori probabilistic information regarding channel sparsity and the feature of the sparsity coefficients are being controlled by its own hyperparameter, and its neighbouring hyperparameters can be exploited for more reliable channel recovery to mitigate the pilot contamination problem. Also, the sampling matrix condition is efficiently overcome based on probabilistic formulation. We have also proposed enhancing the performance of the SBL-based estimator through the principle of thresholding to select the most significant taps to improve channel estimation accuracy. Furthermore, the cramer Rao bound (CRB) has also derived as a reference line.

The remainder of this paper is organized as follows. The multi-cell massive MIMO system model is presented in Section II. The SBL-based Channel Estimator is analyzed in Section III. Section IV presents the simulation results and the conclusions are drawn in Section V.

The following notations are adopted throughout this paper: for any matrix $\boldsymbol{A}$, $A_{i,j}$ denotes the (i,j)th element, while the superscripts $(.)^T$, $(.)^{-1}$ and $(.)^H$ denote the transpose operator, the inverse operator and the conjugate transpose operator, respectively. $tr(.)$ denotes the trace operator. A diagonal matrix with $a_1, ..., a_N$ on the main diagonal is denoted $diag(a_1, ..., a_N)$. Bold font is used to denote matrices and vectors, lower and upper case represents the time domain and frequency domain, respectively. The Frobenius

*This work is supported by Iraqi Higher Committee of Educational Development (HCED).

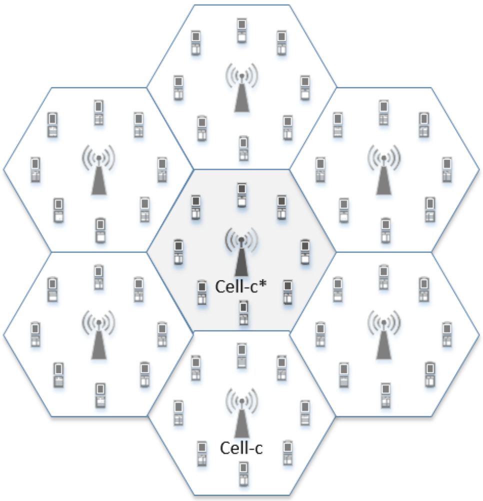

Fig. 1: Illustration of the system model of a multi-cell multi-user massive MIMO.

and spectral norms of a matrix $\boldsymbol{x}$ are denoted by $\|\boldsymbol{x}\|_F$ and $\|\boldsymbol{x}\|_2$ respectively. $E\{.\}$ denotes the expectation of random variables within the brackets. A Gaussian stochastic variable o is then denoted by $o \sim CN(r, q)$, where r is the mean and q is the variance. Furthermore, a random vector $\boldsymbol{x}$ having the prober complex Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is indicated by $\boldsymbol{x} \sim N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{det(\pi\boldsymbol{\Sigma})} e^{-(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$, for simplicity we refer to $N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ as $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. Massive MIMO System Model

Consider a time division duplexing (TDD) multi-cell massive MIMO system with C cells as shown in Fig. 1. Each cell comprises M antennas at the BS and N single-antenna mobile stations. To improve the spectral efficiency, orthogonal frequency division multiplexing (OFDM) is adopted [11],[12].

At the beginning of the transmission, all mobile stations in all cells synchronously transmit V OFDM pilot symbols to their serving base stations. The v-th pilot symbols of user n in the c-th cell is given by $\boldsymbol{X}_c^n[v] = [X_c^n[v,1] X_c^n[v,2] ... X_c^n[V,K]]^T$, where K is the number of subcarriers. Let $H^n_{c^*,c,i}[v,k]$ denotes the uplink channel frequency response of the n-th user in the c-th cell sent by the i-th antenna of cell c^* at the k-th subcarrier of the v-th OFDM symbol. The received signal $Y_{c^*,i}$ at the i-th antenna element in the cell c^* can be expressed as

