

RESEARCH

Open Access



Channel estimation for FDD massive MIMO system by exploiting the sparse structures in angular domain

Yurong Wang¹, Xiaochen Xia^{1,2*} , Kui Xu¹, Wei Xie¹ and Minghai Wang²

Abstract

Massive multiple-input multiple-output (MIMO) is a powerful supporting technology to meet the energy/spectral efficiency and reliability requirement of Internet of Things (IoT) network. However, the gain of massive MIMO relies on the availability of channel state information (CSI). In this paper, we investigate the channel estimation problem for frequency division duplex (FDD) massive MIMO system. By analyzing the sparse property of the downlink massive MIMO channel in the angular domain, a structured prior-based sparse Bayesian learning (SP-SBL) approach is proposed to estimate the downlink channels between base station (BS) and users reliably. The scheme can be implemented without the knowledge of channel statistics and angular information of users. The simulation results show that the proposed scheme outperforms the reference schemes significantly in terms of normalized mean square error (NMSE) for a variety of scenarios with different lengths of pilot sequence, transmit signal-to-noise ratios (SNRs), and angular spreads.

Keywords: FDD massive MIMO, Channel estimation, Structured prior, Sparse Bayesian learning, Normalized mean square error

1 Introduction

As an emerging technology that aims to allow everything to connect, interact, and exchange data, the Internet of Things (IoT) has attracted extensive attention in various areas such as governments, industry, and academia [1]. The massive connectivity, strict energy limitation, and requirement of ultra-reliable transmission are often termed as the most distinct features of IoT [2, 3]. These features make massive multiple-input multiple-output (MIMO) [4] a natural supporting technology for IoT since through coherent processing over the signals of large-scale antenna array, massive MIMO can produce tremendous spectral/energy efficiency gain and improve the reliability of IoT transmission significantly.

However, the coherent processing depends on the reliable estimation of channel state information (CSI) between base station (BS) and users [5–7]. In time division

duplex (TDD) massive MIMO system, by exploiting the channel reciprocity, both the uplink and downlink channels can be obtained using the simple least square (LS) approach [8]. The consumption of pilot resource scales with the number of users. However, there is no channel reciprocity in frequency division duplex (FDD) massive MIMO system. In this case, the downlink channel estimation becomes particularly challenging since the downlink training and corresponding CSI feedback yield an unacceptably high overhead. For example, with the traditional LS channel estimation scheme, it is well-known that the length of the required pilot sequence must be larger than the number of antennas at BS [8], which will degrade the system performance greatly.

In the practical system, the BS is usually elevated at a relatively high altitude with few surrounding scatters. Therefore, the scattering process often occurs in vicinity of user, which results in very narrow signal angle of departure (AoD) at BS [8, 9] for far-field transmission. Based on a physical channel model, several works have shown that the signals from different antennas of BS exhibit high correlation, and hence, the channel between BS and user

*Correspondence: tjuxc@sina.com

¹Institute of Communication Engineering, Army Engineering University of PLA, No. 88, Houbiaoying Street, Nanjing 0086-210007, China

²31682 Army of PLA, Lanzhou, China

can be represented by a sparse vector in some alternate domain (usually called angular domain) [10, 11]. This property can be utilized to design cost-efficient channel estimation schemes [12–14]. The main idea behind is to transform the channel to the angular domain by exploiting the correlations between the channel elements and then estimate the effective low-dimension angular-domain channel with the LS or minimum mean square error (MMSE) method. In this way, the training consumption can be reduced greatly. However, these schemes require the knowledge of channel covariance matrix or angular-domain information, such as the AoDs of the signals, which is not always available. Another path to reduce the training overhead is to formulate the channel estimation as a sparse recovery problem and estimate the angular-domain channel using the algorithms in compressive sensing, such as orthogonal matching pursuit (OMP) [15, 16] and sparse Bayesian learning (SBL) [17]. These schemes do not need the channel covariance matrix and angular information; however, suffer from performance loss when the length of pilot sequence is small. To improve the quality of channel estimation, novel channel estimation schemes based on block l_1/l_2 optimization [18] and variational SBL [19] were proposed recently. These schemes exploit the structures in the channel sparsity to give a robust channel estimation.

In this paper, we investigate the downlink channel estimation problem for FDD massive MIMO system considering the above challenges. The main contributions are as follows:

- 1 The sparse property of downlink massive MIMO channel in angular domain is analyzed theoretically. The results show that the angular-domain channel exhibits two kinds of sparse structures, namely the joint sparsity and burst sparsity.
- 2 By exploiting the two kinds of sparsity, a structured prior-based SBL (SP-SBL) approach is proposed to estimate the downlink channel between the BS and user reliably. The scheme does not require the channel statistics and AoD information.
- 3 Extensive numerical simulations are presented to validate the effectiveness of the proposed scheme. The results show that the proposed scheme outperforms the reference schemes significantly in terms of normalized mean square error (NMSE) for a variety of scenarios with different lengths of pilot sequence, transmit signal-to-noise ratios (SNRs), and angular spreads.

Notations: We use \mathbf{B}^* , \mathbf{B}^T , \mathbf{B}^H , $|\mathbf{B}|$, and $\|\mathbf{B}\|$ to denote conjugate, transpose, conjugate transpose, determinant, and Frobenius norm of matrix \mathbf{B} , respectively. $\mathbf{B} \in \mathbb{C}^{N \times M}$ means \mathbf{B} is an $N \times M$ complex-valued matrix. $\mathcal{CN}(\mathbf{b} | \mathbf{m}, \mathbf{C})$

means that \mathbf{b} is a complex Gaussian variable with mean \mathbf{m} and covariance matrix \mathbf{C} . $[\mathbf{B}]_{ij}$ denotes the $\{i, j\}$ th element of matrix \mathbf{B} . $\mathbb{E}(\cdot)$ denotes the expectation. $\nabla_{\mathbf{b}}(f(\mathbf{b}))$ denotes the gradient of function $f(\mathbf{b})$ w.r.t. the vector \mathbf{b} .

2 Methods

The rest of the paper is organized as follows: Section 3 describes the system model of the FDD massive MIMO system. Section 4 analyzes the sparse property of the angular-domain channel and presents the SP-SBL-based channel estimation scheme. Section 5 presents the simulation results to validate the effectiveness of the proposed scheme. Section 6 draws the conclusions.

3 System model

Consider the massive MIMO system with a BS and K users. It is assumed that both the BS and users are equipped with uniform linear arrays (ULAs) with half wavelength antenna spacing. The numbers of antennas at BS and user are N and M , respectively, which satisfy $M \ll N$. According to the ray-tracing model [8], the downlink channel between BS and user k can be expressed as:

$$\mathbf{H}_k = \int_{\theta_k - \Delta_d}^{\theta_k + \Delta_d} \int_{\varphi_k - \Delta_a}^{\varphi_k + \Delta_a} r_k(\theta, \varphi) \mathbf{a}_{\text{BS}}(\theta) \mathbf{a}_{\text{U}}^H(\varphi) d\theta d\varphi \quad (1)$$

where θ_k denotes AoD w.r.t. the array of BS, and φ_k denotes the angle of arrival (AoA) w.r.t. the array of user k . Δ_d and Δ_a denote the corresponding angular spreads. $r_k(\theta, \varphi)$ denotes the complex channel gain for angle $\{\theta, \varphi\}$. $\mathbf{a}_{\text{BS}}(\theta)$ and $\mathbf{a}_{\text{U}}(\varphi)$ denote the array steering vectors, which are given by:

$$\begin{aligned} \mathbf{a}_{\text{BS}}(\theta) &= \left[1, e^{j\pi \sin \theta}, \dots, e^{j\pi(N-1) \sin \theta} \right]^T \\ \mathbf{a}_{\text{U}}(\varphi) &= \left[1, e^{j\pi \sin \varphi}, \dots, e^{j\pi(M-1) \sin \varphi} \right]^T \end{aligned} \quad (2)$$

In practice, the BS is often deployed at a high place such as the top of high building. As discussed in the introduction, the limited number of scatterers around BS will result in very narrow angular spread Δ_d in far-field propagation. Conversely, the waves arriving at the user are usually uniformly distributed in AoA. Therefore, it is reasonable to assume Δ_a is close to π [9]. Note that the proposed channel estimation scheme in the next section is very general, which is valid for arbitrary Δ_a .

