Sparse Channel Estimation in Millimeter Wave Communications: Exploiting Joint AoD-AoA Angular Spread

Pu Wang, Milutin Pajovic, Philip V. Orlik,
Toshiaki Koike-Akino, and Kyeong Jin Kim
Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge, MA 02139, USA
Emails: {pwang, pajovic, porlik, koike, kkim}@merl.com

Jun Fang
National Key Laboratory on Communications
University of Electronic Science and Technology of China
Chengdu 611731, China
Email: junfang@uestc.edu.cn

***Abstract*—In this paper, channel estimation in millimeter wave (mmWave) communication systems is considered. In contrast to prevailing mmWave channel estimation methods exploiting the sparsity nature of the channel, we move one step further by exploiting the joint AoD-AoA angular spread. By formulating the channel estimation as a block-sparse signal recovery with an underlying two-dimensional cluster feature, we propose a two-dimensional sparse Bayesian learning method without *a priori* knowledge of two-dimensional angular spread patterns. It essentially couples the channel path power at one angular direction with its two-dimensional AoD-AoA neighboring directions. Compared with existing sparse mmWave channel estimation methods, the proposed method is numerically verified to reduce the training overhead and channel estimation error.**

I. Introduction

Millimeter wave (mmWave) communication is a promising technology for future fifth generation (5G) cellular networks. It has the potential to offer gigabit-per-second data rates by exploiting the large bandwidth available at mmWave frequencies [1], [2]. However, communication at such high frequencies suffers from high attenuation and signal absorption. To compensate for the significant path loss, large antenna arrays can be used at the base station (BS) and mobile station (MS) to exploit beam steering to increase the link gain. On one hand, directional precoding and beamforming provide sufficient beamforming gain for mmWave communications. On the other hand, the precoding design requires reliable *channel state information* (CSI) which is challenging to obtain due to the large number of antennas and rapidly varying channel statistics.

The *sparse scattering* nature of the mmWave channel has been utilized in [3] and [4] to reduce the training overhead for mmWave channel estimation. [3] presented a new multi-resolution beamforming codebook and an adaptive compressed channel sensing method for the mmWave channel estimation. The compressed channel sensing scheme divides the training process into several phases, in which the precoding design uses the information from previous phases. However, the requirement of a feedback channel may not be favorable in certain scenarios. On the other hand, [4] focused on the compressed channel sensing scheme with significantly less training signals by taking into account the sparse nature of the mmWave channel.

In this paper, we are still interested in the mmWave channel estimation and move one step further to exploit, in addition to the sparse channel scattering, the joint *angular spread* of path clusters in the angle-of-departure (AoD) and angle-of-arrival (AoA) domain. The joint angular spread induces a two-dimensional block sparse pattern in the resulting complex channel gain matrix, which has been shown in real-world measurements in urban environments [5]–[7]. Specifically, we propose a two-dimensional coupled sparse Bayesian learning (SBL) algorithm to exploit the joint AoD-AoA angular spread. The coupled SBL algorithm treats the channel gain of each path as a random variable and imposes a two-dimensional statistical dependence across the channel path power to favor block-sparse solutions without knowing the block pattern *a priori*. The proposed algorithm encompasses two steps and iterates between them: the Bayesian estimation of the channel gain matrix followed by iteratively updating the prior variance (or, equivalently, the channel path power) by using the expectation-maximization (EM) algorithm. Compared with several existing sparse channel estimation methods, the proposed algorithm shows numerical advantages such as reduced training overhead and lower estimation errors for the mmWave channel estimation.

The rest of the paper is organized as follows. Section II introduces the system model and a sparse representation of the mmWave channel. Section III provides motivations to exploit the joint AoD-AoD spread and formulates the problem in a block-sparse signal recovery framework. In Section IV, we provide the details on deriving the proposed algorithm. Numerical results are provided in Section V, followed by concluding remarks in Section VI.