$$\begin{aligned} Y_{c^*,i}[v,k] &= \sum_{n=1}^{N} H^n_{c^*,c^*,i}[v,k] X^n_{c^*}[v,k] \\ &+ \sum_{c=1,c\neq c^*}^{C} \sum_{n=1}^{N} H^n_{c^*,c,i}[v,k] X^n_c[v,k] + W_{c^*,i}[v,k], \end{aligned} \quad (1)$$

for $1 \leq i \leq M$ and $1 \leq c \leq C$, where $W_{c^*,i}[v,k]$ is the uplink channel's additive white Gaussian noise (AWGN). The set of equations constituted by (1) for $1 \leq i \leq M$ can be written as

$$\begin{aligned} \boldsymbol{Y}_{c^*}[v,k] &= \boldsymbol{X}_{c^*}[v,k] \boldsymbol{H}_{c^*,c^*}[v,k] \\ &+ \sum_{c=1,c\neq c^*}^{C} \boldsymbol{X}_c[v,k] \boldsymbol{H}_{c^*,c}[v,k] + \boldsymbol{W}_{c^*}[v,k], \end{aligned} \quad (2)$$

where $\boldsymbol{Y}_{c^*}[v,k] \in C^{1\times M}$ and $\boldsymbol{W}_{c^*}[v,k] \in C^{1\times M}$ are the two row vectors hosting $Y_{c^*,i}[v,k]$ and $W_{c^*,i}[v,k]$ for $1 \leq i \leq M$, respectively, $\boldsymbol{X}_{c^*}[v,k] \in C^{1\times N}$ and $\boldsymbol{X}_c[v,k] \in C^{1\times N}$ are the two row vectors hosting $X^n_{c^*}[v,k]$ and $X^n_c[v,k]$ for $1 \leq n \leq N$, respectively, while $\boldsymbol{H}_{c^*,c^*}[v,k] \in C^{N\times M}$ and $\boldsymbol{H}_{c^*,c}[v,k] \in C^{N\times M}$ are the two matrices having their n-th row and i-th column elements given by $H^n_{c^*,c^*,i}[v,k]$ and $H^n_{c^*,c,i}[v,k]$, respectively. Assuming that the channel is time-invariant during the channel estimation period, we can drop the OFDM symbol index v from $H^n_{c^*,c,i}[v,k]$. Specifically, $\boldsymbol{H}^n_{c^*,c,i}[k] = \boldsymbol{H}^n_{c^*,c,i}$ for the channel estimation period, where row vector $\boldsymbol{H}^n_{c^*,c,i} = [H^n_{c^*,c,i}[1]\ H^n_{c^*,c,i}[2] ... H^n_{c^*,c,i}[K]] \in C^{1\times K}$ represents the link between the n-th user in the c-th cell over all K OFDM subcarriers. OFDM partitions the multipath channel into a set of parallel and independent sub-channels. The fading coefficient of these sub-channel are the discrete-time Fourier transform of the multipath channel taps. Hence, the term $\boldsymbol{H}^n_{c^*,c,i}$ can be given as [7],[8]

$$\boldsymbol{H}^n_{c^*,c,i} = \boldsymbol{h}^n_{c^*,c,i} \boldsymbol{F}^T, \quad (3)$$

where the row vector $\boldsymbol{h}^n_{c^*,c,i} = [h^n_{c^*,c,i},(1)\ h^n_{c^*,c,i}(2), ..., h^n_{c^*,c,i}(L)] \in C^{1\times L}$ is the CIR between n-th user at c-th cell and the i-th antenna of the serving BS at the c^*th cell, L is the number of the paths, $\boldsymbol{F} \in C^{N\times L}$ represents the matrix comprising the discrete Fourier transform (DFT) matrix, the elements of which are given by $[e^{-j2\pi*(k-1)(l-1)/K}]$ for $1 \leq k \leq K$ and $1 \leq l \leq L$. Via exploiting the advantage of cyclic prefix we can omit the subcarrier index k to simplify our notation. Assuming the total V consecutive OFDM symbols are dedicated to pilot subcarriers, so the received signal associated with K OFDM symbols over $1 \leq v \leq V$ can be written as

$$\boldsymbol{Y}_{c^*} = \boldsymbol{X}_{c^*} \boldsymbol{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C} \boldsymbol{X}_c \boldsymbol{H}_{c^*,c} + \boldsymbol{W}_{c^*}. \quad (4)$$