Let \mathbf{S} denote the $N \times T$ pilot matrix of BS which is broadcasted in T successive symbol times. The power of each pilot symbol is $|\mathbf{S}|_{ij}|^2 = P$. For a practical training scheme, T should be much smaller than N . The received training signal at user k can be expressed as:

$$\mathbf{Z}_k = \mathbf{H}_k^H \mathbf{S} + \mathbf{N}_k \in \mathbb{C}^{M \times T} \quad (3)$$

where $\mathbf{N} \in \mathbb{C}^{M \times T}$ denotes the additive white Gaussian noise (AWGN) matrix whose elements have zero mean and variance σ^2 .

Note that for the full-duplex massive MIMO system with separate antenna configuration [11], the channel reciprocity is also not available which makes the downlink channel estimation very challenging. Since the downlink transmission model of full-duplex massive MIMO is similar with the FDD counterpart in the above, the proposed scheme can be utilized to estimate the downlink channel directly to reduce the pilot consumption.

4 Channel estimation by exploiting the sparse structures

In this section, we first analyze the property of channel sparsity in the angular domain. Based on the theoretical results, we propose a novel SP-SBL approach to estimate the downlink channel between BS and user.

4.1 Joint and burst sparsity in the angular domain

According to [11], the angular-domain channel matrix between BS and user k can be expressed as:

$$\mathbf{X}_k = \mathbf{A}_N^H \mathbf{H}_k \in \mathbb{C}^{N \times M} \quad (4)$$

where \mathbf{A}_N is a shifted discrete Fourier transform (DFT) matrix of dimension N , that is:

$$\mathbf{A}_N = \left[\mathbf{a}_N \left(-\frac{1}{2} \right), \mathbf{a}_N \left(-\frac{1}{2} + \frac{1}{N} \right), \dots, \mathbf{a}_N \left(\frac{1}{2} - \frac{1}{N} \right) \right],$$

$$\mathbf{a}_N(\rho) = \frac{1}{\sqrt{N}} [1, \exp(j2\pi\rho), \dots, \exp(j2\pi(N-1)\rho)]^T. \quad (5)$$

Note that the m th column of \mathbf{X}_k (denoted by $\mathbf{x}_{k,m}$) is the angular-domain channel vector between the BS and the m th antenna of user k . The angular-domain channel is the projection of channel onto the space spanned by the DFT bases. Since a DFT basis is in fact equivalent to an array steering vector with specific AoD, the angular-domain channel can be viewed as the channel response in the angular domain, and the amplitude of each angular-domain channel element indicates the channel strength for the path with specific AoD. Moreover, due to the limited number of scatterers around the BS, the spread of AoD will be very narrow in far-field propagation scenario. As a result, only a small fraction of elements in angular-domain channel matrix has significant amplitude. This results in the sparsity in angular-domain channel.

Mathematically, following a similar analytic approach as that in [10, 11], it can be shown that the angular-domain channel matrix has the following two kinds of sparse structures.

- Joint sparsity among the columns of \mathbf{X}_k : The n th element of $\mathbf{x}_{k,m}$ has a significant amplitude only when

$n \in \Omega_k$, where Ω_k is given by $\Omega_k = \{n | -\frac{1}{2} + \frac{n-1}{N} \in [\frac{1}{2} \sin(\theta_k - \Delta_d), \frac{1}{2} \sin(\theta_k + \Delta_d)]\}$, $n \in \mathbb{N}^+$. Since Ω_k is independent of the column index m , the indexes of the significant elements in each column of \mathbf{X}_k are the same.

- Burst sparsity in each column of \mathbf{X}_k : The indexes of significant elements in $\mathbf{x}_{k,m}$ appear in block.

As discussed in the above, Δ_d is commonly small. Thus, only a small number of elements in $\mathbf{x}_{k,m}$ has a significant value. Physically, the joint sparsity of angular-domain channel is due to the fact that the size of the user's array can be neglected in far-field propagation, and thus, the channels between the BS and all antennas of user can be considered to have similar property in the angular domain. Moreover, it is noted that the AoD of downlink signal varies in a continuous interval $[\theta_k - \Delta_d, \theta_k + \Delta_d]$. This causes the burst sparsity nature in the angular-domain channel. The two kinds of sparse structures are shown in Fig. 1 for an example with $N = 64$ and $M = 4$.

Here, we shall point out that, for the practical system with finite N , the n th element of $\mathbf{x}_{k,m}$ with $n \notin \Omega_k$ is small but not zero as shown in Fig. 1. Therefore, $\mathbf{x}_{k,m}$ is in fact approximately sparse. Note that [16, 18] assumed that the angular-domain channel element is exactly zero if $n \notin \Omega_k$, which we call as exactly sparse channel. Although this assumption simplifies the model, it may change the real system performance in the high SNR region greatly as will be shown in the simulation of Section 5.

4.2 Structured prior-based Bayesian channel estimation

By substituting (4) in (3), the received training signal at user k can be rewritten as:

$$\mathbf{Y}_k = \mathbf{Z}_k^H = \Phi \mathbf{X}_k + \mathbf{N}_k^T \in \mathbb{C}^{T \times M} \quad (6)$$

where $\Phi = \mathbf{S}^H \mathbf{A}_N \in \mathbb{C}^{T \times N}$. Note that as long as \mathbf{X}_k is known, the channel matrix \mathbf{H}_k can be recovered directly based on (4), that is, $\mathbf{H}_k = \mathbf{A} \mathbf{X}_k$ (which is called basis expansion model).

By treating Φ as the sensing matrix, \mathbf{X}_k can be estimated using the conventional SBL [17] or OMP [15] approach from (6). However, the performance will be poor if the number of pilot symbols (i.e., T) is small as will be seen in the simulations. In this paper, we propose a SP-SBL approach to estimate the channel reliably by exploiting the two kinds of sparse structures discussed in the last subsection.

Different from the conventional SBL where a i.i.d. Gaussian prior is utilized, we propose a structured prior for \mathbf{X}_k as follows:

$$p(\mathbf{X}_k | \alpha_k, \gamma_k) = \prod_{m=1}^M CN(\mathbf{x}_{k,m} | 0, \gamma_{k,m}^{-1} \mathbf{D}_k^{-1}) \quad (7)$$

