Received March 20, 2019, accepted April 5, 2019, date of publication April 11, 2019, date of current version April 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910088

Bayesian mmWave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures

KAIHUI LIU[1], XINGJIAN LI[1], JUN FANG[1], (Member, IEEE), AND HONGBIN LI[2], (Fellow, IEEE)

[1]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Jun Fang (junfang@uestc.edu.cn)

This work was supported in part by the National Science Foundation of China, under Grant 61829103 and Grant 61871091.

ABSTRACT We consider the problem of channel estimation for millimeter wave (mmWave) systems, where both the base station and the mobile station employ a single radio frequency (RF) chain to reduce the hardware cost and power consumption. Recent real-world channel measurements reveal that the mmWave channels incur a certain amount of spread over the angular domains due to the scattering clusters. The angular spreads give rise to a joint sparse and low-rank channel matrix in the angular domain. To utilize this joint sparse and low-rank structure, we address the channel estimation problem within a Bayesian framework. Specifically, we adopt a matrix factorization formulation and translate the problem of channel estimation into one of searching for two-factor matrices. To encourage a joint sparse and low-rank solution, independent sparsity-promoting priors are placed on entries of the two-factor matrices, which aims to promote sparse factor matrices with only a few non-zero columns. Based on the proposed prior model, we develop a variational Bayesian inference method for the mmWave channel estimation. The simulation results show that our proposed method presents a considerable performance improvement over the state-of-the-art compressed sensing-based channel estimation methods.

INDEX TERMS mmWave channel estimation, angular spread, joint sparse and low-rank, compressed sensing.

I. INTRODUCTION

Millimeter-wave (mmWave) communication is a promising technology to cope with the ever-increasing need for bandwidth and capacity in future wireless networks [1]–[4]. Due to the high attenuation and severe signal absorption at the mmWave frequency bands, mmWave communication systems have to employ large antenna arrays at both the base station and the mobile station, and exploit beam steering to provide an adequate link budget [5], [6]. In this setup, accurate channel estimation is essential for the proper operation of directional precoding/beamforming in mmWave systems. Nevertheless, channel estimation in mmWave systems is challenging due to the use of hybrid precoding structures and the large number of antennas. In particular, hybrid precoding structures employed in mmWave systems prevent the digital baseband from directly accessing the entire channel dimension, which is referred to as the channel subspace sampling limitation [7].

To address this difficulty, most previous studies exploited the sparse scattering nature of mmWave channels and formulated mmWave channel estimation as a compressed sensing problem [7]–[15]. These methods can help achieve substantial reduction of the training overhead. To overcome the grid mismatch issue arising from conventional compressed sensing techniques, super-resolution (off-grid) compressed sensing methods were developed to improve the channel estimation accuracy [16], [17]. In addition to the sparse scattering characteristic, several real-world measurement campaigns in dense-urban propagation environments suggest that mmWave channels spread in the form of clusters of paths over the angular domains [18]–[21]. This angular spread naturally results in a structured sparsity pattern that can be exploited to enhance the estimation performance [22]. Furthermore, it was pointed out in [13] that, due to the spatial correlation and the unsymmetric angular spreads over different domains,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Xu.

mmWave channels may exhibit a useful joint sparse and low-rank structure whereby the rank is far smaller than the sparsity level of the channel. To utilize this joint sparse and low-rank structure, a two-stage compressed sensing scheme was proposed and it was shown that the proposed scheme achieves a lower sample complexity than a conventional compressed sensing method that exploits only the sparse structure of mmWave channels [13].

In this paper, we continue the effort towards developing efficient channel estimation schemes for mmWave communications. To exploit the joint sparse and low-rank structure of mmWave channels, we cast the channel estimation problem within a Bayesian framework. Specifically, we adopt a matrix factorization formulation which converts channel estimation into a problem of searching for these two factor matrices. Independent sparsity-promoting priors are placed on entries of the two factor matrices, which helps promote sparse factor matrices with only a few non-zero columns and in turn encourage a simultaneously sparse and low-rank channel structure. Based on the proposed prior model, we develop a variational Bayesian inference method for mmWave channel estimation.

It should be noted that although the estimation of low-rank matrices or sparse matrices from compressed linear measurements has been extensively studied in various settings [23]–[27], much less research has addressed the case where the matrix of interest is characterized by simultaneous low-rank and sparse structures. A recent study reveals that a convex formulation combining the ℓ_1 norm with the nuclear norm to exploit both types of structures may not perform better than exploiting only one type of the structures [28]. To break the sample complexity barrier of simultaneously sparse and low-rank matrix estimation, a two-stage convex method was developed by assuming a nested structure of the measurement operator [29]. Inspired by [29], a variant of the two-stage scheme was developed for mmWave channel estimation, which has been shown to achieve a near-optimal sample complexity [13]. Recently, a nonconvex method, named sparse power factorization (SPF), was proposed for joint sparse and low-rank matrix estimation [30]. It decomposes the matrix as a product of two factor matrices and then applies the alternating minimization scheme over the factor matrices. Although SPF was shown to have an optimal sample complexity, it requires an accurate knowledge of the rank of the underlying matrix to be estimated, which is usually unavailable for practical problems. In contrast to the SPF, our proposed Bayesian method can automatically infer the rank of the matrix of interest. To our best knowledge, our work presents the first attempt to address the joint sparse and low-rank matrix estimation problem from a Bayesian framework.

The rest of the paper is organized as follows. The system and the channel model are discussed in Section II. In Section III, we propose a hierarchical Gaussian prior model to capture the joint sparse and low-rank structure of the underlying mmWave channel. A variational Bayesian method is developed in Section IV for mmWave channel estimation. Simulation results are provided in Section V, followed by concluding remarks in Section VI.

II. SYSTEM MODEL

Consider a point-to-point mmWave system consisting of a transmitter (e.g. the mobile station) and a receiver (e.g. the base station), where the transmitter is equipped with N_t antennas and the receiver is equipped with N_r antennas. To reduce the hardware cost and power consumption, both the transmitter and the receiver employ a single RF chain for transmit beamforming or receive combining. At time instant t, suppose the transmitter uses $\boldsymbol{f}(t) \in \mathbb{C}^{N_t}$ as the beamforming vector, and the receiver employs $\boldsymbol{z}(t) \in \mathbb{C}^{N_r}$ to combine the received signal. The received signal at the receiver can be expressed as

$$y(t) = \boldsymbol{z}^H(t)\boldsymbol{H}\boldsymbol{f}(t)s(t) + w(t) \quad \forall t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $s(t)$ is the transmitted symbol, and $w(t)$ denotes the additive complex Gaussian noise with zero mean and variance σ^2. For simplicity, the transmitted symbol is set to $s(t) = 1$. Since $\boldsymbol{z}(t)$ and $\boldsymbol{f}(t)$ are implemented using analog phase shifters, their entries are of constant modulus. The problem of interest is to estimate the channel matrix $\boldsymbol{H}$ from the received signal $\{y(t)\}$. In particular, we wish to obtain a reliable channel estimate by using as few measurements as possible.