The channel coefficient is modelled as $h^n_{c^*,c,i}[\ell] = \sqrt{\phi^n_{c^*,c,i}}[\ell] g^n_{c^*,c,i}[\ell]$ for $1 \leq \ell \leq L$, where $\phi^n_{c^*,c,i}$ captures the path-loss and shadowing (large-scale fading), while the term $g^n_{c^*,c,i}$ is assumed to be independent identical distribution (i.i.d) of unknown random variables with $CN(0,1)$ (small-scale fading) [3].

The received signal of (4) can be re-written as

$$\mathbf{Y}_{c^*} = \mathbf{X}^n_{c^*} \mathbf{F} \mathbf{h}_{c^*,c^*} + \mathbf{Z}_{c^*}, \quad (5)$$

where the term $\mathbf{Z}_{c^*} = \sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c \mathbf{F} \mathbf{h}'^n_{c^*,c} + \mathbf{W}_{c^*}$ in (5) represents the net sum of inter-cell interference plus the receiver

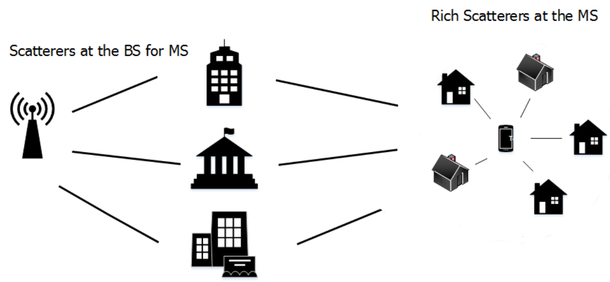

Fig. 2: Illustration of the rich scatterers wireless channel and the resulting channel impulse response is sparse.

noise. The variance interference σ_I^2 of the inter-cell interference term caused during pilot transmission can be expressed as

$$\sigma_I^2 = E\{(\sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c^n \mathbf{F} \mathbf{h}_{c^*,c})(\sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c \mathbf{F} \mathbf{h}_{c^*,c})^H\}. \quad (6)$$

We define the measurement matrix $\mathbf{A}_{c^*} = \mathbf{X}_{c^*}\mathbf{F}$, then (5) can be rewritten as

$$\mathbf{Y}_{c^*} = \mathbf{A}_{c^*}\mathbf{h}_{c^*,c^*} + \mathbf{Z}_{c^*}. \quad (7)$$

Based on the physical properties of outdoor electromagnetic propagation, the channel impulse response (CIR) in wireless communications usually possesses several significant channel taps as it shown in Fig. 2, i.e. the CIR are sparse, So, the number of non-zero channel taps is much smaller than the channel length, hence CS techniques can be applied for sparse channel estimation [13]-[15]. This sparsity feature can be exploited to reduce the necessary channel parameters needing to be estimated. In this case, we can address the pilot contamination problem by using fewer pilots than the unknown channel coefficients [13]-[15].

III. SBL-Based Channel Estimator

In this section, the pattern-coupled sparse Bayesian learning method is presented in the context of massive MIMO channel estimation. Based on Bayesian channel estimation philosophy, estimated unknown parameters of interest are an expectation of the posterior probability. As such, to obtain the estimated channel, we need to infer the posterior probability of the unknown parameters.