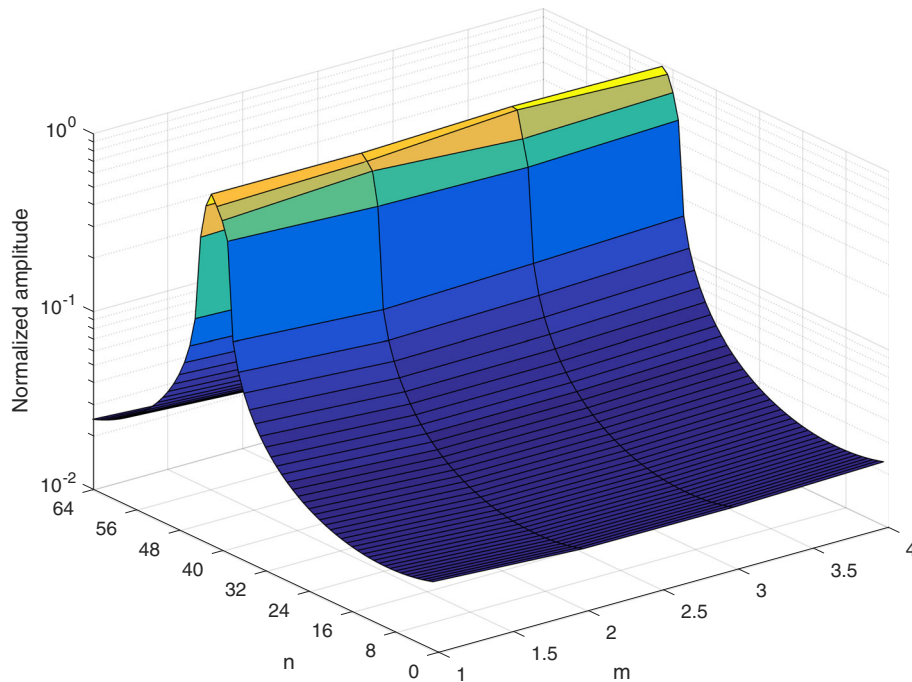


Fig. 1 Normalized amplitudes for the elements of angular-domain channel matrix \mathbf{X}_k , where $N = 64$, $M = 4$, and $\theta_k = 15.5^\circ$. The angular spread for AoD is $\delta_d = 5^\circ$. It is assumed that the waves arriving at the user are usually uniformly distributed in AoA [12]. Therefore, Δ_σ is equal to π

where $\gamma_k = [\gamma_{k,1}, \dots, \gamma_{k,M}]^T$ and $\alpha_k = [\alpha_{k,1}, \dots, \alpha_{k,N}]^T$. \mathbf{D}_k an $N \times N$ diagonal matrix whose n th diagonal element is modeled as:

$$[\mathbf{D}_k]_{n,n} = \begin{cases} \alpha_{k,N} + \alpha_{k,1} + \alpha_{k,2}, & n = 1 \\ \alpha_{k,n-1} + \alpha_{k,n} + \alpha_{k,n+1}, & 2 \leq n \leq N - 1 \\ \alpha_{k,N-1} + \alpha_{k,N} + \alpha_{k,1}, & n = N \end{cases} \quad (8)$$

For the prior model in (7) and (8), the m th column of angular-domain channel matrix \mathbf{X}_k has precision matrix $\gamma_{k,m} \mathbf{D}_k$. Thus, precision matrices for different columns of \mathbf{X}_k differ only with a scaling factor $\gamma_{k,m}$. This property captures the joint sparsity among the columns of \mathbf{X}_k . The parameter $\gamma_{k,m}$ is utilized to model the relative difference between the amplitudes of $\{\mathbf{x}_{k,m}\}_{m=1}^M$. By exploiting the joint sparsity, the recovery errors which predict zero for one element of $\mathbf{x}_{k,m}$ and predict non-zero for the element of $\mathbf{x}_{k,m'}$ ($m \neq m'$) in the same position can be mitigated effectively.

Additionally, the precision of the n th element of $\mathbf{x}_{k,m}$ can be expressed as $\gamma_{k,m} (\alpha_{k,n-1} + \alpha_{k,n} + \alpha_{k,n+1})$. Thus, the precisions of adjacent elements in $\mathbf{x}_{k,m}$ are mutually coupled. When $\alpha_{k,n} \rightarrow \infty$, the estimations of n th, $(n-1)$ th, and $(n+1)$ th elements of $\mathbf{x}_{k,m}$ will be driven to zero simultaneously. Therefore, the zero elements (and hence non-zero elements) in the estimation of $\mathbf{x}_{k,m}$ will appear in block, which just captures the block sparsity structure.

By exploiting burst sparsity, the situation that a significant element of $\mathbf{x}_{k,m}$ is predicted as zero (or near zero), isolated in $\mathbf{x}_{k,m}$, can be reduced greatly. Note that the precisions of the first and the last elements of $\mathbf{x}_{k,m}$ are also coupled. This structure captures a basic property of angular-domain channel, i.e., the elements at the beginning and that at the end of $\mathbf{x}_{k,m}$ tend to be close to zero or have large amplitude simultaneously.

According to (6), the likelihood function of the received training signal \mathbf{Y}_k can be expressed as:

$$p(\mathbf{Y}_k | \mathbf{X}_k) = \prod_{m=1}^M \mathcal{CN}(\mathbf{y}_{k,m} | \Phi \mathbf{x}_{k,m}, \sigma^2 \mathbf{I}_T) \quad (9)$$

Using (7), (9), and the property of complex Gaussian distribution, the posterior distribution of $\mathbf{x}_{k,m}$ can be derived as:

$$\begin{aligned} p(\mathbf{X}_k | \mathbf{Y}_k) &\propto p(\mathbf{Y}_k | \mathbf{X}_k) p(\mathbf{X}_k | \alpha_k, \gamma_k) \\ &= \prod_{m=1}^M \mathcal{CN}(\mathbf{x}_{k,m} | \mu_{k,m}, \Sigma_{k,m}) \end{aligned} \quad (10)$$

where the posterior mean and covariance matrix of $\mathbf{x}_{k,m}$ are given by:

$$\begin{aligned} \mu_{k,m} &= \frac{1}{\sigma^2} \Sigma_{k,m} \Phi^H \mathbf{y}_{k,m} \\ \Sigma_{k,m} &= \left(\mathbf{D}_k^{-1} + \frac{1}{\sigma^2} \Phi^H \Phi \right)^{-1} \end{aligned} \quad (11)$$

To give a fully Bayesian treatment, similar to conventional SBL, we introduce a gamma hyperprior for α_k :

$$\begin{aligned} p(\alpha_k) &= \prod_{n=1}^N \text{Gam}(\alpha_{k,n} | a, b) \\ &= \prod_{n=1}^N \frac{1}{\Gamma(a)} b^a \alpha_{k,n}^a \exp(-b\alpha_{k,n}), \end{aligned} \quad (12)$$

where a and b are fixed parameters. Commonly, a and b are set as small values to impose a non-informative prior. In this paper, a and b are chosen as $a = b = 10^{-4}$. Moreover, we introduce a Dirichlet hyperprior for γ_k :

$$p(\gamma_k) = C(\mathbf{u}) \prod_{m=1}^M \gamma_{k,m}^{u_m} \quad (13)$$

where $\mathbf{u} = [u_0, \dots, u_M]$ is a fixed parameter with $u_0 = \sum_{m=1}^M u_m$. $C(\mathbf{u}) = \frac{\Gamma(u_0)}{\Gamma(u_1) \dots \Gamma(u_M)}$ is the normalization constant. Since the expectation of $\gamma_{k,m}$ with respect to the distribution (13) is given by $\mathbb{E}[\gamma_{k,m}] = \frac{u_m}{u_0}$, we can interpret \mathbf{u} as the parameter which gives an initial guess on the relative difference between the amplitudes of $\{\mathbf{x}_{k,m}\}_{m=1}^M$.

As in the convention of sparse Bayesian learning framework, the hyperpriors are utilized to imbed the prior knowledge of the parameters into the estimation algorithm. Moreover, as shown in [17], the utilization of gamma hyperprior is helpful to produce a sparse solution which is desired in our problem.

As long as α_k and γ_k are obtained, the maximum posterior (MAP) estimation of \mathbf{X}_k can be given by its posterior mean in (11). Therefore, in the following, we focus on finding the optimal α_k and γ_k by solving the MAP problem:

$$\begin{aligned} \{\hat{\alpha}_k, \hat{\gamma}_k\} &= \arg \max_{\alpha_k, \gamma_k} p(\alpha_k, \gamma_k | \mathbf{Y}_k) \\ &= \arg \min_{\alpha_k, \gamma_k} -\log p(\alpha_k, \gamma_k, \mathbf{Y}_k). \end{aligned} \quad (14)$$

Different from the conventional SBL, solving the above problem directly is quite challenging due to the utilization of the structured prior. To address this problem, we resort to the expectation maximization (EM) [20] algorithm to find a computationally efficient solution.