A conventional approach is to exploit the sparse scattering property of mmWave channels and formulate channel estimation as a sparse signal recovery problem. The mmWave channel is typically modeled as [7]

$$\boldsymbol{H} = \sum_{l=1}^{\tilde{L}} \alpha_l \boldsymbol{a}_r(\theta_l)\boldsymbol{a}_t^H(\phi_l) \tag{2}$$

where $\tilde{L}$ is the number of paths, α_l is the complex gain of the lth path, $\theta_l \in [0, 2\pi]$ and $\phi_l \in [0, 2\pi]$ are the associated azimuth AoA and azimuth AoD respectively, and $\boldsymbol{a}_r \in \mathbb{C}^{N_r}$ ($\boldsymbol{a}_t \in \mathbb{C}^{N_t}$) is the array response vector associated with the receiver (transmitter). In this paper, we assume that uniform linear arrays (ULA) are used at the transmitter and receiver. Due to the sparse scattering characteristic, the mmWave channels have a sparse representation in the beam space domain, i.e.

$$\boldsymbol{H} = \boldsymbol{A}_r \boldsymbol{H}_v \boldsymbol{A}_t^H \tag{3}$$

where $\boldsymbol{A}_r \triangleq [\boldsymbol{a}_r(\psi_1) \ \ldots \ \boldsymbol{a}_r(\psi_{N_1})]$ is an overcomplete matrix ($N_1 \geq N_r$) with each column a steering vector parameterized by a pre-discretized AoA, $\boldsymbol{A}_t \triangleq [\boldsymbol{a}_t(\omega_1) \ \ldots \ \boldsymbol{a}_t(\omega_{N_2})]$ is an overcomplete matrix (i.e. $N_2 \geq N_t$) with each column a steering vector parameterized by a pre-discretized AoD, and $\boldsymbol{H}_v \in \mathbb{C}^{N_1 \times N_2}$ is a sparse matrix with $\tilde{L}$ non-zero entries corresponding to the channel path gains $\{\alpha_l\}$.

In addition to the sparse scattering characteristic, real-world measurement campaigns in dense-urban propagation

environments (e.g. [18]–[21]) reveal that mmWave channels spread in the form of clusters of paths over the angular domains including the angle of arrival (AoA), angle of departure (AoD), and elevation. As shown in [13], in the presence of angular spreads, the mmWave channel exhibits a simultaneously sparse and low-rank structure in which the rank is far smaller than the sparsity level of the channel. Specifically, due to the spatial correlation observed from real-world channel measurements, the mmWave channel can be approximately modeled as [13]

$$\boldsymbol{H}=\sum_{l=1}^{L}\Big(\sum_{i=1}^{I}\alpha_{l,i}\boldsymbol{a}_r(\theta_l-\vartheta_{l,i})\Big)\Big(\sum_{j=1}^{J}\beta_{l,j}\boldsymbol{a}_t^H(\phi_l-\varphi_{l,j})\Big) \quad (4)$$

where L is the number of clusters, θ_l and ϕ_l denote the mean AoA and AoD of each cluster, respectively, and $\vartheta_{l,i}$ and $\varphi_{l,j}$ represent the relative AoA and AoD shift from the mean angle, respectively. Accordingly, the mmWave channel matrix and its beam space version are related as

$$\begin{aligned}\boldsymbol{H} &= \sum_{l=1}^{L}\boldsymbol{A}_r\boldsymbol{a}_l\boldsymbol{b}_l^T\boldsymbol{A}_t^H=\boldsymbol{A}_r\Big(\sum_{l=1}^{L}\boldsymbol{a}_l\boldsymbol{b}_l^T\Big)\boldsymbol{A}_t^H\\ &\triangleq \boldsymbol{A}_r\boldsymbol{H}_v\boldsymbol{A}_t^H\end{aligned} \quad (5)$$

where $\boldsymbol{a}_l\in\mathbb{C}^{N_1}$ and $\boldsymbol{b}_l\in\mathbb{C}^{N_2}$ are the virtual representation over the AoA and AoD domain, respectively. Since the angular spread occupies only a small portion of the whole angular domain, both $\boldsymbol{a}_l$ and $\boldsymbol{b}_l$ are sparse vectors with only a few nonzero entries concentrated around the mean AoA and AoD associated with the lth cluster. Also, notice that $\boldsymbol{H}_v$ has a low-rank structure with $\text{rank}(\boldsymbol{H}_v)=L$. Hence the beam space channel $\boldsymbol{H}_v$ has a simultaneously low-rank and sparse structure in which its rank is far smaller than its number of nonzero entries.

III. PROPOSED HIERARCHICAL PRIOR MODEL

To exploit the simultaneously sparse and low-rank structure of the underlying mmWave channel, we, in the following, employ a matrix factorization formulation and enforce sparsity constraints on the factor matrices. The beam space channel matrix $\boldsymbol{H}_v$ is expressed as a product of two factor matrices:

$$\boldsymbol{H}_v=\sum_{n=1}^{N}\boldsymbol{c}_n\boldsymbol{d}_n^T=\boldsymbol{C}\boldsymbol{D}^T \quad (6)$$

where $\boldsymbol{C}\triangleq[\boldsymbol{c}_1\ \ldots\ \boldsymbol{c}_N]\in\mathbb{C}^{N_1\times N}$, $\boldsymbol{D}\triangleq[\boldsymbol{d}_1\ \ldots\ \boldsymbol{d}_N]\in\mathbb{C}^{N_2\times N}$, $N\triangleq\min\{N_1,N_2\}$. To encourage a joint sparse and low-rank structure, we place independent sparsity-promoting priors on entries of the factor matrices $\boldsymbol{C}$ and $\boldsymbol{D}$, which aims to yield sparse factor matrices with only a few nonzero columns.

Here we employ a two-layer hierarchical Gaussian-inverse Gamma prior model to promote sparse factor matrices. Let

$$\begin{aligned}\boldsymbol{c}&\triangleq\text{vec}(\boldsymbol{C})=[\boldsymbol{c}_1^T\ \ldots\ \boldsymbol{c}_N^T]^T\\ \boldsymbol{d}&\triangleq\text{vec}(\boldsymbol{D})=[\boldsymbol{d}_1^T\ \ldots\ \boldsymbol{d}_N^T]^T\end{aligned}$$

In the first layer, the entries of $\boldsymbol{c}$ and $\boldsymbol{d}$ are assumed to follow a Gaussian distribution

$$p(\boldsymbol{c}|\boldsymbol{\alpha})\sim\prod_{i=1}^{N_1N}\mathcal{CN}\big(c_i;0,\alpha_i^{-1}\big) \quad (7)$$

$$p(\boldsymbol{d}|\boldsymbol{\beta})\sim\prod_{j=1}^{N_2N}\mathcal{CN}\big(d_j;0,\beta_j^{-1}\big) \quad (8)$$

where $\boldsymbol{\alpha}\triangleq\{\alpha_i\}$ and $\boldsymbol{\beta}\triangleq\{\beta_j\}$ are non-negative hyperparameters controlling the sparsity of the entries in $\boldsymbol{c}$ and $\boldsymbol{d}$, respectively. The second layer specifies Gamma distributions as hyperpriors over the hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, i.e.

$$\begin{aligned}p(\boldsymbol{\alpha})&=\prod_{i=1}^{N_1N}\text{Gamma}(\alpha_i|a,b)=\prod_{i=1}^{N_1N}\Gamma(a)^{-1}b^a\alpha_i^{a-1}e^{-b\alpha_i}\\ p(\boldsymbol{\beta})&=\prod_{i=1}^{N_2N}\text{Gamma}(\beta_i|a,b)=\prod_{i=1}^{N_2N}\Gamma(a)^{-1}b^a\beta_i^{a-1}e^{-b\beta_i}\end{aligned} \quad (9)$$

where $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$ is the Gamma function, a and b are set to small values, e.g., 10^{-10}, effectively making the Gamma distribution a non-informative prior.