A. Bayesian Inference Model

Following the pattern-coupled sparse Bayesian learning model and based on Bayes' rule [10], the full posterior distribution of $\boldsymbol{h}_{c^*,c^*}$ over unknown parameters of interest for the problem at hand is proportional to the prior probability and the likelihood of the unknown parameters, that can be computed as

$$P(\boldsymbol{h}_{c^*,c^*}|\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{Y}_{c^*}) = P(\boldsymbol{h}_{c^*,c^*}|\boldsymbol{\alpha})P(\boldsymbol{Y}_{c^*}|\boldsymbol{h}_{c^*,c^*}), \quad (8)$$

where $\boldsymbol{\gamma}$ represents the inverse of the net sum of the noise and interference covariance matrices and $\boldsymbol{\alpha}$ are non-negative hyperparameters controlling the sparsity of the channel $\boldsymbol{h}_{c^*,c^*}$. According to probability theory, the term $P(\boldsymbol{Y}_{c^*}|\boldsymbol{h}_{c^*,c^*})$ can be written as

$$P(\boldsymbol{Y}_{c^*}|\boldsymbol{h}_{c^*,c^*}) = (\frac{1}{\sqrt{2\pi}\boldsymbol{\gamma}^{-1}})exp(-\frac{||\boldsymbol{Y}_{c^*} - \boldsymbol{h}_{c^*,c^*}\boldsymbol{A}_{c^*}||_2^2}{2\boldsymbol{\gamma}^{-1}}), \quad (9)$$

while, in the pattern-coupled model, the Gaussian prior for each channel coefficient $P(\boldsymbol{h_{c^*,c^*}}|\boldsymbol{\alpha})$ is given by

$$P(\boldsymbol{h_{c^*,c^*}}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} P(H_{c^*,c^*,i}[v,k]|\alpha_i,\alpha_{i+1},\alpha_{i-1}) \quad (10)$$

$$\begin{aligned} =&(2\pi)^{\frac{-M}{2}}\prod_{i=1}^{M}((\alpha_i,\beta\alpha_{i+1},\beta\alpha_{i-1}))^{\frac{1}{2}} \\ &exp[\frac{-1}{2}(H^n_{c^*,c^*,i}[v,k])^T((\alpha_i,\beta\alpha_{i+1},\beta\alpha_{i-1}))H^n_{c^*,c^*,i}[v,k]], \\ &i=1,...,M \end{aligned} \quad (11)$$

where $0 \leq \beta \leq 1$ is a parameter indicating the pattern relevance between the channel coefficient $H^n_{c^*,c^*,i}[v,k]$ and its neighboring coefficients $\{H^n_{c^*,c^*,i+1}[v,k], H^n_{c^*,c^*,i-1}[v,k]\}$. For $\beta = 0$, the Gaussian prior distribution in (10) is reduced to the prior for the conventional sparse Bayesian learning (which represents the uncorrelated channel scenario).

B. Proposed Algorithm of the Pattern-Coupled Hierarchical Model

We now proceed to perform Bayesian inference for the proposed pattern-coupled SBL-based estimator. The posterior $P(\boldsymbol{h}_{c^*,c^*}|\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{Y}_{c^*}) \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ follows a Gaussian distribution with its mean and covariance given respectively by

$$\boldsymbol{\mu} = \boldsymbol{\gamma}\boldsymbol{\Sigma}\boldsymbol{A_{c^*}}^H\boldsymbol{Y}_{c^*}, \quad (12)$$

$$\boldsymbol{\Sigma} = (\boldsymbol{D} + \boldsymbol{\gamma}(\boldsymbol{A_{c^*}})^H\boldsymbol{A}_{c^*})^{-1}, \quad (13)$$

where $\boldsymbol{D}$ is a diagonal matrix with its ith diagonal element is given by $[\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}]$, for $i = 1, ..., M$. The maximum a posterior (MAP) estimate of $\boldsymbol{h}_{c^*,c^*}$ is the mean of its posterior distribution, i.e.,

$$\hat{\boldsymbol{h}}_{c^*,c^*} = \boldsymbol{\mu} = ((\boldsymbol{A_{c^*}})^H\boldsymbol{A}_{c^*} + \boldsymbol{\gamma}^{-1}\boldsymbol{D})^{-1}(\boldsymbol{A_{c^*}})^H\boldsymbol{Y}_{c^*}. \quad (14)$$