II. Sparse mmWave Channel Model

Consider a mmWave communication system with N_T transmitters at the BS and N_R receivers at MS. At time instant k, the BS applies a precoder/beamformer $\mathbf{p}_k$ to transmit a symbol s. Without loss of generality, $s = 1$. Correspondingly, the MS applies a combiner $\mathbf{q}_k$ to generate the received signal y_k:

$$y_k = \sqrt{\rho}\mathbf{q}_k^H \mathbf{H}\mathbf{p}_k s + \mathbf{q}_k^H \mathbf{v}, \quad k = 1, 2, \cdots, K, \qquad (1)$$

where ρ is the average transmitted power, $\mathbf{H} \in \mathcal{C}^{N_T \times N_R}$ is the channel matrix, and $\mathbf{v}$ is the white Gaussian noise with an

978-1-4673-8999-0/17/$31.00 ©2017 IEEE

unknown variance σ^2, and K is the number of training signals. Here, we consider to use either phase shifters or switches. For the phase shifter, $\mathbf{p}_k$ and $\mathbf{q}_k$ are selected from the elements $\{\pm 1, \pm j\}$, while $\mathbf{p}_k$ and $\mathbf{q}_k$ are binary selection vectors with a few ones (e.g., a single one) and zeros elsewhere for the switch.

Assuming a geometric channel model with N_s scatterers between the BS and the MS, the channel matrix $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \sum_{i=1}^{N_s} \alpha_i \boldsymbol{\alpha}_{\text{BS}}(\theta_i)\boldsymbol{\alpha}_{\text{MS}}^H(\phi_i) \tag{2}$$

where α_i is the complex path gain associated with the i-th path, θ_i and ϕ_i are the associated angular AoD and AoA, respectively, $\boldsymbol{\alpha}_{\text{BS}}(\theta_i)$ and $\boldsymbol{\alpha}_{\text{MS}}(\phi_i)$ denote the array response vectors associated with the BS and the MS, respectively. In the case of a uniform linear array (ULA), the steering vectors can be written in terms of the spatial frequency ω and ψ

$$\begin{aligned}\boldsymbol{\alpha}_{\text{BS}}(\theta) = \boldsymbol{\alpha}_{\text{BS}}(\omega) &= \frac{1}{\sqrt{N_T}}\left[1, e^{j\omega}, \cdots, e^{j(N_T-1)\omega}\right]^T, \\ \boldsymbol{\alpha}_{\text{MS}}(\phi) = \boldsymbol{\alpha}_{\text{MS}}(\psi) &= \frac{1}{\sqrt{N_R}}\left[1, e^{j\psi}, \cdots, e^{j(N_R-1)\psi}\right]^T, \end{aligned} \tag{3}$$

where $\omega = 2\pi d \sin(\theta)/\lambda$ and $\psi = 2\pi d \sin(\phi)/\lambda$ with d denoting the inter-element spacing of the array and λ the wavelength. In this paper, we consider the problem of estimating the channel path gain α_i and its associated spatial frequencies, i.e., ω and ψ from the K measurements $y_k, k = 1, 2, \cdots, K$.

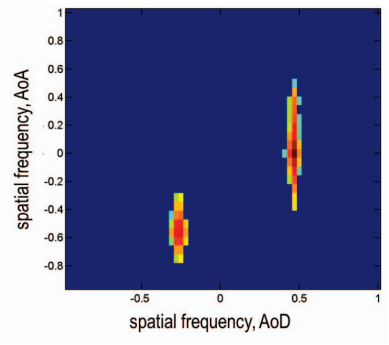
This sparse channel feature can be utilized to formulate the channel estimation problem as a sparse signal recovery along with the channel sensing scheme in (1). To see this, we decompose the channel matrix into an overcomplete virtual channel model

$$\mathbf{H} = \mathbf{A}_{\text{MS}}\mathbf{H}_v\mathbf{A}_{\text{BS}}^H \tag{4}$$

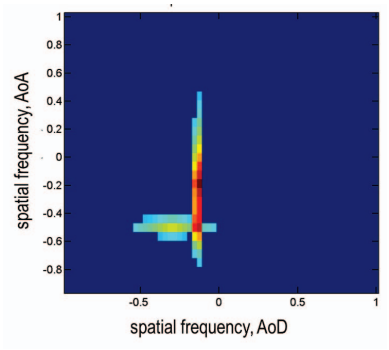
where $\mathbf{A}_{\text{MS}} = [\boldsymbol{\alpha}_{\text{MS}}(\psi_1), \cdots, \boldsymbol{\alpha}_{\text{MS}}(\psi_{N_{\text{A}}})]$ is an overcomplete matrix (i.e, $N_{\text{A}} \gg N_s$) with each column representing the steering vector at a pre-discretized AoA spatial frequency, $\mathbf{A}_{\text{BS}} = [\boldsymbol{\alpha}_{\text{BS}}(\omega_1), \cdots, \boldsymbol{\alpha}_{\text{BS}}(\omega_{N_{\text{D}}})]$ represents the steering matrix at pre-discretized AoD spatial frequencies with $N_{\text{D}} \gg N_s$, and $\mathbf{H}_v$ is a sparse matrix with N_s non-zero entries corresponding to the channel path gain in (1). The overcomplete virtual channel representation of (4) further assumes that the spatial frequencies of the N_s scatterers fall into the pre-discretized grids of the AoD ($[\omega_1, \cdots, \omega_{N_{\text{D}}}]$) and AoA ($[\psi_1, \cdots, \psi_{N_{\text{A}}}]$).