We assume entries of the additive noise $\{w(t)\}$ are independent and identically distributed (i.i.d.) random variables following a Gaussian distribution with zero mean and variance $\sigma^2=\gamma^{-1}$. To learn γ, a Gamma hyperprior is placed over γ, i.e.

$$p(\gamma)=\text{Gamma}(\gamma|e,f)=\Gamma(e)^{-1}f^e\gamma^{e-1}e^{-f\gamma} \quad (10)$$

The parameters e and f are set to small values, e.g. 10^{-10}. For clarity, the proposed hierarchical prior model is depicted in Fig. 1.

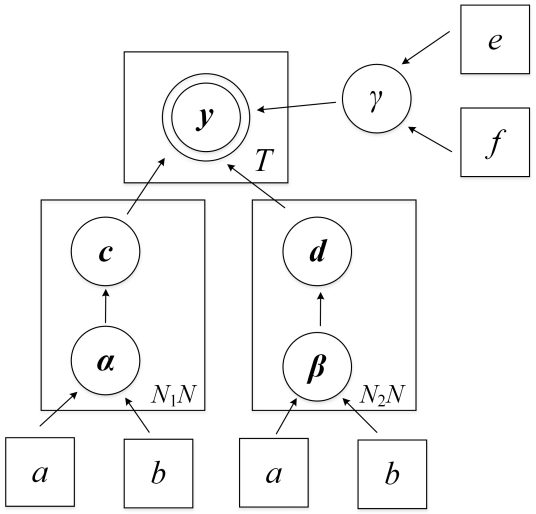


FIGURE 1. Proposed hierarchical Gaussian prior model, in which double circles denote the observable variable, single circles denote the hidden variable, and the boxes denote pre-specified hyperparameters.

To understand why our proposed prior model has the potential to encourage a low-rank structure, note that even if both

$\boldsymbol{c}_n$ and $\boldsymbol{d}_n$ are nonzero sparse vectors, their out-product $\boldsymbol{c}_n\boldsymbol{d}_n^T$ is equal to zero as long as their support sets are disjoint, which is very likely the case since independent sparsity priors are placed on entries of $\boldsymbol{c}_n$ and $\boldsymbol{d}_n$. Note that our prior model is different from the prior model proposed in [26]. In [26], the sparsity constraint is placed on the columns, instead of individual entries, of the factor matrices to promote structured-sparse factor matrices consisting of only a few non-zero columns. As a result, the prior leads to a low-rank but not necessarily a sparse matrix.

IV. VARIATIONAL BAYESIAN INFERENCE

As discussed earlier, our objective is to estimate the mmWave channel matrix $\boldsymbol{H}$ from the received signal $\{y(t)\}$. Let $\tilde{\boldsymbol{z}}(t) \triangleq \boldsymbol{A}_r^H \boldsymbol{z}(t)$ and $\tilde{\boldsymbol{f}}(t) \triangleq \boldsymbol{A}_t^H \boldsymbol{f}(t)$, the received signal $y(t)$ can be expressed as

$$y(t) = \tilde{\boldsymbol{z}}^H(t)\boldsymbol{H}_v\tilde{\boldsymbol{f}}(t) + w(t) \quad \forall t = 1, \ldots, T \tag{11}$$

We now develop a variational Bayesian method for estimating $\boldsymbol{H}_v$. Let $\boldsymbol{y} \triangleq \{y(t)\}$ denote the observed data, and $\boldsymbol{z} \triangleq \{\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma\}$ denote the hidden variables in our graphical model. The rationale of variational Bayesian is to obtain $q(\boldsymbol{z})$, an approximate of the posterior distribution $p(\boldsymbol{z}|\boldsymbol{y})$, via maximizing the evidence lower bound defined as

$$L(q) = \int q(\boldsymbol{z}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{z})}{q(\boldsymbol{z})} d\boldsymbol{z} \tag{12}$$

To maximize $L(q)$ with respect to $q(\boldsymbol{z})$, a factorized form of $q(\boldsymbol{z})$ over the component variables in $\boldsymbol{z}$ is usually assumed [31], i.e.

$$q(\boldsymbol{z}) = q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma) \tag{13}$$

With such a factorized form, we can maximize the evidence lower bound $L(q)$ in an alternating fashion, which yields

$$\begin{aligned}
\ln q_c(\boldsymbol{c}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\
\ln q_d(\boldsymbol{d}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_c(\boldsymbol{c})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\
\ln q_\alpha(\boldsymbol{\alpha}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\
\ln q_\beta(\boldsymbol{\beta}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)} + \text{const} \\
\ln q_\gamma(\gamma) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})} + \text{const}
\end{aligned} \tag{14}$$

where the "const" in (14) are normalizing constants to make sure the terms on the left-hand side of (14) are probability density functions. Details of the Bayesian inference are provided next.

Update of $q_c(\boldsymbol{c})$: Before proceeding, we substitute the factorized form of $\boldsymbol{H}_v$ (6) into (11) and express $y(t)$ as

$$\begin{aligned}
y(t) &= \tilde{\boldsymbol{z}}^H(t)\boldsymbol{H}_v\tilde{\boldsymbol{f}}(t) + w(t) \\
&= \tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}\boldsymbol{D}^T\tilde{\boldsymbol{f}}(t) + w(t) \\
&= \left(\tilde{\boldsymbol{f}}^T(t)\boldsymbol{D}\right) \otimes \tilde{\boldsymbol{z}}^H(t)\Big) \operatorname{vec}(\boldsymbol{C}) + w(t) \\
&= (\tilde{\boldsymbol{f}}^T(t) \otimes \tilde{\boldsymbol{z}}^H(t))(\boldsymbol{D} \otimes \boldsymbol{I}_{N_1})\boldsymbol{c} + w(t)
\end{aligned} \tag{15}$$

Collecting all measurements $\{y(t)\}_{t=1}^T$ and stacking them into a vector, we have

$$\begin{aligned}
\boldsymbol{y} &= \begin{bmatrix} \tilde{\boldsymbol{f}}^T(1) \otimes \tilde{\boldsymbol{z}}^H(1) \\ \vdots \\ \tilde{\boldsymbol{f}}^T(T) \otimes \tilde{\boldsymbol{z}}^H(T) \end{bmatrix} (\boldsymbol{D} \otimes \boldsymbol{I}_{N_1})\boldsymbol{c} + \boldsymbol{w} \\
&\triangleq \boldsymbol{\Phi}_{\boldsymbol{c}}\boldsymbol{c} + \boldsymbol{w}
\end{aligned} \tag{16}$$

where $\boldsymbol{w} \triangleq [w(1), \ldots, w(T)]^T$.