To obtain the term $\hat{\boldsymbol{h}}_{c^*,c^*}$, we need to jointly estimate the hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, which can be achieved by exploiting the expectation-maximization (EM) approach (we refer interested readers to [10] for detailed derivations). So, the new estimate of $\boldsymbol{\alpha}^{(t+1)}$ and $\boldsymbol{\gamma}^{(t+1)}$ can be given as

$$\begin{aligned} \alpha_i^{(t+1)} =& 10^{-4}/0.5(\hat{\mu}_i^2 + \hat{\Sigma}_{i,i}) + \beta(\hat{\mu}_{i+1}^2 + \hat{\Sigma}_{i+1,i+1}) \\ &+ \beta(\hat{\mu}_{i-1}^2 + \hat{\Sigma}_{i-1,i-1}) + 10^{-4},\ i = 1,...,M, \end{aligned} \quad (15)$$

$$\begin{aligned} \boldsymbol{\gamma}^{(t+1)} =& M + 2*10^{-4}/||\boldsymbol{Y}_{c^*} - \boldsymbol{h}_{c^*,c^*}\boldsymbol{A}_{c^*}||_2^2 + (\boldsymbol{\gamma}^{(t)})^{-1} \\ &\sum_i(1 - \hat{\Sigma}_{i,i}(\alpha_i^{(t)} + \beta\alpha_{i-1}^{(t)} + \beta\alpha_{i+1})^{(t)}) \\ &+ 2*10^{-4},\ i = 1,...,M, \end{aligned} \quad (16)$$

The procedure for implementation of the proposed technique are summarized in algorithm 1.

Algorithm 1 SBL-based Channel Estimator

INPUTS: Pilot Signal $\boldsymbol{X}_{c^*}$, observation matrix $\boldsymbol{Y}_{c^*}$ and the measurement matrix $\boldsymbol{A}_{c^*} = \boldsymbol{F}\boldsymbol{X}_{c^*}$

Initial Configuration:

1: Select a specific convergence value $\boldsymbol{\epsilon}$.

2: Select a start value for $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\gamma}^{(t)}$

3: $t = 0$

4:**While** $\|(\hat{\boldsymbol{h}}_{c^*,c^*})^{(t+1)} - (\hat{\boldsymbol{h}}_{c^*,c^*})^{(t)}\| \leq \boldsymbol{\epsilon}$ **do**

5: Obtain a new estimate for $\boldsymbol{\alpha}^{(t+1)}$ and $\boldsymbol{\gamma}^{(t+1)}$ as in (15) and (17), respectively.

6: Compute $\boldsymbol{\Sigma} = (\boldsymbol{D} + \boldsymbol{\gamma}(\boldsymbol{A}_{c^*})^H \boldsymbol{A}_{c^*})^{-1}$

7: Compute $\hat{\boldsymbol{h}}_{c^*,c^*} = \boldsymbol{\mu} = \boldsymbol{\gamma}\boldsymbol{\Sigma}\boldsymbol{A}_{c^*}{}^H \boldsymbol{Y}_{c^*}$

8: $t \leftarrow t + 1$

8: **end while**

OUTPUTS: Return the Estimated Channel $\hat{\boldsymbol{h}}_{c^*,c^*}$

C. Enhanced SBL-Based Estimator

In contrast to the proposed SBL-based estimator, the performance of the proposed SBL-based estimator can be improved through the principle of thresholding, which can be applied to retain the most significant taps. The proposed algorithm therefore implements a threshold approach by conserving the channel taps that have energies above a threshold value of ϱ and setting the other taps to zero. The value of ϱ is the energy of the CIR.