Note that $\text{vec}(\mathbf{H}) = \left(\mathbf{A}_{\text{BS}}^* \otimes \mathbf{A}_{\text{MS}}^H\right)\mathbf{h}$ with $\mathbf{h} = \text{vec}(\mathbf{H}_v) \in \mathcal{C}^{N_{\text{D}}N_{\text{A}}\times 1}$ denoting the sparse vector with N_s non-zero entries. Plugging (4) back to (1) yields

$$\begin{aligned} y_k &= \sqrt{\rho}\left(\mathbf{p}_k^T \otimes \mathbf{q}_k^H\right)\text{vec}(\mathbf{H}) + \mathbf{q}_k^H\mathbf{v}, \quad k = 1, 2, \cdots, K, \\ &= \sqrt{\rho}\left(\mathbf{p}_k^T \otimes \mathbf{q}_k^H\right)\left(\mathbf{A}_{\text{BS}}^* \otimes \mathbf{A}_{\text{MS}}^H\right)\mathbf{h} + \mathbf{q}_k^H\mathbf{v} \\ &= \sqrt{\rho}\left(\mathbf{p}_k^T \otimes \mathbf{q}_k^H\right)\boldsymbol{\Psi}\mathbf{h} + \mathbf{q}_k^H\mathbf{v}, \end{aligned} \tag{5}$$



(a) Scenario 1



(b) Scenario 2

Fig. 1. Channel path power profiles in two scenarios of joint AoD-AoA angular spread: (a) **Scenario 1**: two separated path clusters with AoA spreads larger than AoD spreads; (b) **Scenario 2**: two intersecting path clusters with one spreading over the AoD and the other spreading along the AoA. Colors represent the average power.

where $\boldsymbol{\Psi} = \mathbf{A}_{\text{BS}}^* \otimes \mathbf{A}_{\text{MS}}^H$. Grouping all K measurements, we have

$$\begin{aligned}\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} &= \sqrt{\rho}\begin{bmatrix} \mathbf{p}_1^T \otimes \mathbf{q}_1^H \\ \mathbf{p}_2^T \otimes \mathbf{q}_2^H \\ \vdots \\ \mathbf{p}_K^T \otimes \mathbf{q}_K^H \end{bmatrix}\boldsymbol{\Psi}\mathbf{h} + \begin{bmatrix} \mathbf{q}_1^H\mathbf{v} \\ \mathbf{q}_2^H\mathbf{v} \\ \vdots \\ \mathbf{q}_K^H\mathbf{v} \end{bmatrix} \\ &= \mathbf{A}\mathbf{h} + \mathbf{e}. \end{aligned} \tag{6}$$

To recover the sparse vector $\mathbf{h}$, existing solutions such as the greedy methods (e.g., orthogonal matching pursuit (OMP)) and the ℓ_1-norm regularized least square solutions (e.g., the LASSO and BPDN) [8] can be utilized to exploit the sparsity in $\mathbf{h}$ while reducing the number of training signals K.

III. Exploiting Joint AoD-AoA Spread as Two-Dimensional Block-Sparse Recovery

In this paper, in addition to the sparse channel scattering, we further exploit the *joint AoD-AoA spread* of the mmWave channel. Several real-world measurements in dense-urban propagation environments reveal that mmWave channel may spread in the form of cluster of paths over the angular domains including the AoD, AoA and elevation [5]–[7]. The angular spread may exacerbate as the spatial resolution becomes finer when the number of antennas at the BS and MS