4.3 Solving the SP-SBL using expectation maximization

Instead of solving the MAP problem in (14) directly, the EM algorithm tries to find the optimal α_k and γ_k by maximizing the expected complete-data log-likelihood function, that is:

$$\begin{aligned} \{\hat{\alpha}_k, \hat{\gamma}_k\} &= \arg \max_{\alpha_k, \gamma_k} Q(\alpha_k, \gamma_k) \\ Q(\alpha_k, \gamma_k) &= \mathbb{E}_{\mathbf{X}_k | \mathbf{Y}_k, \hat{\alpha}_k^{\text{old}}, \hat{\gamma}_k^{\text{old}}} [\ln p(\alpha_k, \gamma_k, \mathbf{X}_k, \mathbf{Y}_k)] \end{aligned} \quad (15)$$

where the expectation is w.r.t. the posterior distribution of \mathbf{X}_k given by (10). $\hat{\alpha}_k^{\text{old}}$ and $\hat{\gamma}_k^{\text{old}}$ denote the latest estimations of α_k and γ_k , respectively. Each iteration of the algorithm consists of an expectation step (E-step) and a maximization step (M-step).

In the E-step of the i th iteration, the posterior distribution of \mathbf{X}_k is computed approximately using the estimations of α_k and γ_k in the $(i-1)$ th iteration (denoted by $\hat{\alpha}_k^{(i-1)}$ and $\hat{\gamma}_k^{(i-1)}$) based on (10), which is then used to evaluate the expected complete-data log-likelihood function.

In the M-step of the i th iteration, the estimation of α_k and γ_k is updated to maximize the expected complete-data log-likelihood function obtained in E-step.

4.3.1 Update of γ_k

By substituting (10) and (13) into (15) and discarding the terms irrelevant to γ_k , it can be shown that the expected complete-data log-likelihood function reduces to"

$$\begin{aligned} Q(\hat{\alpha}_k^{(i-1)}, \gamma_k) &= \sum_{m=1}^M u_m \ln \gamma_{k,m} + \sum_{m=1}^M \ln |\gamma_{k,m} \hat{\mathbf{D}}_k| \\ &\quad - \sum_{m=1}^M \text{tr} \left(\gamma_{k,m} \hat{\mathbf{D}}_k \left(\hat{\Sigma}_{k,m} + \hat{\mu}_{k,m} \hat{\mu}_{k,m}^H \right) \right) + \text{const}. \end{aligned} \quad (16)$$

Note that α_k has been fixed at its estimation after the last iteration, i.e., $\alpha_k = \hat{\alpha}_k^{(i-1)}$. $\hat{\mathbf{D}}_k$ can be computed using (8) by replacing α_k with $\hat{\alpha}_k^{(i-1)}$. $\hat{\Sigma}_{k,m}$ and $\hat{\mu}_{k,m}$ can be computed using (11) by replacing $\{\alpha_k, \gamma_k\}$ with $\{\hat{\alpha}_k^{(i-1)}, \hat{\gamma}_k^{(i-1)}\}$. The first-order derivative of $Q(\hat{\alpha}_k^{(i-1)}, \gamma_k)$ w.r.t $\gamma_{k,m}$ can be expressed as:

$$\frac{dQ(\hat{\alpha}_k^{(i-1)}, \gamma_k)}{d\gamma_{k,m}} = \frac{N + u_m}{\gamma_{k,m}} - \text{tr} \left(\hat{\mathbf{D}}_k \left(\hat{\Sigma}_k + \hat{\mu}_k \hat{\mu}_k^H \right) \right) \quad (17)$$

By setting (17) to zero, we can obtain the estimation for $\gamma_{k,m}$ in the i th iteration:

$$\hat{\gamma}_{k,m}^{(i-1)} = \frac{N + u_m}{\text{tr} \left(\hat{\mathbf{D}}_k \left(\hat{\Sigma}_k + \hat{\mu}_k \hat{\mu}_k^H \right) \right)} \quad (18)$$

4.3.2 Update of α_k

By substituting (10) and (12) into (15) and discarding the terms irrelevant to α_k , it can be shown that the expected complete-data log-likelihood function reduces to:

$$\begin{aligned}
Q(\alpha_k, \hat{\gamma}_k^{(i-1)}) &= \sum_{n=1}^N (a \log \alpha_{k,n} - b \alpha_{k,n}) \\
&\quad - \text{tr} \left(\hat{\gamma}_{k,m}^{(i-1)} \mathbf{D}_k \left(\hat{\Sigma}_{k,m} + \hat{\mu}_{k,m} \hat{\mu}_{k,m}^H \right) \right) \\
&\quad + \sum_{m=1}^M \ln \left| \hat{\gamma}_{k,m}^{(i-1)} \mathbf{D}_k \right| + \text{const.}
\end{aligned} \tag{19}$$

where γ_k has been fixed at its estimation after the last iteration, i.e., $\gamma_k = \hat{\gamma}_k^{(i-1)}$. The first-order derivative of $Q(\alpha_k, \hat{\gamma}_k^{(i-1)})$ w.r.t $\alpha_{k,n}$ can be expressed as:

$$\begin{aligned}
\frac{dQ(\alpha_k, \hat{\gamma}_k^{(i-1)})}{d\alpha_{k,n}} &= \frac{a}{\alpha_{k,n}} - b - \sum_{m=1}^M \hat{\gamma}_{k,m}^{(i-1)} q_{k,m,n} \\
&\quad + M(\omega_{k,n-1} + \omega_{k,n} + \omega_{k,n+1})
\end{aligned} \tag{20}$$

with

$$\begin{aligned}
\omega_{k,n} &= \frac{1}{\alpha_{k,n-1} + \alpha_{k,n} + \alpha_{k,n+1}} \\
q_{k,m,n} &= \sum_{j=-1}^1 \left(\left[\hat{\Sigma}_{k,m} \right]_{n+j,n+j} + |\hat{\mu}_{k,n+j}|^2 \right)
\end{aligned} \tag{21}$$

In (21), the derivative w.r.t. $\alpha_{k,n}$ is correlated with $\alpha_{k,n-2}$, $\alpha_{k,n-1}$, $\alpha_{k,n+1}$, and $\alpha_{k,n+2}$. This makes the closed-form solution for $\alpha_{k,n}$ unavailable. Nevertheless, we can obtain a valid solution using the gradient-based algorithm as follows:

$$\hat{\alpha}_k \leftarrow \hat{\alpha}_k - \delta \nabla_{\alpha_k} Q(\alpha_k, \hat{\gamma}_k^{(i-1)}) \tag{22}$$

where δ is the stepsize, and the gradient can be directly computed using (20).

By virtue of the EM's properties, the above algorithm will always converge since each iteration is guaranteed to reduce the target function. However, the update formula of α_k in (22) is still complex due to the lack of analytic expressions. Moreover, (22) gives little insight into the basic mechanism of SP-SBL. In the following, we derive a closed-form solution for α_k by adding some heuristic assumptions.