The approximate posterior distribution of $q_c(\boldsymbol{c})$ can be calculated as

$$\begin{aligned}
\ln q_c(\boldsymbol{c}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\
&= \langle \ln p(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{d}, \gamma) \rangle + \langle \ln p(\boldsymbol{c}|\boldsymbol{\alpha}) \rangle + \text{const} \\
&= -\langle \gamma \rangle \Big\langle (\boldsymbol{y} - \boldsymbol{\Phi}_{\boldsymbol{c}}\boldsymbol{c})^H (\boldsymbol{y} - \boldsymbol{\Phi}_{\boldsymbol{c}}\boldsymbol{c}) \Big\rangle - \boldsymbol{c}^H \langle \boldsymbol{A} \rangle \boldsymbol{c} + \text{const} \\
&= -(\boldsymbol{c} - \boldsymbol{\mu}_c)^H \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{c} - \boldsymbol{\mu}_c) + \text{const}
\end{aligned} \tag{17}$$

where

$$\boldsymbol{\mu}_c \triangleq \langle \gamma \rangle \boldsymbol{\Sigma}_c \langle \boldsymbol{\Phi}_c \rangle^H \boldsymbol{y} \tag{18}$$

$$\boldsymbol{\Sigma}_c \triangleq \Big(\langle \gamma \rangle \langle \boldsymbol{\Phi}_c^H \boldsymbol{\Phi}_c \rangle + \langle \boldsymbol{A} \rangle \Big)^{-1} \tag{19}$$

$$\boldsymbol{A} \triangleq \operatorname{diag}(\boldsymbol{\alpha}) \tag{20}$$

Therefore, $q_c(\boldsymbol{c})$ follows a complex Gaussian distribution, i.e.

$$q_c(\boldsymbol{c}) = \mathcal{CN}(\boldsymbol{c}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{21}$$

Update of $q_d(\boldsymbol{d})$: Similarly, we express $y(t)$ as

$$\begin{aligned}
y(t) &= \tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}\boldsymbol{D}^T\tilde{\boldsymbol{f}}(t) + w(t) \\
&= \tilde{\boldsymbol{f}}^T(t)\boldsymbol{D}\boldsymbol{C}^T\tilde{\boldsymbol{z}}^*(t) + w(t) \\
&= \left((\tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}) \otimes \tilde{\boldsymbol{f}}^T(t)\right) \operatorname{vec}(\boldsymbol{D}) + w(t) \\
&= (\tilde{\boldsymbol{z}}^H(t) \otimes \tilde{\boldsymbol{f}}^T(t))(\boldsymbol{C} \otimes \boldsymbol{I}_{N_2})\boldsymbol{d} + w(t)
\end{aligned} \tag{22}$$

Therefore $\boldsymbol{y}$ can be written as

$$\begin{aligned}
\boldsymbol{y} &= \begin{bmatrix} \tilde{\boldsymbol{z}}^H(1) \otimes \tilde{\boldsymbol{f}}^T(1) \\ \vdots \\ \tilde{\boldsymbol{z}}^H(T) \otimes \tilde{\boldsymbol{f}}^T(T) \end{bmatrix} (\boldsymbol{C} \otimes \boldsymbol{I}_{N_2})\boldsymbol{d} + \boldsymbol{w} \\
&\triangleq \boldsymbol{\Phi}_{\boldsymbol{d}}\boldsymbol{d} + \boldsymbol{w}
\end{aligned} \tag{23}$$

The approximate posterior distribution of $q_d(\boldsymbol{d})$ can be calculated as

$$\begin{aligned}
\ln q_d(\boldsymbol{d}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z}) \rangle_{q_d(\boldsymbol{c})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\
&= \langle \ln p(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{d}, \gamma) \rangle + \langle \ln p(\boldsymbol{d}|\boldsymbol{\beta}) \rangle + \text{const} \\
&= -\langle \gamma \rangle \Big\langle (\boldsymbol{y} - \boldsymbol{\Phi}_{\boldsymbol{d}}\boldsymbol{d})^H (\boldsymbol{y} - \boldsymbol{\Phi}_{\boldsymbol{d}}\boldsymbol{d}) \Big\rangle - \boldsymbol{d}^H \langle \boldsymbol{B} \rangle \boldsymbol{d} + \text{const} \\
&= -(\boldsymbol{d} - \boldsymbol{\mu}_d)^H \boldsymbol{\Sigma}_d^{-1} (\boldsymbol{d} - \boldsymbol{\mu}_d) + \text{const}
\end{aligned} \tag{24}$$

where

$$\boldsymbol{\mu}_d \triangleq \langle \gamma \rangle \boldsymbol{\Sigma}_d \langle \boldsymbol{\Phi}_d \rangle^H \boldsymbol{y} \tag{25}$$

$$\boldsymbol{\Sigma}_d \triangleq \Big(\langle \gamma \rangle \langle \boldsymbol{\Phi}_d^H \boldsymbol{\Phi}_d \rangle + \langle \boldsymbol{B} \rangle \Big)^{-1} \tag{26}$$

$$\boldsymbol{B} \triangleq \operatorname{diag}(\boldsymbol{\beta}) \tag{27}$$

Therefore, $q_d(\boldsymbol{d})$ follows a complex Gaussian distribution, i.e.

$$q_d(\boldsymbol{d}) = \mathcal{CN}(\boldsymbol{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d) \tag{28}$$

Update of $q(\boldsymbol{\alpha})$: The variational optimization of $q_\alpha(\boldsymbol{\alpha})$ yields

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}) &= \langle \ln p(\boldsymbol{y}, z)\rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\beta(\boldsymbol{\beta})q_\gamma(\gamma)} + \text{const} \\ &= \langle \ln p(\boldsymbol{c}|\boldsymbol{\alpha})\rangle + \langle \ln p(\boldsymbol{\alpha})\rangle + \text{const} \\ &= \sum_{i=1}^{N_1 N} \ln \alpha_i - \sum_{i=1}^{N_1 N} \langle |c_i|^2\rangle \alpha_i \\ &\quad + (a-1)\sum_{i=1}^{N_1 N} \ln \alpha_i - b\sum_{i=1}^{N_1 N} \alpha_i + \text{const} \\ &\triangleq (a_1 - 1)\sum_{i=1}^{N_1 N} \ln \alpha_i - \sum_{i=1}^{N_1 N} b_{1i}\alpha_i + \text{const}\end{aligned} \tag{29}$$

where

$$a_1 \triangleq a + 1 \tag{30}$$

$$b_{1i} \triangleq b + \langle |c_i|^2\rangle \tag{31}$$

Therefore, $q_\alpha(\boldsymbol{\alpha})$ follows a Gamma distribution, i.e.

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{i=1}^{N_1 N} \text{Gamma}(\alpha_i; a_1, b_{1i}) \tag{32}$$

Update of $q(\boldsymbol{\beta})$: The variational optimization of $q_\beta(\boldsymbol{\beta})$ yields

$$\begin{aligned}\ln q_\beta(\boldsymbol{\beta}) &= \langle \ln p(\boldsymbol{y}, z)\rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)} + \text{const} \\ &= \langle \ln p(\boldsymbol{d}|\boldsymbol{\beta})\rangle + \langle \ln p(\boldsymbol{\beta})\rangle + \text{const} \\ &= \sum_{i=1}^{N_2 N} \ln \beta_i - \sum_{i=1}^{N_2 N} \langle |d_i|^2\rangle \beta_i \\ &\quad + (a-1)\sum_{i=1}^{N_2 N} \ln \beta_i - b\sum_{i=1}^{N_2 N} \beta_i + \text{const} \\ &\triangleq (a_2 - 1)\sum_{i=1}^{N_2 N} \ln \beta_i - \sum_{i=1}^{N_2 N} b_{2i}\beta_i + \text{const}\end{aligned} \tag{33}$$

where

$$a_2 \triangleq a + 1 \tag{34}$$

$$b_{2i} \triangleq b + \langle |d_i|^2\rangle \tag{35}$$

Therefore, $q_\beta(\boldsymbol{\beta})$ follows a Gamma distribution, i.e.

$$q_\beta(\boldsymbol{\beta}) = \prod_{i=1}^{N_2 N} \text{Gamma}(\beta_i; a_2, b_{2i}) \tag{36}$$