D. CRB For SBL-Based Estimator

To quantify the best performance that can be achieved by the proposed algorithm, in this section, we derive the CRB of the pattern-coupled SBL channel estimation. The CRB on the covariance of any estimator $\hat{\boldsymbol{\theta}}$ can be given as $E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H\} \geq J^{-1}(\boldsymbol{\theta})$, where $J(\boldsymbol{\theta})$ is the Fisher information matrix (FIM) corresponding to the observation f, and can be given as

$$J(\boldsymbol{\theta}) = E(\frac{\partial}{\partial \boldsymbol{\theta}} logl(\boldsymbol{\theta}, f))(\frac{\partial}{\partial \boldsymbol{\theta}} logl(\boldsymbol{\theta}, f))^T, \quad (17)$$

where $l(\boldsymbol{\theta}, f)$ is the likelihood function corresponding to the observation f, parametrized by $\boldsymbol{\theta}$ [16].

Theorem 1: The closed form expression of the Bayesian CRB for the proposed SBL can be given as

$$J(\boldsymbol{h}_{c^*,c^*}) \geq (\frac{1}{(\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1})} + \frac{\boldsymbol{A}_{c^*}(\boldsymbol{A}_{c^*})^H}{\boldsymbol{\gamma}})^{-1}. \quad i = 1, ..., M. \quad (18)$$

Proof: See Appendix A.

IV. Simulation Results

In this section, we conduct experiments to evaluate the performance of the proposed algorithm and compare it to existing methods. The simulation parameters can be summarized as follows, $M = 100$ antennas, $N = 10$ users, $L = 10$ taps and

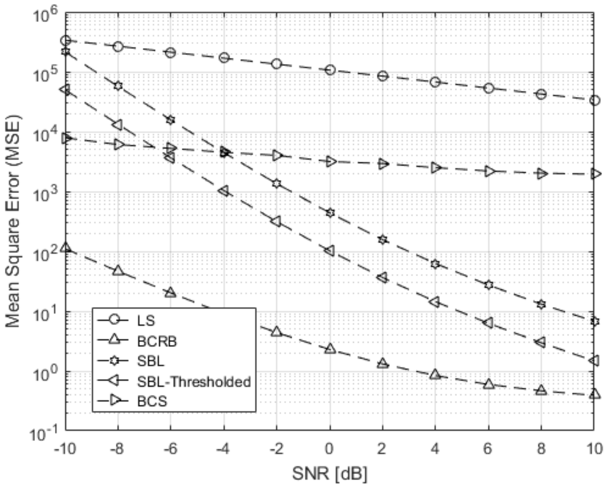

Fig. 3: Relative MSE performance comparison between SBL, Modified SBL, BCS, and the LS versus SNR.

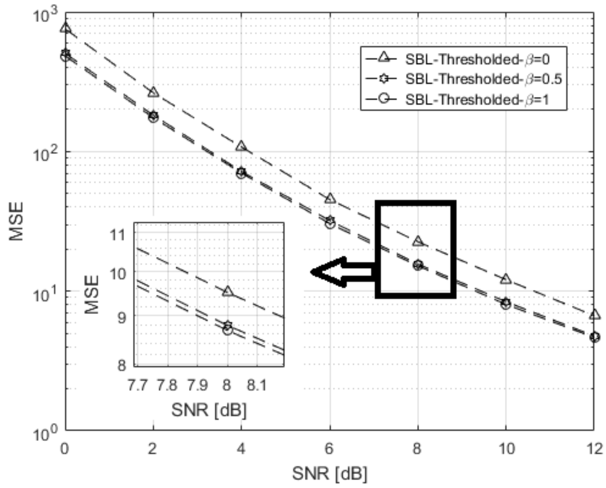

Fig. 4: Relative MSE of the Pattern-Coupled SBL for $\beta = \{0, 0.5$ and $1\}$.