When updating $\alpha_{k,n}$, we temporarily assume that the n th element of $\mathbf{x}_{k,m}$ has the same variance with its neighbors, i.e., the $(n-1)$ th and $(n+1)$ th elements of $\mathbf{x}_{k,m}$. In this case, we will have $\alpha_{k,n-2} = \alpha_{k,n-1} = \alpha_{k,n} = \alpha_{k,n+1} = \alpha_{k,n+2}$. This assumption is reasonable if we do not have much prior knowledge about α_k . Note that this does not mean we will obtain an estimation of α_k with $\hat{\alpha}_{k,n-2}^{(i)} = \hat{\alpha}_{k,n-1}^{(i)} = \hat{\alpha}_{k,n}^{(i)} = \hat{\alpha}_{k,n+1}^{(i)} = \hat{\alpha}_{k,n+2}^{(i)}$ because, with the incoming training signal, the estimation of $\alpha_{k,n}$ is rectified and is expected to get close to the true value. Under this assumption, (20) becomes:

$$\frac{dQ(\alpha_k, \hat{\gamma}_k^{(i-1)})}{d\alpha_{k,n}} = \frac{a}{\alpha_{k,n}} - b + \frac{M}{\alpha_{k,n}} - \sum_{m=1}^M \hat{\gamma}_{k,m}^{(i-1)} q_{k,m,n} \tag{23}$$

By setting (23) to zero, we can obtain:

$$\hat{\alpha}_{k,n}^{(i)} \approx (a + M) \left(\sum_{m=1}^M \hat{\gamma}_{k,m}^{(i-1)} q_{k,m,n} + b \right)^{-1} \tag{24}$$

The numerical results show that the update formula in (24) gives rise to the similar performance with that using gradient-based update in (22). Moreover, the solution in (24) provides a clear insight into the difference between SP-SBL and conventional SBL. With conventional SBL, the update of the precision for the n th element of $\mathbf{x}_{k,m}$ is only related to the posterior mean and variance of itself. In contrast, in SP-SBL, the update of $\alpha_{k,n}$ is effected by the posterior means and variances of $(n-1)$ th, n th, and $(n+1)$ th elements of $\mathbf{x}_{k,m}$ for all $m = 1, \dots, M$ due to the utilization of joint and burst sparsity. Therefore, the SP-SBL is expected to achieve better performance.

After the downlink channel estimation, the estimated CSI should be sent back to the BS in order to perform the downlink beamforming. In this process, additional error can be introduced by quantization, noise, and feedback delay. However, the results in [21] showed that the error due to the imperfect feedback can be made much smaller than that due to the estimation error in downlink training phase. Therefore, as in [11], we optimistically neglect the additional error due to the feedback in this paper.

4.4 Computation complexity analysis

The main computation load in each iteration is due to the $N \times N$ matrix inversion when updating the posterior mean and covariance matrix of $\mathbf{x}_{k,m}$ in E-step. By using the matrix inversion lemma, the calculation of each matrix inversion has complexity $\mathcal{O}(T^3)$. Therefore, the overall computational complexity of the algorithm scales with $\mathcal{O}(GKMT^3)$, where G is the number of EM iterations. Since we consider the situation that T is much smaller than N , the computation complexity will not pose a significant problem.

4.5 Extension to FDD massive MIMO system with hybrid analog-digital processing

In the practical system, the hybrid analog-digital processing may be utilized at the BS to reduce the implementation complexity. In the following, we will show that the proposed scheme can still be applied in this case. With hybrid analog-digital processing at the BS, the received training signal at user k can be expressed as:

$$\mathbf{Z}_k = \mathbf{H}_k^H \mathbf{W} \mathbf{S} + \mathbf{N}_k \tag{25}$$

where $\mathbf{W} \in \mathbb{C}^{N \times F}$ is the arbitrary analog beamforming matrix utilized by BS, and $F \leq N$ denotes the number of radio frequency chain at the BS. The pilot signal \mathbf{S} is an $F \times T$ matrix in this case. Note that we have reuse some notations to avoid introducing too many new definitions. Using the definition of angular-domain channel in (4), we can rewrite (25) as:

$$\mathbf{Y}_k = \mathbf{Z}_k^H = (\mathbf{W}\mathbf{S})^H \mathbf{A}_N \mathbf{X}_k + \mathbf{N}_k^T \quad (26)$$

Note that, by treating $(\mathbf{W}\mathbf{S})^H \mathbf{A}_N$ as the sensing matrix, (26) is in the same form with (6). Therefore, the proposed scheme can be utilized directly to estimation the channel matrix from (26).

5 Simulation results and discussion

This section presents the simulation results to validate the proposed scheme. Without loss of the generality, the large-scale fading and noise variance are normalized to 1. For illustration, the number of users is set to $K = 8$. The central frequency is set to 2.4 GHz, and the antenna spacing is equal to the half of wavelength. Similar to [22], the rows of sensing matrix Φ are designed as the length- N Zadoff-Chu sequence [23] with shifting step 7. In particular, the first row of Φ is given by $\frac{1}{\sqrt{N}} \left[1, e^{j\frac{\nu\pi 1^2}{N}}, e^{j\frac{\nu\pi 2^2}{N}}, \dots, e^{j\frac{\nu\pi (N-1)^2}{N}} \right]$. The second row of Φ is obtained by cyclicly right shifting the first row with a step of 7, which is given by $\frac{1}{\sqrt{N}} \left[e^{j\frac{\nu\pi (N-7)^2}{N}}, e^{j\frac{\nu\pi (N-6)^2}{N}}, \dots, e^{j\frac{\nu\pi (N-1)^2}{N}}, 1, e^{j\frac{\nu\pi 1^2}{N}}, e^{j\frac{\nu\pi 2^2}{N}}, \dots, e^{j\frac{\nu\pi (N-8)^2}{N}} \right]$. Moreover, $\nu = 7$ is used in the simulations. The rest rows of Φ are generated in a similar way. The AoD θ_k and AoA φ_k for different users are generated randomly from the interval $[-90^\circ, 90^\circ]$. It is assumed that the waves arriving at the user are uniformly distributed in AoA [12]. Therefore, Δ_a is equal to π , and the complex channel gain reduces to $r_k(\theta, \varphi) = r_k(\theta)$. It is assumed that the $r_k(\theta)$ for different θ are uncorrelated. For each sample of θ , $r_k(\theta)$ is the complex Gaussian distributed with zero mean and variance $U_k(\theta)$, where $U_k(\theta)$ is the power angle spectrum. We model $U_k(\theta)$ as the truncated Laplacian distribution centered at θ_k [12]. The simplified update rule for α_k in (24) is utilized to reduce the complexity. The parameters in the hyperprior models (12) and (13) are set as $a = b = 10^{-4}$, $u_1 = \dots = u_M$, and $u_0 = M$.

We consider four reference schemes in the simulations, i.e., the conventional SBL based on the original algorithm in [17], OMP [15], block l_1/l_2 minimization [18], and variational SBL [19]. Note that the block l_1/l_2 minimization and variational SBL can also exploit the joint sparsity among the angular-domain channel vectors between BS and different antennas of user.

We consider the NMSE performance which is defined as follows:

$$\text{NMSE} = \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \frac{\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2}{\|\mathbf{H}_k\|^2} \right], \quad (27)$$

All figures are obtained by averaging the results for 10^3 independent channel realizations.

Figure 2 shows the NMSE performance for different lengths of pilot sequence T , where the transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB and the angular spread is $\Delta_d = 5^\circ$. It is seen that the performances of conventional SBL and OMP are poor when the length of pilot sequence is small. By exploiting the joint sparsity among the angular-domain channel vectors between BS and different antennas of user, the block l_1/l_2 minimization and variational SBL can provide better performance. Moreover, the proposed scheme based on SP-SBL achieves the best NMSE performance since it exploits the joint sparsity and burst sparsity simultaneously.