Update of $q(\gamma)$: The variational optimization of $q_\gamma(\gamma)$ yields

$$\begin{aligned}\ln q_\gamma(\gamma) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{z})\rangle_{q_c(\boldsymbol{c})q_d(\boldsymbol{d})q_\alpha(\boldsymbol{\alpha})q_\beta(\boldsymbol{\beta})} + \text{const} \\ &= \langle \ln p(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{d}, \gamma)\rangle + \langle \ln p(\gamma)\rangle + \text{const} \\ &= T\ln\gamma - \gamma\sum_{t=1}^{T}\left\langle |y(t) - \tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}\boldsymbol{D}^T\tilde{\boldsymbol{f}}(t)|^2\right\rangle \\ &\quad + (e-1)\ln\gamma - f\gamma + \text{const} \\ &\triangleq (e_1 - 1)\ln\gamma - f_1\gamma + \text{const}\end{aligned} \tag{37}$$

where

$$e_1 \triangleq e + T \tag{38}$$

$$f_1 \triangleq f + \sum_{t=1}^{T}\left\langle |y(t) - \tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}\boldsymbol{D}^T\tilde{\boldsymbol{f}}(t)|^2\right\rangle \tag{39}$$

Therefore, $q_\gamma(\gamma)$ follows a Gamma distribution, i.e.

$$q_\gamma(\gamma) = \text{Gamma}(\gamma; e_1, f_1) \tag{40}$$

Calculation of Some Expectations: Some of the expectations and moments used during the update are summarized as follows. Define

$$\boldsymbol{\Phi}_{FZ} \triangleq \begin{bmatrix} \tilde{\boldsymbol{f}}^T(1)\otimes\tilde{\boldsymbol{z}}^H(1) \\ \vdots \\ \tilde{\boldsymbol{f}}^T(T)\otimes\tilde{\boldsymbol{z}}^H(T)\end{bmatrix}, \boldsymbol{\Phi}_{ZF} \triangleq \begin{bmatrix} \tilde{\boldsymbol{z}}^H(1)\otimes\tilde{\boldsymbol{f}}^T(1) \\ \vdots \\ \tilde{\boldsymbol{z}}^H(T)\otimes\tilde{\boldsymbol{f}}^T(T)\end{bmatrix} \tag{41}$$

$$\boldsymbol{R}_{FZ} \triangleq \boldsymbol{\Phi}_{FZ}^H\boldsymbol{\Phi}_{FZ} \quad \boldsymbol{R}_{ZF} \triangleq \boldsymbol{\Phi}_{ZF}^H\boldsymbol{\Phi}_{ZF} \tag{42}$$

$$\begin{aligned}\boldsymbol{R}_{FZ}^{i,j} &\triangleq \boldsymbol{R}_{FZ}[(i-1)N_1 + 1 : iN_1, (j-1)N_1 + 1 : jN_1] \\ \boldsymbol{R}_{ZF}^{i,j} &\triangleq \boldsymbol{R}_{ZF}[(i-1)N_2 + 1 : iN_2, (j-1)N_2 + 1 : jN_2] \\ \boldsymbol{R}_c^{i,j} &\triangleq \langle \boldsymbol{C}[i,:]^H\rangle\langle \boldsymbol{C}[j,:]\rangle + \langle \boldsymbol{C}[i,:]^H\boldsymbol{C}[j,:]\rangle \\ \boldsymbol{R}_d^{i,j} &\triangleq \langle \boldsymbol{D}[i,:]^H\rangle\langle \boldsymbol{D}[j,:]\rangle + \langle \boldsymbol{D}[i,:]^H\boldsymbol{D}[j,:]\rangle\end{aligned} \tag{43}$$

The expectations $\langle \boldsymbol{\Phi}_c^H\boldsymbol{\Phi}_c\rangle$ and $\langle \boldsymbol{\Phi}_d^H\boldsymbol{\Phi}_d\rangle$ used in the update of $q_c(\boldsymbol{c})$ and $q_d(\boldsymbol{d})$ can be calculated as

$$\langle \boldsymbol{\Phi}_c^H\boldsymbol{\Phi}_c\rangle = \sum_{i=1}^{N_2}\sum_{j=1}^{N_2}(\boldsymbol{1}_{N\times N}\otimes\boldsymbol{R}_{FZ}^{i,j})\odot(\boldsymbol{R}_d^{i,j}\otimes\boldsymbol{1}_{N_1\times N_1}) \tag{44}$$

$$\langle \boldsymbol{\Phi}_d^H\boldsymbol{\Phi}_d\rangle = \sum_{i=1}^{N_1}\sum_{j=1}^{N_1}(\boldsymbol{1}_{N\times N}\otimes\boldsymbol{R}_{ZF}^{i,j})\odot(\boldsymbol{R}_c^{i,j}\otimes\boldsymbol{1}_{N_2\times N_2}) \tag{45}$$

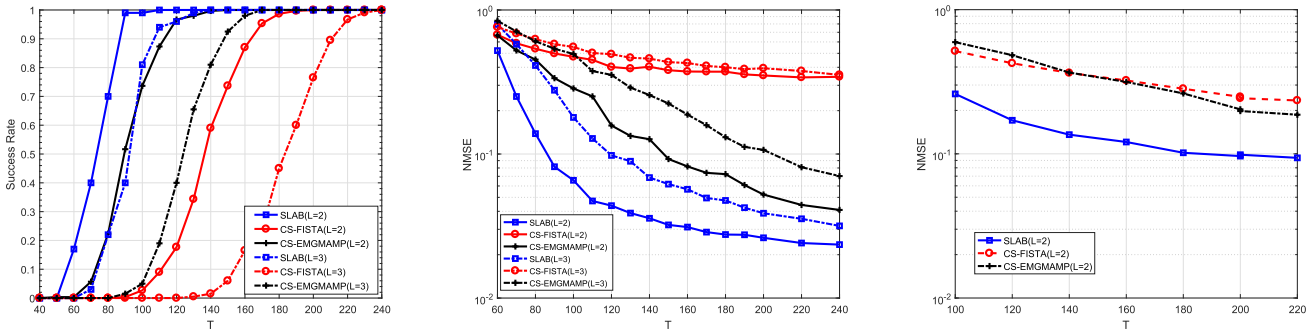
and the expectation in (39) can be calculated as

$$\sum_{t=1}^{T}\left\langle |y(t) - \tilde{\boldsymbol{z}}^H(t)\boldsymbol{C}\boldsymbol{D}^T\tilde{\boldsymbol{f}}(t)|^2\right\rangle \tag{46}$$

$$\begin{aligned}&= \sum_{t=1}^{T}|y(t)|^2 - 2\text{Re}\{y^*(t)\tilde{\boldsymbol{z}}^H(t)\langle \boldsymbol{C}\rangle\langle \boldsymbol{D}^T\rangle\tilde{\boldsymbol{f}}(t)\} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\boldsymbol{z}}^H(t)\langle \boldsymbol{C}(:,j)\boldsymbol{C}(:,i)^H\rangle\tilde{\boldsymbol{z}}(t)\tilde{\boldsymbol{f}}^H(t)\langle \boldsymbol{D}^*(:,i)\boldsymbol{D}(:,j)^T\rangle\tilde{\boldsymbol{f}}(t)\end{aligned} \tag{47}$$

For clarity, the proposed variational Bayesian learning approach for mmWave channel estimation is summarized in Algorithm 1. Note that the algorithm is considered to be convergent when the normalized difference between two consecutive estimates of $\boldsymbol{H}_v$ is less than 10^{-2}.