increases, which is highly likely when the massive MIMO is fused with the mmWave transmission. In [6] and [7], spatial channel models have been statistically proposed from real-world measurements at 28 and 73 GHz in New York city, and they provide a realistic assessment of mmWave micro- and pico-cellular networks in a dense urban deployment. Specifically, the angular spread (or angular dispersion) of path clusters has been explicitly studied in terms of the root mean-squared (rms) beamspread in the different angular (AoA, AoD, and elevation) dimensions. The measured (rms) angular spreads at 28 GHz and 73 GHz are listed in Table 1 of [7]. The AoA spreads are 15.5° and 15.4°, respectively, for the two carrier frequencies, while corresponding AoD spreads are 10.2° and 10.5°, respectively. Fig. 1 shows the path cluster power profiles of two scenarios generated using the proposed statistical channel model with fitted large-scale parameters in [7]. Fig. 1 (a) shows a case of two separated path clusters with the AoA angular spread relatively larger than the AoD spread, while Fig. 1 (b) shows a scenario where two path clusters are intersected with one cluster spreading over the AoD and the other spreading over the AoA.

To explore the angular spread in $\mathbf{H}_v$ (and, subsequently, $\mathbf{h}$), [9] first looked at the one-dimensional AoA spread in $\mathbf{H}_v$ for the mmWave channel estimation. To further exploit the two-dimensional joint AoD-AoA spread, the channel estimation can be straightforwardly formulated as a two-dimensional block-sparse signal recovery problem, where the block-sparsity is now in the joint AoD-AoA domain. Block-sparse signal recovery has been studied recently in the literature [10]–[14]; for instance, the group LASSO [10], the mixed ℓ_2/ℓ_1 program [11], and the block OMP [12]. However, these methods require *a priori* knowledge of *block patterns* including the block size and its location. For the problem concerned here, such information is usually not available in advance.

Adaptive block-sparse signal recovery algorithms [13], [14] with capability of learning the block pattern are more suitable for the problem here. These algorithms introduce block/graphical priors to model the statistical dependencies between atoms. However, these priors usually lead to intractable posterior distribution and computationally intensive sampling methods have to be used. In [15], we proposed a simple yet flexible coupled sparse Bayesian framework for the one-dimensional block-sparse signal recovery. It introduces statistical dependencies among bases (i.e., atoms) via a hierarchical Gaussian prior model and allows deriving a closed-form Bayesian MAP/MMSE estimator for the signal parameter estimation and an approximated closed-form EM algorithm for automatically updating the hyperparameters. This has been recently extended to the two-dimensional block-sparse signal recovery for the inverse synthetic aperture imaging [16].

IV. Proposed Channel Estimation Method

In the section, we first formulate the joint AoD-AoA channel estimation problem with a hierarchical Gaussian signal model with coupled prior variances, and then derive Bayesian inference algorithms.

A. Two-Dimensional Coupled Hierarchical Signal Model

1) Hierarchical Gaussian Model: Recall the signal model in (6) where $\mathbf{A} \in \mathcal{C}^{K\times N_D N_A}$, $\mathbf{h}$ is the vectorized complex channel gain matrix

$$\mathbf{h} = [h_1, \cdots, h_N]^T = \text{vec}\{\mathbf{H}_v\}, \quad \mathbf{H}_v \in \mathcal{C}^{N_A \times N_D}, \tag{7}$$

with $N = N_D N_A$, and $\mathbf{e} \in \mathcal{C}^{K\times 1}$ is the noise vector. Then a standard hierarchical Gaussian prior model is assumed for the channel gain vector $\mathbf{h}$ [17]. First, the entries of $\mathbf{h}$, i.e., the complex channel gain h_n, are assumed to be, conditioned on its variance, independently Gaussian distributed with zero mean and variance λ_n:

$$h_n|\alpha_n \sim \mathcal{CN}(0, \alpha_n^{-1}). \tag{8}$$

where $\alpha_n = 1/\lambda_n$ denotes the precision (i.e., the reciprocal of the variance) of h_n. It is worthy noting that the prior variance λ_n represents the channel gain power associated with the n-th path. It is also seen from (8) that the channel path gain h_n becomes zero when its corresponding α_n goes to infinity (or the corresponding variance λ_n goes to zero). Invoking the independence across the entries, we have

$$p(\mathbf{h}|\boldsymbol{\alpha}) = \prod_{n=1}^{N} p(h_n|\alpha_n). \tag{9}$$

To allow statistical learning on α_n, the hierarchical Gaussian prior model further assumes that the prior precisions are also random variables that follow the i.i.d. Gamma distribution

$$p(\alpha_n|\mu,\nu) = \frac{1}{\Gamma(\mu)}\nu^\mu \alpha_n^{\mu-1} e^{-\nu\alpha_n}, \text{with } \alpha_n \geq 0. \tag{10}$$