Figure 3 shows the NMSE performance for different transmit SNRs of BS, where the length of pilot sequence is $T = 16$ and the angular spread is $\Delta_d = 5^\circ$. It is seen that the performances of all schemes converge to NMSE floors for large SNR. The reason is that the angular-domain channel is approximately sparse as discussed in Section 4.1. That is, the n th element of $\mathbf{x}_{k,m}$ with $n \notin \Omega_k$ is small but not zero. With small T , recovering all these small elements of $\mathbf{x}_{k,m}$ with high accuracy is difficult. As a result, with the increasing of the SNR, the effect of the mismatch between the true values of these small elements and their estimates becomes significant when compared with AWGN, which incurs NMSE floor in the high SNR region¹. Note that the NMSE floor does not occur in

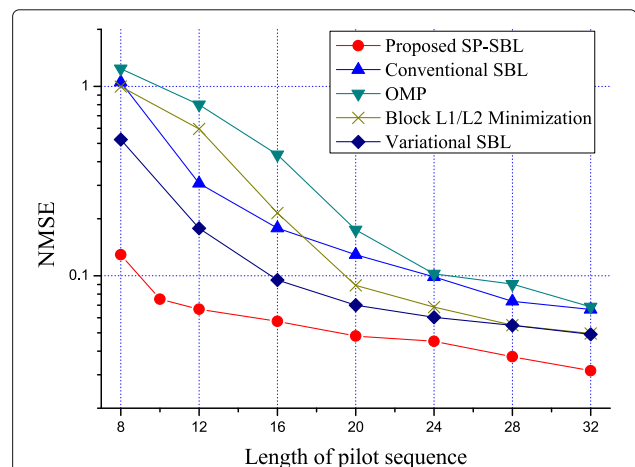


Fig. 2 NMSE performance for different lengths of pilot sequence T , where $N = 64$ and $M = 4$. The transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB and the angular spread is $\Delta_d = 5^\circ$

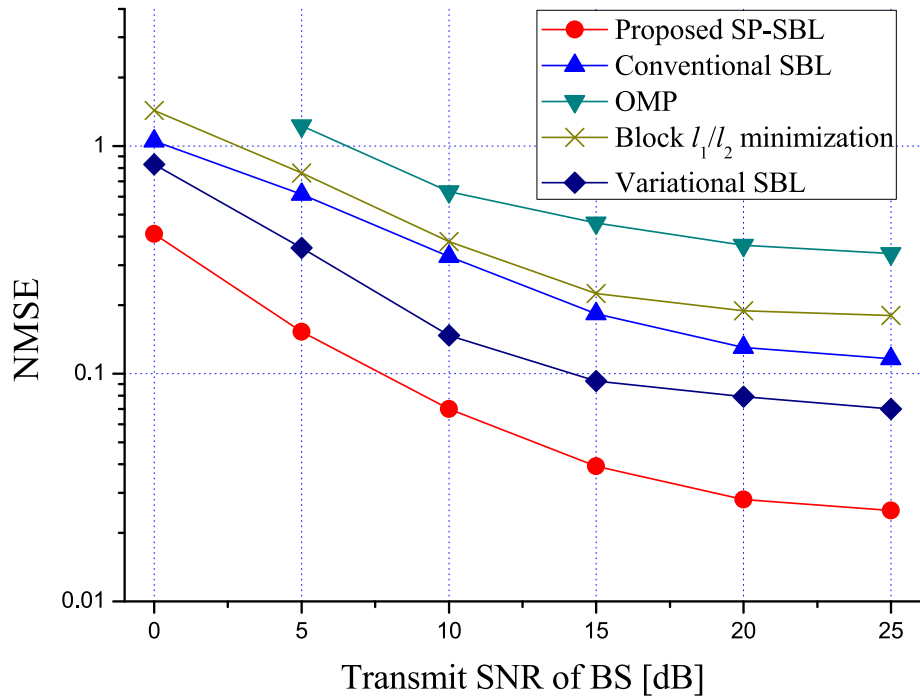


Fig. 3 NMSE performance for different transmit SNRs of BS, where $N = 64$ and $M = 4$. The length of pilot sequence is $T = 16$ and the angular spread is $\Delta_d = 5^\circ$

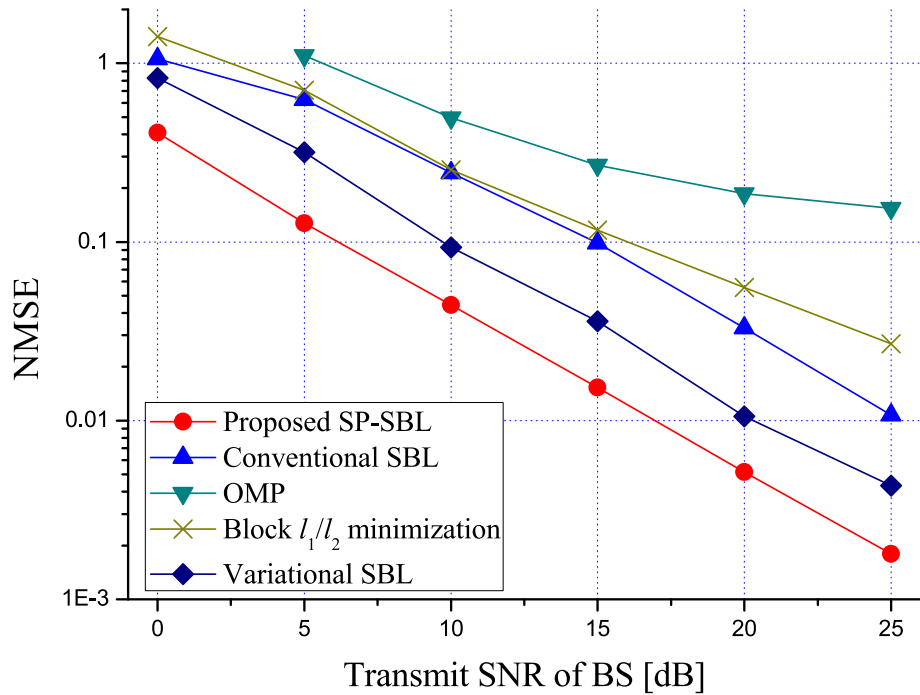


Fig. 4 NMSE performance for different transmit SNRs of BS, where $N = 64$ and $M = 4$. The length of pilot sequence is $T = 16$ and the angular spread is $\Delta_d = 5^\circ$. The angular-domain channel is assumed exactly sparse

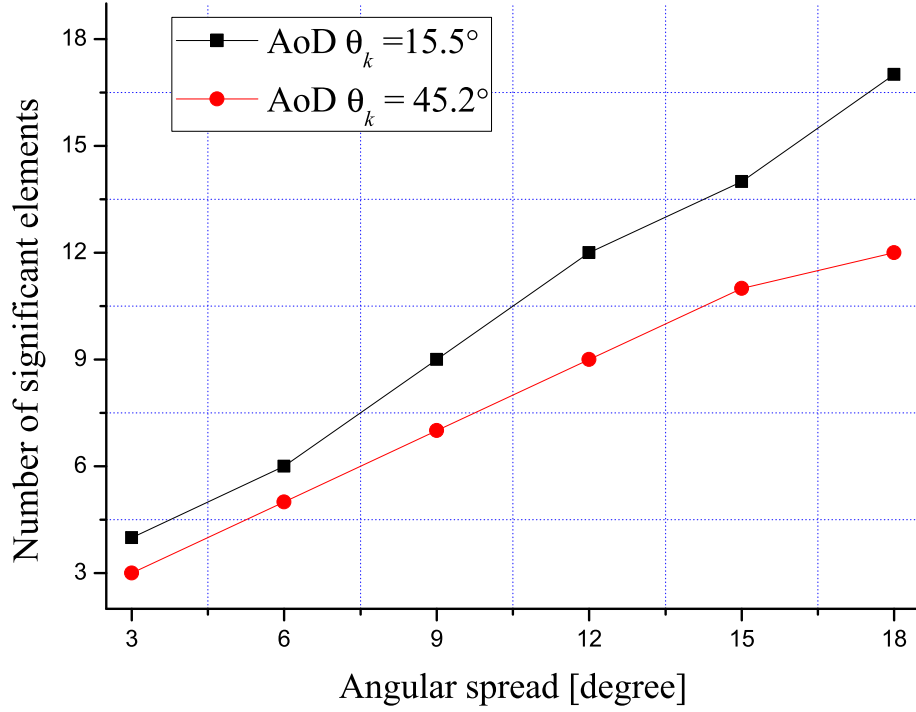


Fig. 5 Number of significant elements in $\mathbf{x}_{k,m}$, where $N = 64$ and $M = 4$