We analyze the computational complexity of our proposed method. Note that the major computational task of our proposed method involves calculating $\boldsymbol{\Sigma}_c \in \mathbb{C}^{N_1N_2\times N_1N_2}$ and $\boldsymbol{\Sigma}_d \in \mathbb{C}^{N_1N_2\times N_1N_2}$, both of which require to perform inverse of an $N_1N_2\times N_1N_2$ matrix. Therefore our proposed algorithm has a computational complexity of $\mathcal{O}(N_1^3N_2^3)$. To reduce the



(a) (b) (c)

FIGURE 2. Geometric channel model: Success rates and NMSEs of respective algorithms vs. T (a) Success rates vs. T (b) NMSEs vs. T when SNR = 10dB. (c) NMSEs vs. T when SNR = 5dB.

Algorithm 1 Proposed Variational Bayesian Algorithm

Input: $\boldsymbol{y}$, $\{\tilde{\boldsymbol{z}}(t)\}_{t=1}^{T}$ and $\{\tilde{\boldsymbol{f}}(t)\}_{t=1}^{T}$.
Output: $q_c(\boldsymbol{c})$, $q_d(\boldsymbol{d})$, $q_\alpha(\boldsymbol{\alpha})$, $q_\beta(\boldsymbol{\beta})$, and $q_\gamma(\gamma)$.
Initialize $\langle \boldsymbol{c} \rangle$, $\langle \boldsymbol{d} \rangle$, $\langle \boldsymbol{\alpha} \rangle$, $\langle \boldsymbol{\beta} \rangle$, $\langle \gamma \rangle$, $\langle \boldsymbol{c}^H \boldsymbol{c} \rangle$, and $\langle \boldsymbol{d}^H \boldsymbol{d} \rangle$;
while not converge **do**
Update $q_c(\boldsymbol{c})$ via (21), with $q_d(\boldsymbol{d})$, $q_\alpha(\boldsymbol{\alpha})$, $q_\beta(\boldsymbol{\beta})$, and $q_\gamma(\gamma)$ fixed;
Update $q_d(\boldsymbol{d})$ via (28), with $q_c(\boldsymbol{c})$, $q_\alpha(\boldsymbol{\alpha})$, $q_\beta(\boldsymbol{\beta})$, and $q_\gamma(\gamma)$ fixed;
Update $q_\alpha(\boldsymbol{\alpha})$ via (32), with $q_c(\boldsymbol{c})$, $q_d(\boldsymbol{d})$, $q_\beta(\boldsymbol{\beta})$, and $q_\gamma(\gamma)$ fixed;
Update $q_\beta(\boldsymbol{\beta})$ via (36), with $q_c(\boldsymbol{c})$, $q_d(\boldsymbol{d})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\gamma(\gamma)$ fixed;
Update $q_\gamma(\gamma)$ via (40), with $q_c(\boldsymbol{c})$, $q_d(\boldsymbol{d})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\beta(\boldsymbol{\beta})$ fixed;
end while

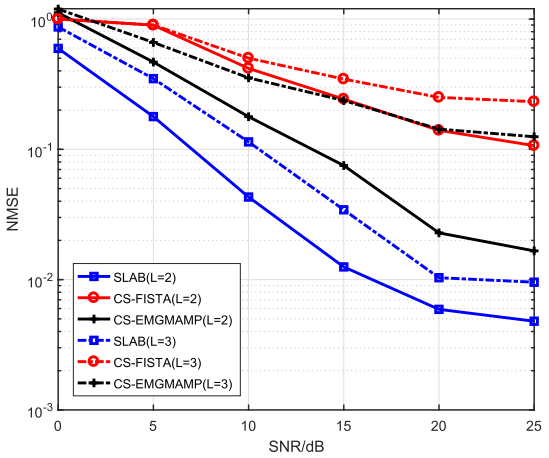


FIGURE 3. Geometric channel model: NMSEs of respective algorithms vs. SNR.

computational complexity, similar to [32], we can resort to the generalized approximate massage passing (GAMP) technique to obtain approximate posterior distributions of $q_c(\boldsymbol{c})$ and $q_d(\boldsymbol{d})$, thus circumventing cumbersome matrix inverse operations. This will be a topic for our future work. As a comparison, we consider the computational complexity of some conventional compressed sensing-based channel estimation methods. For the fast iterative shrinkage-thresholding algorithm [33], the main computational task is dominated by evaluating the proximal operator per iteration, which has a complexity of $\mathcal{O}(N_1^2 N_2^2)$. The approximate message passing (AMP)-based Bayesian method [10], [34] involves some simple matrix-vector multiplications, which has a computational complexity of $\mathcal{O}(TN_1N_2)$.

V. SIMULATION RESULTS

In this section, we carry out experiments to illustrate the performance of our proposed Bayesian method. The proposed method is referred to as the joint Sparse and Low-rAnk Bayesian learning (SLAB) algorithm. We compare our proposed method with the approximate message passing (AMP)-based Bayesian method developed in [34], [10], and the fast iterative shrinkage-thresholding algorithm (FISTA) proposed in [33]. These two conventional compressed sensing methods are respectively referred to as CS-EMGMAMP and CS-FISTA. Note that it was shown in [13] that the CS-EMGMAMP method empirically outperforms the two-stage compressed sensing method [13] which was devised to exploit the joint sparse and low-rank structure of mmWave channels. Therefore the two-stage compressed sensing method was not included for comparison. In our experiments, the beamforming vectors $\{\boldsymbol{f}(t)\}_{t=1}^{T}$ and the combining vectors $\{\boldsymbol{z}(t)\}_{t=1}^{T}$ are randomly generated with its entries independently drawn from the unit circle.

We consider the scenario where both the transmitter and the receiver employ a uniform linear array (ULA) with $N_t = N_r = 32$ antennas. The distance between neighboring antenna elements is assumed to be half the wavelength of the signal. The mmWave channel is generated according to the geometric channel model (4), where the number of clusters L is set to $L = 2$ and $L = 3$. When $L = 2$, the mean AoAs/AoDs for these two clusters are set to $\theta_1 = \phi_1 = \pi/6, \theta_2 = \phi_2 = -\pi/6$, respectively. For the case of $L = 3$,

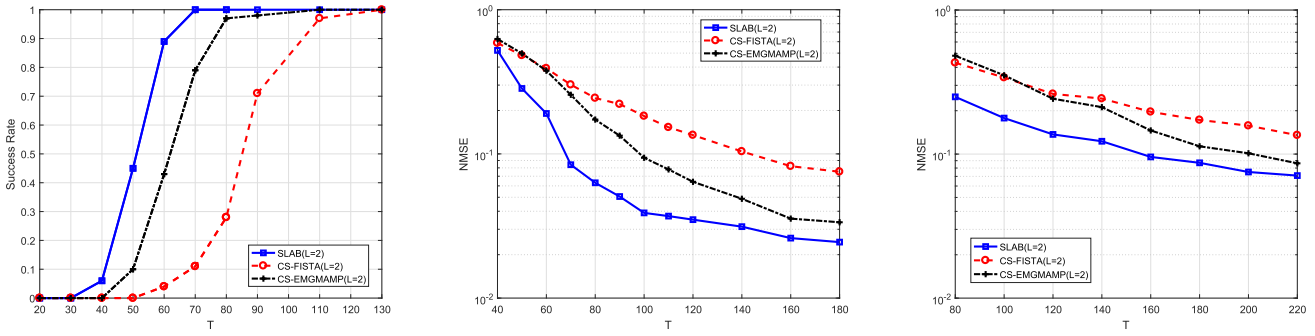


FIGURE 4. Geometric channel: Success rates and NMSEs of respective algorithms vs. T for the non-symmetrical case: $N_r = 16$ and $N_t = 32$. (a) Success rates vs. T (b) NMSEs vs. T when SNR = 10dB. (c) NMSEs vs. T when SNR = 5dB.