where $\Gamma(\mu) = \int_0^\infty t^{\mu-1}e^{-t}dt$ is the Gamma function, μ is the shape parameter, and ν is the inverse shape parameter. We might fix these small values to the hyper-hyperparameters: e.g., $\mu = \nu = 10^{-4}$ to make these priors non-informative. It is easy to show that, with the Gamma distribution on α_n of (10) and the conditional Gaussian distribution of (8), the prior distribution of h_n reduces to the Student-t distribution [17]

$$p(h_n) = \int p(h_n|\alpha_n)p(\alpha_n)d\alpha_n \sim e^{\left(\mu + h_n^2/2\right)^{(-\nu+0.5)}}, \tag{11}$$

which inherently promotes a sparse solution on $\mathbf{h}$.

2) Two-Dimensional Coupled Building Block: To exploit the statistical dependencies across entries, we can introduce a coupled pattern on the hyperparameters, i.e., prior precisions. More precisely, the prior of each entry not only involves its own hyperparameter, but also the hyperparameters of its immediate neighbors on the 2-D AoD-AoA grids. Specifically, we have

$$p(\mathbf{h}|\boldsymbol{\alpha}) = \prod_{n=1}^{N} p(h_n|\alpha_n, \alpha_{\mathbb{N}(n)}). \tag{12}$$

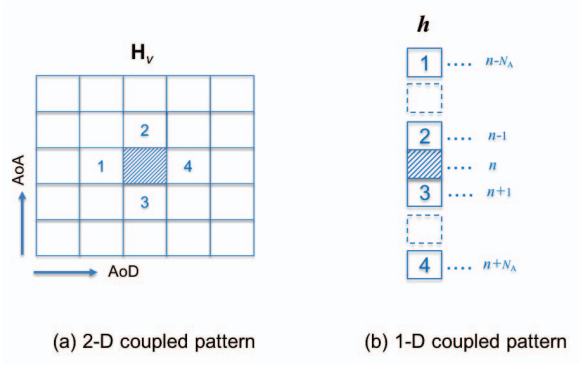


Fig. 2. The building block of (a) 2-D coupled pattern and (b) its equivalent 1-D coupled pattern for the clustered sparse mmWave channel estimation. N_A is the number of discretized grids in the AoA domain.

where $\mathbb{N}(n)$ denotes the indices of a defined neighboring grids to the grid n and

$$h_n|\alpha_n, \alpha_{\mathbb{N}(n)} \sim \mathcal{CN}(0, \left(\alpha_n + \beta f(\alpha_{\mathbb{N}(n)})\right)^{-1}) \tag{13}$$

where β is the coupling coefficient and $f(\cdot)$ is a function of the prior precisions at the neighboring grids $\mathbb{N}(n)$ of the index n. For instance, for a given grid n, we define its neighboring grids in the AoD-AoA domain as the four step-one grids (numbered from 1 to 4) in Fig. 2 (a). Due to the vectorization, the neighboring index set in the vector $\mathbf{h}$ is, hence, $\mathbb{N}(n) = \{n - N_A, n-1, n+1, n+N_A\}$, as shown in Fig. 2 (b). The coupling coefficient β quantitatively describes the statistical dependence between each entry h_n and its 2-D neighboring entries $h_{\mathbb{N}(n)}$. When $\beta = 0$, the coupled building block reduces to the conventional uncoupled model of (8).

The choice of the coupling function $f(\cdot)$ can be flexible. One choice of $f(\cdot)$ is a simple summation over the neighboring precisions [15], [16],

$$f(\alpha_{\mathbb{N}(n)}) = \alpha_{n-N_A} + \alpha_{n+N_A} + \alpha_{n-1} + \alpha_{n+1}. \tag{14}$$

To further capture the smooth decay of the channel power along the AoD and AoA domains, one may use the following coupled total variation functions

$$f(\alpha_{\mathbb{N}(n)}) = |\alpha_{n-N_A} - \alpha_{n+N_A}| + |\alpha_{n-1} - \alpha_{n+1}|, \tag{15}$$

$$f(\alpha_{\mathbb{N}(n)}) = |\alpha_{n-N_A} - \alpha_n| + |\alpha_n - \alpha_{n+N_A}| + |\alpha_{n-1} - \alpha_n| + |\alpha_n - \alpha_{n+1}|, \tag{16}$$

where $|\alpha_{n-N_A} - \alpha_n|$ captures the *channel power variation* along the AoD domain, $|\alpha_{n-1} - \alpha_n|$ along the AoA domain, and the summation between these two terms imposes the dependence of the four neighboring grids on the current grid.