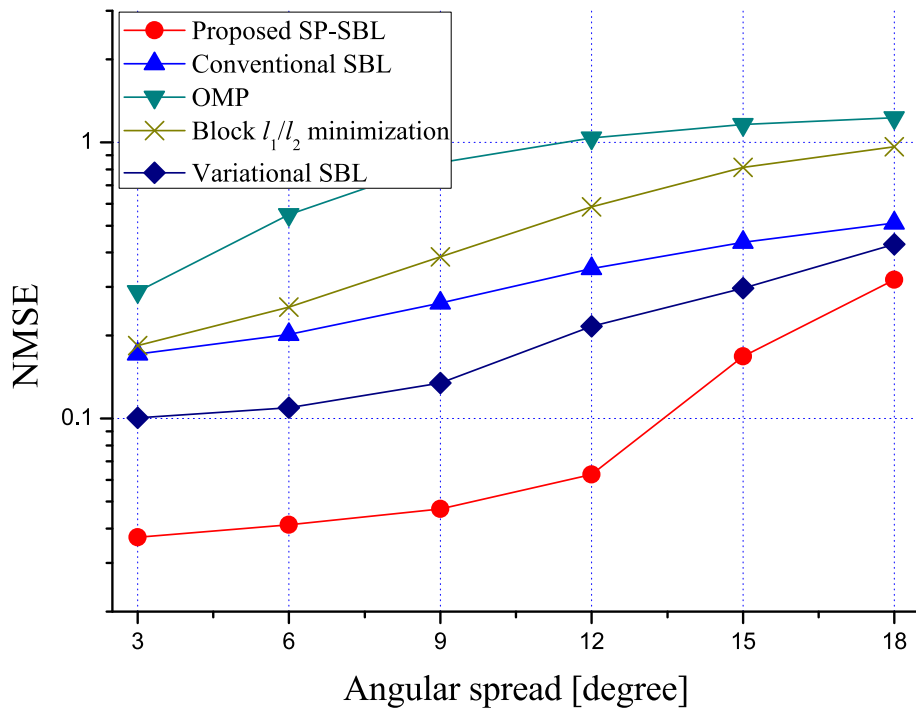


Fig. 6 NMSE performance for different angular spreads Δ_d , where $N = 64$ and $M = 4$. The length of pilot sequence is $T = 16$ and the transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB

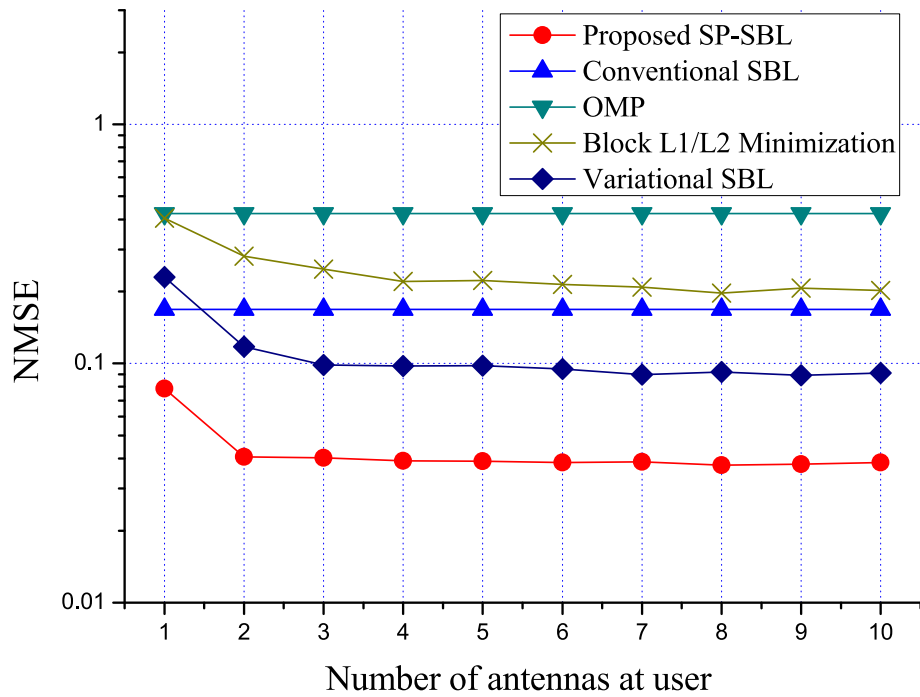


Fig. 7 NMSE performance for different numbers of antennas at users, i.e., M , where $N = 64$. The length of pilot sequence is $T = 16$. Transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB. The angular spread is $\Delta_d = 5^\circ$

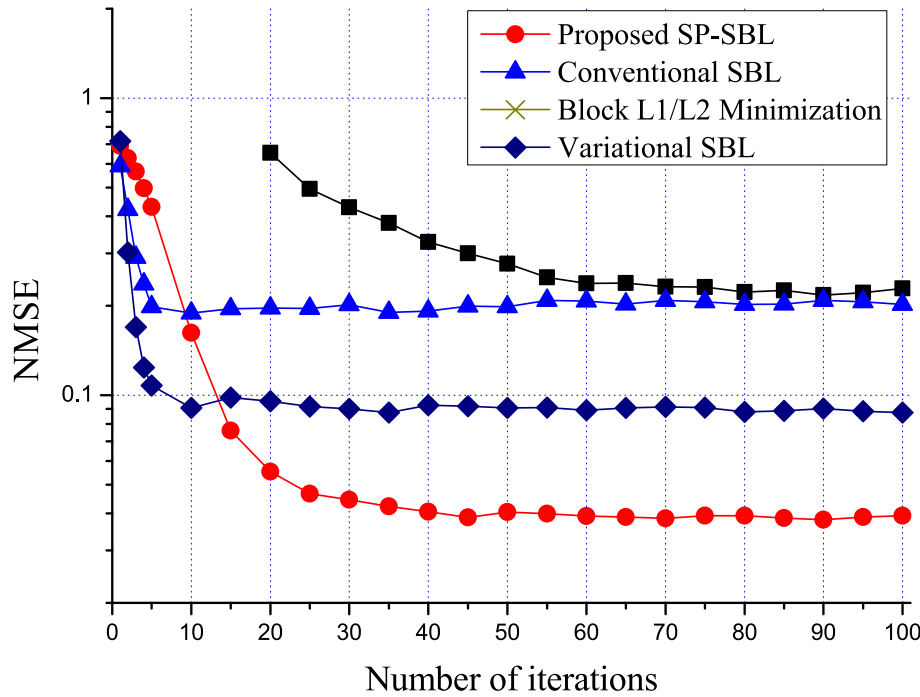


Fig. 8 NMSE performance for different numbers of iterations, where $N = 64$. The length of pilot sequence is $T = 16$. Transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB. The angular spread is $\Delta_d = 5^\circ$

the simulations of [16, 18]. This is because these works assume that there is no channel power leakage outside Ω_k . Although this assumption simplifies the model, the resultant simulations cannot reflect the real performance in the high SNR region very well. For illustration, we also present the NMSE performance for all schemes in Fig. 4 under exactly sparse angular-domain channel model considered in [16, 18], where the n th element of $\mathbf{x}_{k,m}$ is set to zero as long as $n \notin \Omega_k$ when generating the channel matrix. From the figure, we can see that the performance floors disappear as expected.