$K = 100$ subcarriers. The simulation scenario is influenced by strong pilot contamination ($\phi^n_{c^*,c^*,i} = 1$ and $\phi^n_{c^*,c,i} = 0.7$). The simulation results are obtained by averaging over 1000 channel realizations.

Fig. 3 demonstrates the relative MSE performance comparison of least square (LS) with no pilot contamination, SBL-based, thresholded-SBL, Bayesian compressed sensing (BCS) [17] channel estimation techniques along with the BCRB reference line. The results showed that the proposed SBL approach provided significant performance enhancement over LS and BCS with respect to estimation accuracy as a result of exploiting the correlation between the antennas. In addition, the results showed that the thresholding approach strengthened the estimation accuracy of conventional SBL as the CIR possessed so many taps without a significant energy. By setting the threshold and neglecting these taps, a huge portion of the noise and interference from pilot contamination would be eliminated.

Fig. 4 shows the relative MSE performance comparison of SBL-based channel estimation with different settings for $\beta = \{0, 0.5, \text{ and } 1\}$. It can be observed that estimation accuracy is improved when employing the antennas correlation on a large scale.

V. Conclusion

In this paper, we proposed a SBL-based channel estimation algorithm for multi-cell massive MIMO systems. The simulation results revealed that the SBL-based channel estimation algorithm had a tremendous advantage over conventional estimators. Furthermore, the proposed technique can be enhanced by thresholding the CIR to a specific value. In addition, the results demonstrated that the estimation accuracy is enhanced by employing the correlation between antennas on a large scale.

Appendix A
Proof of Theorem 1

Following (19), we can write the FIM as

$$J(\boldsymbol{h}_{c^*,c}) \geq E(\frac{\partial^2 log(p_{\boldsymbol{h}_{c^*,c}/Y_{c^*}}(\boldsymbol{h}_{c^*,c}, Y_{c^*}))}{\partial^2 \boldsymbol{h}_{c^*,c}}). \tag{19}$$

Based on Bayes' rule in (8), the FIM can be decomposed into two terms

$$\begin{aligned} &E(\frac{\partial^2 log(p_{\boldsymbol{h}_{c^*,c}/Y_{c^*}}(\boldsymbol{h}_{c^*,c}, Y_{c^*}))}{\partial^2 \boldsymbol{h}_{c^*,c}}) = \\ &E(\frac{\partial^2 log(p_{Y_{c^*}/\boldsymbol{h}^n_{c^*,c}}(Y_{c^*}, \boldsymbol{h}_{c^*,c}))}{\partial^2 \boldsymbol{H}_{c^*,c}}) + \\ &E(\frac{\partial^2 log(p_{\boldsymbol{h}_{c^*,c}}(\boldsymbol{h}_{c^*,c}))}{\partial^2 \boldsymbol{h}_{c^*,c}}). \end{aligned} \tag{20}$$

So, (22) can be expressed in matrix form as

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P. \tag{21}$$

where $\mathbf{J}$, $\mathbf{J}_D$ and $\mathbf{J}_P$ represent the Bayesian FIM, data information matrix and prior information matrix, respectively.

Using (9), the data information matrix $\mathbf{J}_D$ can be given as

$$\mathbf{J}_D = E(\frac{\partial^2 log(p_{Y_{c^*}/\boldsymbol{h}_{c^*,c}}(Y_{c^*}, \boldsymbol{h}_{c^*,c}))}{\partial^2 \boldsymbol{h}_{c^*,c}}) = \frac{\boldsymbol{A}_{c^*}(\boldsymbol{A}_{c^*})^H}{\gamma}. \tag{22}$$

Considering (10), the prior information matrix $\mathbf{J}_D$ can be given as

$$\mathbf{J}_P = E(\frac{\partial^2 log(p_{\boldsymbol{h}_{c^*,c}}(\boldsymbol{h}_{c^*,c}))}{\partial^2 \boldsymbol{h}_{c^*,c}}) = (\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1})^{-1}. \tag{23}$$

References

[1] Nokia Solutions and Networks, ”Technology Vision 2020 White Paper”, pp. 116, 2013.