B. Coupled Sparse Bayesian Channel Estimation

With the above signal model, we use the SBL framework in [17] to develop the channel estimation algorithm which is capable of exploiting the joint AoD-AoA spread. In general, the SBL framework has two steps. First the inference on the basis coefficient vector $\mathbf{h}$ is derived from the measurements $\mathbf{y}$, provided that the hyperparameters (the prior precision α_n and the noise variance σ^2) are known. Then the inference on the hyperparameters can be obtained by using the EM algorithm.

1) Bayesian Estimation of $\mathbf{h}$*:* First, one can compute the posterior distribution of $\mathbf{h}$ as

$$p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}) \propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\boldsymbol{\alpha}) \tag{17}$$

where $\mathbf{y}|\mathbf{h} \sim \mathcal{CN}(\mathbf{A}\mathbf{h}, \sigma^2\mathbf{I})$ according to (6) and $p(\mathbf{h}|\boldsymbol{\alpha})$ is given by (14). It is easy to infer that the posterior distribution of $\mathbf{h}$ is a complex Gaussian distribution given by

$$\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{18}$$

where

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\mathbf{A}^H\mathbf{y},$$
$$\boldsymbol{\Sigma} = \left(\sigma^{-2}\mathbf{A}^H\mathbf{A} + \boldsymbol{\Lambda}\right)^{-1} \tag{19}$$

with $\boldsymbol{\Lambda}$ is a diagonal matrix with its n-th diagonal entry given as

$$[\boldsymbol{\Lambda}]_{nn} = \alpha_n + \beta f(\alpha_{\mathbb{N}(n)}). \tag{20}$$

From (18), the Bayesian MAP estimate of $\mathbf{h}$, also the MMSE estimate here, is given by the posterior mean

$$\hat{\mathbf{h}} = \boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\mathbf{A}^H\mathbf{y}, \tag{21}$$

provided that the prior precisions $\boldsymbol{\alpha}$ (via $\boldsymbol{\Lambda}$ in $\boldsymbol{\Sigma}$) and the noise variance σ^2 are known.

2) An EM Update of Hyperparameters: The second step is to derive an update rule for the prior precisions[1]. To this purpose, we adopt the EM algorithm to estimate the hyperparameters $\{\alpha_n\}_{n=1}^N$, which maximizes the posterior probability $p(\boldsymbol{\alpha}|\mathbf{y})$ by treating the channel vector $\mathbf{h}$ as a hidden variable and maximizing the expected value of the complete log-posterior of $\boldsymbol{\alpha}$. Specifically, it has two steps: the expectation (E) step and the maximization (M) step.

The E-step computes the expected value of the complete log-posterior of $\boldsymbol{\alpha}$, also defined as the Q-function, over the hidden variable $\mathbf{h}$, provided the measurements $\mathbf{y}$ and the current estimate of $\boldsymbol{\alpha}^{(\ell)}$ from the previous iteration. From the above signal model, the Q-function is given as

$$Q(\boldsymbol{\alpha}) = E_{\mathbf{h}|\mathbf{y},\boldsymbol{\alpha}^{(\ell)}} \log p(\boldsymbol{\alpha}|\mathbf{h}) \propto \int \log\left[p(\mathbf{h}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})\right] p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}) d\mathbf{h} \tag{22}$$

With $p(\mathbf{h}|\boldsymbol{\alpha})$ of (14), $p(\alpha_n)$ of (10) and the posterior distribution of (17) and defining $\eta_n = \alpha_n + \beta f(\alpha_{\mathbb{N}(n)})$ we have

$$Q(\boldsymbol{\alpha}) \propto \log p(\boldsymbol{\alpha}) + \sum_{n=1}^{N}\left[\log\eta_n - \eta_n \int p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha})|h_n|^2 d\mathbf{h}\right]$$
$$\overset{(a)}{\propto} \sum_{n=1}^{N}\left[(\mu - 1)\log\alpha_n - \nu\alpha_n + \log\eta_n - \eta_n(|\mu_n|^2 + [\boldsymbol{\Sigma}]_{nn})\right] \tag{23}$$

where (a) is due to the Gamma distribution of $\boldsymbol{\alpha}$ in (10) and the posterior distribution of $\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}$ in (18).