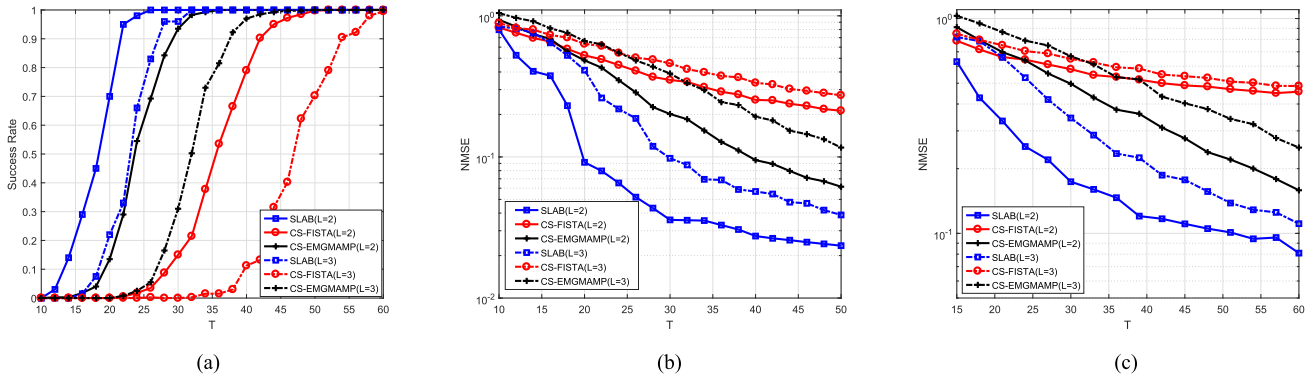


FIGURE 5. Geometric channel model: Success rates and NMSEs of respective algorithms vs. T, where the number of RF chains at the receiver is set to 4. (a) Success rates vs. T (b) NMSEs vs. T when SNR = 10dB. (c) NMSEs vs. T when SNR = 5dB.

the mean AoAs/AoDs for these three clusters are set to $\theta_1 = \phi_1 = \pi/6, \theta_2 = \phi_2 = -\pi/6, \theta_3 = \phi_3 = -10.8\pi/180$, respectively. The angular spreads (over the AoA and the AoD) of each cluster are set to $\delta_\theta = \delta_\phi = 16\pi/180$, and the relative AoA/AoD shifts are uniformly generated within the angular spreads, i.e. $\vartheta_{l,i} \in [\theta_l - \delta_\theta/2, \theta_l + \delta_\theta/2]$, $\varphi_{l,i} \in [\phi_l - \delta_\phi/2, \phi_l + \delta_\phi/2]$. The number of paths per cluster is set to 100, i.e. I and J in (4) are set to 10. In our simulations, we discretize the AoA/AoD domains into 32×32 grid points, i.e. $N_1 = N_2 = 32$. In this case, δ_θ and δ_ϕ span across about 5 grid points on the AoA and the AoD domain, respectively.

To evaluate the recovery performance of respective algorithms, two metrics are considered, namely, the normalized mean squared error (NMSE) and the success rate. The NMSE is calculated as

$$\text{NMSE} = E\left[\frac{\|\hat{\boldsymbol{H}} - \boldsymbol{H}\|_F^2}{\|\boldsymbol{H}\|_F^2}\right] \tag{48}$$

where $\hat{\boldsymbol{H}}$ denotes the estimate of the true channel $\boldsymbol{H}$. The success rate is computed as the ratio of the number of successful trials to the total number of independent runs. A trial is considered successful if the normalized reconstruction error $\|\hat{\boldsymbol{H}} - \boldsymbol{H}\|_F^2/\|\boldsymbol{H}\|_F^2$ is no greater than 10^{-2}.

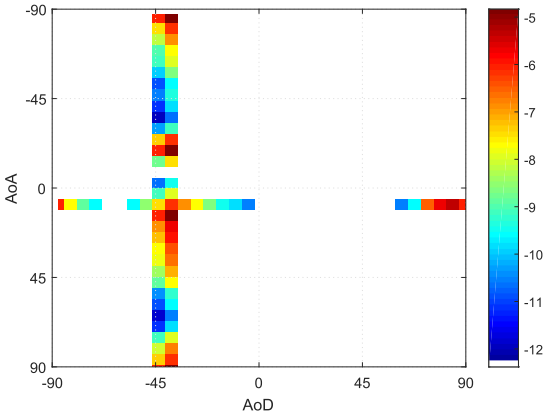
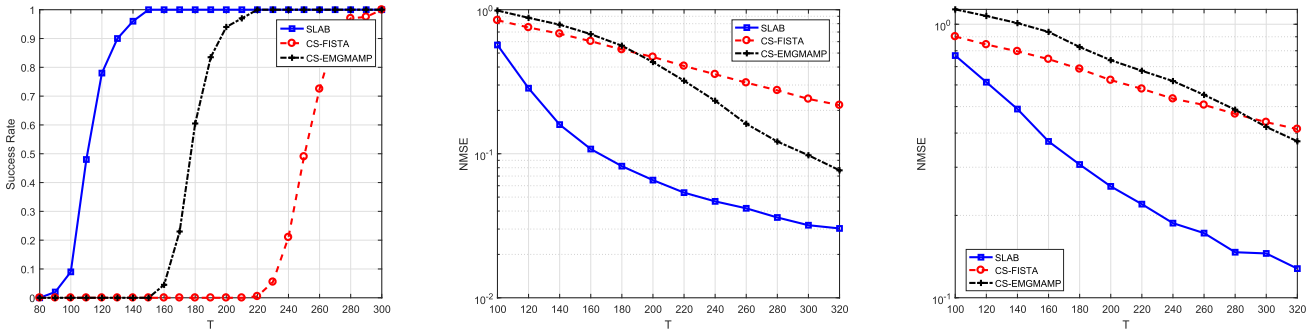


FIGURE 6. The RX power angular profile of the simulated channel.

We first consider the noiseless case. Fig. 2(a) depicts the success rates of respective algorithms as a function of the number of measurements T, where both $L = 2$ and $L = 3$ are considered. We see that our proposed method SLAB presents a significant performance improvement over the CS-FISTA and a clear performance advantage over the CS-EMGMAMP. This performance improvement is primarily due to the fact that our proposed method exploits the joint sparse and low-rank structure of mmWave channels, whereas the other two



(a) (b) (c)

FIGURE 7. Simulated channel: Success rates and NMSEs of respective algorithms vs. T (a) Success rates vs. T (b) NMSEs vs. T when SNR = 10dB. (c) NMSEs vs. T when SNR = 5dB.

methods only utilize the sparsity pattern of mmWave channels. As a result, our proposed method has the potential to achieve a lower sample complexity than the other two methods. Fig. 2(b) and Fig. 2(c) plot the NMSEs of respective algorithms vs. T in the noisy case, where the signal-to-noise ratio (SNR), defined as $10\log(\|\boldsymbol{H}\|_F^2/(N_r N_t \sigma^2))$, is set to 10dB and 5dB, respectively. It can be seen that our proposed method achieves a higher estimation accuracy than the other two methods. Fig. 3 plots the NMSEs of different methods as the SNR varies, where we set $T = 120$. We see that our proposed algorithm yields decent performance even in a low SNR regime, and outperforms the other two methods across the whole SNR region.