[2] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfr ojd, and T. Svensson, “The Role of Small Cells, Coordinated Multipoint, and Massive MIMO in 5G”, IEEE Communications Magazine, vol. 52, no. 5, pp. 4451, May 2014.

[3] 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 Transaction on Vehicular Technology, pp. 99, 2015.

[4] S. Nguyen, “Compressive Sensing for Multi-channel and Large-scale MIMO Networks”,PhD. thesis, Dept. of Elect. And Comp. Eng., Concordia Univ., Montreal, Quebec, Canada, 13 Jan 2014.

[5] C. Qi, G. Yue, L. Wu, Y. Huang and A. Nallanathan, “Pilot Design Schemes for Sparse Channel Estimation in OFDM Systems”, IEEE Trans. Veh. Technol., vol. 64, no. 4, pp. 1493-1505, 2015.

[6] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, “Directions-of-arrival estimation through Bayesian compressive sensing strategies”,IEEE Trans. Antennas Propagat., vol.61, no. 7, pp. 38283838, July 2013.

[7] R. G. Baraniuk, “Compressive sampling”, IEEE Signal Process. Mag., vol. 24, no. 4, pp. 118-124, July 2007.

[8] Hien Quoc Ngo et al.: “Analysis of the Pilot Contamination Effect in Very Large Multicell Multiuser MIMO Systems for Physical Channel Models”, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech, May 2011.

[9] J. Fang, Y. Shen, H. Li and P. Wang, “Pattern-Coupled Sparse Bayesian Learning for Recovery of Block-Sparse Signals”, IEEE Transaction on Signal Processing, vol. 63, no. 2, pp. 360-372, 2015.

[10] Y. Shen, H. Duan, J. Fang, H. Li, “Pattern-coupled sparse bayesian learning for recovery of block-sparse signals”, Acoustics Speech and Signal Processing (ICASSP) 2014 IEEE International Conference on, pp. 1896-1900, May 2014.

[11] J. Zhang, B. Zhang, S. Chen et al., “Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems IEEE J. Sel. Topics Signal Processing”, pp. 759-772, 2014.

[12] M. R. Nakhai, “Multicarrier transmission”, IET Signal Process., vol. 2, no. 1, pp. 114, Mar. 2008.

[13] M. Masood, L. Afify, and Tareq Y. Al-Naffouri, “Efficient Coordinated Recovery of Sparse Channels in Massive MIMO,” IEEE Transactions on Signal Processing, vol. 63, no. 1, Jan. 2015.

[14] W. Ding, F. Yang, W. Dai and J. Song, “Time-frequency joint sparse channel estimation for MIMO-OFDM systems”, IEEE Commun. Lett., vol. 19, no. 1, pp. 58-61, Jan. 2015.

[15] C. Qi, G. Yue, L. Wu, Y. Huang and A. Nallanathan, “Pilot design schemes for sparse channel estimation in OFDM systems”, IEEE Trans. Veh. Technol., vol. 64, no. 4, pp. 1493-1505, Apr. 2015.

[16] A. K. Jagannatham and B. D. Rao, “Whitening-Rotation-Based Semi-Blind MIMO Channel Estimation”, IEEE Trans. on Signal Processing, vol. 54, no. 3, pp. 861-869, 2006.

[17] H. Al-Salihi and M. R. Nakhai, “Bayesian Compressed Sensing-based Channel Estimation for Massive MIMO Systems”, European Conference on Networks and Communications (EUCNC), Athens -EuCNC/EURASIP Student Best Paper Award, 27-30 June 2016.