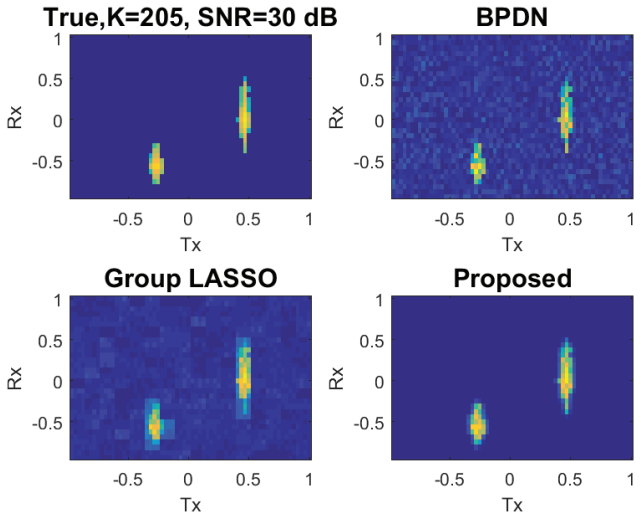


Fig. 3. **Scenario 1** — 2 path clusters and 57 paths: the estimated channel gain matrix over the two-dimensional AoD-AoA domain with $K = 205$ training signals and SNR = 30 dB.

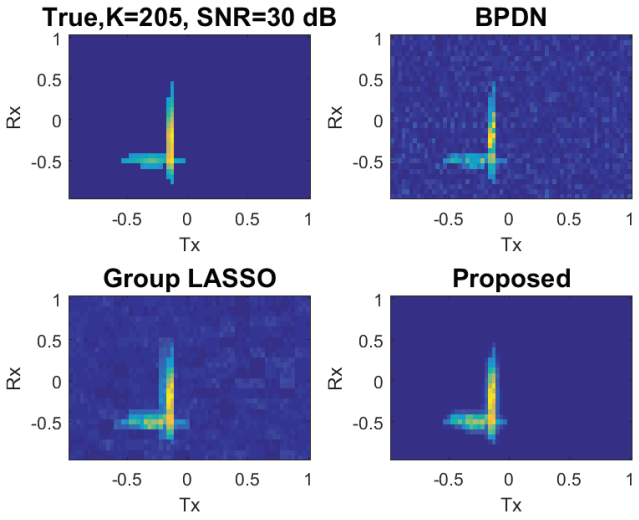


Fig. 5. **Scenario 2** — 2 path clusters and 67 paths: the estimated channel gain matrix over the two-dimensional AoD-AoA domain with $K = 205$ training signals.

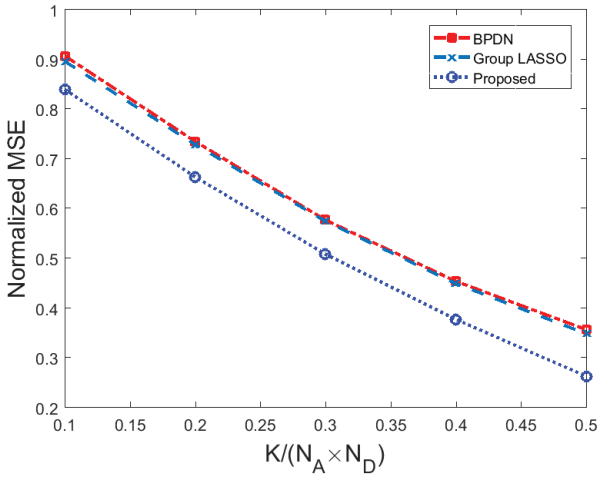


Fig. 4. The normalized MSE versus the number of training signals for Scenario 1.

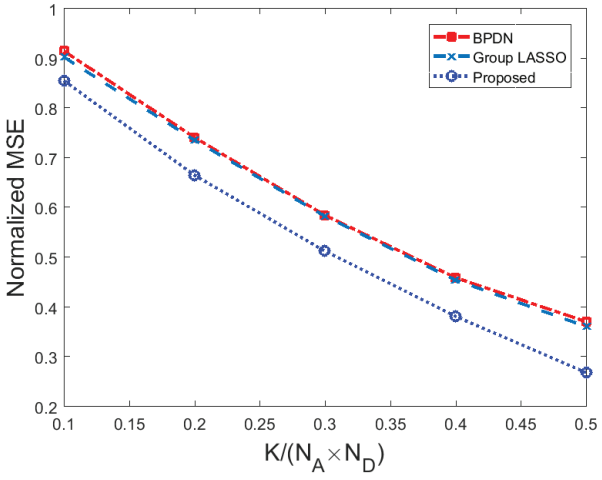


Fig. 6. The normalized MSE versus the number of training signals for Scenario 2.