We also consider a non-symmetrical scenario where $N_r = 16$ and $N_t = 32$. Fig. 4 depicts the success rates and NMSEs of respective algorithms as a function of the number of measurements T. The results in Fig. 4, again, demonstrate the superiority of the SLAB method. Our proposed method can be easily extended to the scenario where multiple RF chains are employed at the receiver. Suppose there are R RF chains at the receiver. In this case, the receiver can collect R measurements at each time instant. Fig. 5 depicts the success rates and the NMSEs of respective algorithms vs. T, where we set $R = 4$. Our results show that the time complexity can be significantly reduced at the expenses of increasing the number of RF chains.

Next, we examine the performance of our proposed method with a more practically relevant channel. We simulate a channel according to recent real-world mmWave channel measurements. Specifically, the RX power angular profile of the simulated channel is depicted in Fig. 6, which is similar to the RX power angular profile measured at a typical TX-RX location pair at 28GHz [21]. From Fig. 6, we see that, due to the angular spreads over the AoA and AoD domains, the mmWave channel exhibits a joint sparse and low-rank structure in which the rank of the channel is far smaller than the number of dominant entries of the beam space channel. Fig. 7 plots the success rates for the noiseless case and the NMSEs for the noisy case as a function of the number of measurements T, where, in the noisy case, the SNR is set to 5dB and 10dB, respectively. From Fig. 7, we see that our proposed method achieves significant performance improvements over the competing algorithms. The performance advantage is more pronounced compared to the results reported in previous figures. This is probably because the simulated channel has a more prominent low-rank structure as compared with its number of nonzero entries, which is a case more favorable for our proposed method. Fig. 8 shows the NMSEs of respective methods under different SNRs, where the number of measurements is set to $T = 200$. We observe that our proposed method yields a lower estimation error in the moderate and high SNR regime.

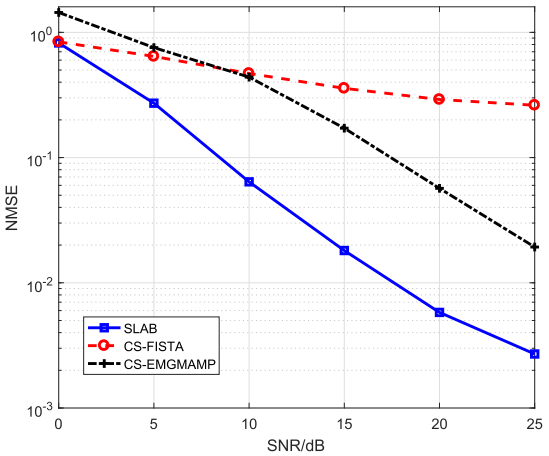


FIGURE 8. Simulated channel: NMSEs of respective algorithms vs. SNR.

VI. CONCLUSION

We studied the problem of channel estimation for mmWave systems, where both the base station and the mobile station employ a single RF chain for analog beamforming/combining. To exploit the simultaneously sparse and low-rank structure arising from angular spreads, we addressed the channel estimation problem within a Bayesian framework, where a matrix factorization formulation is adopted to translate the channel estimation problem into one of searching for two factor matrices, and independent sparsity-promoting

priors are placed on entries of the two factor matrices to encourage a joint sparse and low-rank solution. A variational Bayesian inference method was then developed based on the proposed prior model. Simulation results were provided to demonstrate the superiority of our proposed method over state-of-the-art compressed sensing-based mmWave channel estimation methods.

REFERENCES

[1] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[4] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1425–1431, Jul. 2017.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[6] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[8] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 2909–2913.

[9] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 273–277.

[10] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun.*, Stockholm, Sweden, Jun./Jul. 2015, pp. 146–150.

[11] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[12] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[13] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[14] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[15] C.-R. Tsai, Y.-H. Liu, and A.-Y. Wu, "Efficient compressive channel estimation for millimeter-wave large-scale antenna systems," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2414–2428, May 2018.

[16] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.

[17] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.

[18] M. Samimi *et al.*, "28 GHz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in New York city," in *Proc. IEEE 77th Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013, pp. 1–6.

[19] H. Zhao *et al.*, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *Proc. IEEE Int. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 5163–5167.

[20] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[21] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[22] P. Wang, M. Pajovic, P. V. Orlik, T. Koike-Akino, K. J. Kim, and J. Fang, "Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.

[23] H. Zhang, Z. Lin, and C. Zhang, "Completing low-rank matrices with corrupted samples from few coefficients in general basis," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4748–4768, Aug. 2016.

[24] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[25] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[26] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.

[27] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[28] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, May 2015.

[29] S. Bahmani and J. Romberg, "Near-optimal estimation of simultaneously sparse and low-rank matrices from nested linear measurements," *Inf. Inference, A J. IMA*, vol. 5, no. 3, pp. 331–351, Sep. 2016.

[30] K. Lee, Y. Wu, and Y. Bresler, "Near-optimal compressed sensing of a class of sparse low-rank matrices via sparse power factorization," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1666–1698, Mar. 2018.

[31] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Jan. 2008.

[32] L. Yang, J. Fang, H. Duan, H. Li, and B. Zeng, "Fast low-rank Bayesian matrix completion with hierarchical Gaussian prior models," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2804–2817, Jun. 2018.

[33] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[34] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

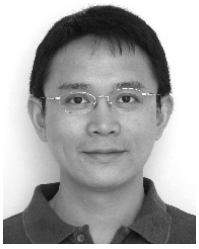
KAIHUI LIU is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interests include the mathematics of data science, particularly the interplay of convex and non-convex optimization, statistical learning and estimation, signal processing, computational imaging, and wireless communication.

XINGJIAN LI received the B.Sc. degree from the University of Electronic Science and Technology of China, in 2015, where he is currently pursuing the Ph.D. degree. His current research interests include compressed sensing, millimeter wave, and massive MIMO communications.

JUN FANG (M'08) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

In 2006, he was a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic of Science and Technology of China. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference.

Dr. Fang received the IEEE Jack Neubauer Memorial Award, in 2013 for the Best Systems Paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He serves as a Senior Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, and an Associate Technical Editor for the *IEEE Communications Magazine*.



HONGBIN LI (M'99–SM'08–F'18) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor. He was a Summer Visiting Faculty Member of the Air Force Research Laboratory, in the summers of 2003, 2004, and 2009, respectively. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the IEEE Jack Neubauer Memorial Award, in 2013 for the Best Systems Paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the Outstanding Paper Award from the IEEE AFICON Conference, in 2011, the Harvey N. Davis Teaching Award, in 2003, the Jess H. Davis Memorial Award for Excellence in Research from the Stevens Institute of Technology, in 2001, and the Sigma Xi Graduate Research Award from the University of Florida, in 1999. He has been a member of the IEEE SPS Signal Processing Theory and Methods, since 2011, the Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC (2006–2012). He has been involved in various conference organization activities, including serving as the General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, NJ, USA, in 2012. He has been an Associate Editor of the *Signal Processing* (Elsevier), since 2013, the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2009 and since 2014), the IEEE SIGNAL PROCESSING LETTERS (2005–2006), and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2003–2006), as well as, a Guest Editor for the IEEE *Journal of Selected Topics in Signal Processing* and the *EURASIP Journal on Applied Signal Processing*.