Then, we consider the performance for different angular spreads Δ_d ranged from 3 to 18°. This corresponds to the scenarios with scattering ring of 30 m and the distance between BS and user varying from about 100 to 500 m. Figure 5 shows the number of significant elements in the angular-domain channel $\mathbf{x}_{k,m}$, i.e., the number of elements that contain 90% of the channel power, varies from 3 to 17 when the angular spread increases from 3 to 18°. Figure 6 shows the NMSE performance for different angular spreads, where the transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB and the length of pilot sequence is $T = 16$. Again, the proposed scheme based on SP-SBL achieves the best performance. Moreover, we note that the NMSEs of all schemes degrade when the angular spread increases. This is because the number of significant elements in angular-domain channel becomes larger as shown in Fig. 5. In this case, to maintain the estimation performance, the capacity of all schemes must be enhanced by increasing the training samples.

Figure 7 shows the NMSE performance for different numbers of antennas at user, i.e., M . The length of pilot sequence is $T = 16$. The transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB. The angular spread is $\Delta_d = 5^\circ$. It is seen that the NMSEs of conventional SBL and OMP-based schemes are independent of M . In contrast, the performances of block l_1/l_2 minimization, variational SBL, and SP-SBL-based schemes are improved gradually with the increasing of M , which is just the benefit of exploiting the joint sparsity.

Figure 8 shows the NMSE performance for different numbers of iterations. The length of pilot sequence is $T = 16$. The transmit SNR of BS is $\frac{NP}{\sigma^2} = 15$ dB. The angular spread is $\Delta_d = 5^\circ$. It is seen that the NMSE performance converges after 40 EM iterations. Moreover, the number of iterations required for SP-SBL is smaller than that of block l_1/l_2 minimization and greater than that of conventional SBL and variational SBL. The computation complexities in each iteration for SP-SBL, block l_1/l_2 minimization, conventional SBL, and variational SBL are $\mathcal{O}(KMT^3)$, $\mathcal{O}(KMNT)$ [18], $\mathcal{O}(KMT^3)$ [17], and $\mathcal{O}(KMT^3)$ [19], respectively. Note that the OMP needs no iteration and its total computation complexity is $\mathcal{O}(\eta KMNT)$ [15], where η denotes the number of significant elements in $\mathbf{x}_{k,m}$. Therefore, from the analysis above,

the SP-SBL has the highest computation complexity but converges to best NMSE.

6 Conclusions

This paper proposes a SP-SBL approach for downlink channel estimation in FDD massive MIMO system. By exploiting the two kinds of sparse structures, the scheme can substantially reduce the pilot resource assumption while maintaining the estimation performance. Through numerical simulations, it is shown that the proposed scheme outperforms the reference schemes significantly in terms of NMSE for a variety of scenarios with different lengths of pilot sequence, transmit SNRs, and angular spreads.

Endnote

¹To address this problem, a possible solution is to exploit other inherent structures of angular-domain channel in the prior model, for example, the additional correlation between the phases and amplitudes of the elements in angular-domain matrix if they exist. In general, the problem is quite challenging and will be considered as future study.

Abbreviations

AoA: Angle of arrival; AoD: Angle of departure; AWGN: Additive white Gaussian noise; BS: Base station; CSI: Channel state information; DFT: Discrete Fourier transform; EM: Expectation maximization; FDD: Frequency division duplex; IoT: Internet of Things; LS: Least square; MIMO: Multiple-input multiple-output; MMSE: Minimum mean square error; NMSE: Normalized mean square error; OMP: Orthogonal matching pursuit; SBL: Sparse Bayesian learning; SNR: Signal-to-noise ratio; SP-SBL: Structured prior-based sparse Bayesian learning; ULA: Uniform linear array

Funding

This work is supported by National Natural Science Foundation of China (No. 61671472) and Jiangsu Province Natural Science Foundation (BK20160079).

Availability of data and materials

Not applicable.

Authors' contributions

XX provided the main idea and is responsible for the manuscript writing. YW conducted the simulations. KX, WX, and MW assisted with the revision of the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 December 2018 Accepted: 18 January 2019

Published online: 05 February 2019

References

1. J. A. Stankovic, Research directions for the Internet of Things. *IEEE Internet Things J.* **1**(1), 3-9 (2014)
2. T. Lv, Z. Lin, P. Huang, J. Zeng, Optimization of the energy-efficient relay-based massive IoT network. *IEEE Internet Things J.* **5**(4), 3043-3058 (2018)

3. C. Li, S. Zhang, P. Liu, F. Sun, J. M. Cioffi, L. Yang, Overhearing protocol design exploiting intercell interference in cooperative green networks. *IEEE Trans. Veh. Technol.* **65**(1), 441–446 (2016)
4. T. L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wirel. Commun.* **9**(11), 3590–3600 (2010)
5. C. Li, L. Yang, W. Zhu, Two-way, MIMO relay precoder design with channel state information. *IEEE Trans. Commun.* **58**(12), 3358–3363 (2010)
6. C. Li, H. J. Yang, F. Sun, J. M. Cioffi, L. Yang, Multiuser overhearing for cooperative two-way multiantenna relays. *IEEE Trans. Veh. Technol.* **65**(5), 3796–3802 (2016)
7. C. Li, P. Liu, C. Zou, F. Sun, J. M. Cioffi, L. Yang, Spectral-efficient cellular communications with coexistent one- and two-hop transmissions. *IEEE Trans. Veh. Technol.* **65**(8), 6765–6772 (2016)
8. A. Adhikary, J. Nam, J.-Y. Ahn, G. Caire, Joint spatial division and multiplexing: the large-scale array regime. *IEEE Trans. Info. Theory.* **59**(10), 6441–6463 (2013)
9. C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, C. Xiao, Beam division multiple access transmission for massive MIMO communications. *IEEE Trans. Commun.* **63**(6), 2170–2184 (2015)
10. H. Xie, F. Gao, S. Zhang, S. Jin, A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model. *IEEE Trans. Veh. Technol.* **66**(4), 3170–3184 (2017)
11. X. Xia, K. Xu, D. Zhang, Y. Xu, Y. Wang, Beam-domain full-duplex massive MIMO: realizing co-time co-frequency uplink and downlink transmission in the cellular system. *IEEE Trans. Veh. Technol.* **66**(10), 8845–8862 (2017)
12. L. You, X. Gao, X. Xia, N. Ma, Y. Peng, Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels. *IEEE Trans. Wireless Commun.* **14**(6), 3352–3366 (2015)
13. L. You, X. Gao, A. L. Swindlehurst, W. Zhong, Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots. *IEEE Trans. Signal Process.* **64**(6), 1461–1476 (2016)
14. L. Zhao, D. W. K. Ng, J. Yuan, Multi-user precoding and channel estimation for hybrid millimeter wave systems. *IEEE J. Sel. Areas Commun.* **35**(7), 1576–1590 (2017)
15. J. Lee, G. T. Gil, Y. H. Lee, Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. *IEEE Trans. Commun.* **64**(6), 2370–2386 (2016)
16. X. Rao, V. K. N. Lau, Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. Signal Process.* **62**(12), 3261–3271 (2014)
17. S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing. *IEEE Trans. Signal Process.* **56**(6), 2346–2356 (2008)
18. C. C. Tseng, J. Y. Wu, T. S. Lee, Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block ℓ_1 -minimization. *IEEE Trans. Commun.* **64**(3), 1055–1067 (2016)
19. X. Cheng, J. Sun, S. Li, Channel estimation for FDD multi-user massive MIMO: a variational Bayesian inference-based approach. *IEEE Trans. Wirel. Commun.* **16**(11), 7590–7602 (2017)
20. A. Dempster, Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B.* **39**(1), 1–38 (1977)
21. G. Caire, N. Jindal, M. Kobayashi, N. Ravindran, Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Trans. Inf. Theory.* **56**(6), 2845–2866 (2010)
22. Y. Ding, S. E. Chiu, B. Rao, Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs. *IEEE J. Sel. Top. Sig. Process.* **12**(3), 499–513 (2018)
23. D. Chu, Polyphase codes with good periodic correlation properties. *IEEE Trans. Inf. Theory.* **18**(4), 531–532 (1972)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