Next, the M-step is to maximize the above Q-function to find a new estimate of $\boldsymbol{\alpha}$, i.e.,

$$\boldsymbol{\alpha}^{(\ell+1)} = \arg\max_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(\ell)}) \tag{24}$$

Due to the coupled structure in (23), i.e., $\log(\alpha_n+\beta f(\alpha_{\mathbb{N}(n)}))$, the maximization of the Q-function cannot be decoupled into a number of separable optimizations as the traditional SBL [17]. Even for the simple coupled function of (14), an exact solution to (24) cannot be found in closed form. As a result, gradient descend methods can be used to search for the optimal solution[2].

V. Numerical Results

We now present simulation results to compare the proposed channel estimation algorithm with existing algorithms. Particularly, we consider a) the basis pursuit denoising (BPDN) or the LASSO with the SPGL1 solver [8], [19] and (b) the group LASSO (implemented by using the SPGL1-Group solver) with given block patterns. We consider a system model consisting of 64 ULA antennas at the BS and 32 ULA antennas at MS. The

[1]For better exposure of our derivations, we focus here on the derivation for updating the prior precisions only. The derivation can be straightforwardly extended to the case when the noise variance is unknown.

[2]Alternatively, the coupling among different α's in each iteration may be approximately untangled with an *iteration decoupling* scheme by replacing those variables involving α_n by their estimates from the previous iteration [18].

mmWave channel is assumed to follow a geometric channel model with the AoAs and AoDs distributed in $[-\pi/2, \pi/2]$. To illustrate the performance, we run 100 independent Monte-Carlo simulations for the two scenarios with angular path power profiles given in Fig 2. For each Monte-Carlo run, the two-dimensional channel gain matrix $\mathbf{H}_v$ (and hence the corresponding channel vector $\mathbf{h}$) is generated as a complex sub-Gaussian matrix with zero mean and variance given by the angular path power specified in Fig. 1. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \|\mathbf{Ah}\|_2^2/(K\sigma^2)$ where σ^2 is the variance of the noise vector $\mathbf{e}$.

First, we consider the scenario in Fig 2 (a), referred to as Scenario 1 here, with two separated path clusters spreading in the AoA domain. Fig. 3 shows the estimated channel gain matrix by averaging the result from 100 Monte-Carlo runs with $K = 205$ training signals and $\text{SNR} = 30$ dB. In this case, the number of significant paths is 57. It is clear to see that the proposed method is able to preserve the two-dimensional path clusters better than the existing channel estimation methods as the proposed channel estimation method favors the clustered solution. Then we vary the number of training signals K and the normalized MSE is shown as a function of K in Fig. 4.

Next, we consider the scenario in Fig 2 (b) referred to as Scenario 2 here. Fig. 5 shows the estimated channel gain matrix over the two-dimensional AoD-AoA domain from 100 Monte-Carlo runs when $K = 205$. The number of significant paths is 67 in this case. Again, it is seen that the proposed method gives the more clustered channel gain matrix than the other considered methods. Fig. 6 shows the performance in terms of the MSE as a function of K. Compared with the result in Fig. 4 for Scenario 1, Fig. 6 shows slightly higher MSE with the same number of training signals due to the slightly higher number of significant paths.

VI. Conclusion

In this paper, we have proposed a sparse channel estimation for mmWave communication systems and exploits the joint angular spread of the mmWave channel matrix over the two-dimensional AoD-AoA domain. By formulating the sparse channel estimation as a block-sparse recovery problem with an underlying two-dimensional coupled structure, we proposed the coupled sparse Bayesian learning which, building on a hierarchical Gaussian prior model, uses the Bayesian inference for the channel matrix estimation and makes use of the EM algorithm to update the hyperparameters associated with the hierarchical signal model. Compared with existing sparse mmWave channel estimation methods, the proposed method can further reduce the training overhead and improve the channel estimation error.

VII. Acknowledgements

We would like to thank Dr. Hassan Mansour at MERL for insightful discussions on block sparse recovery methods and pointing out the SPGL1 solver.

References

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

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

[3] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, October 2014.

[4] A. Alkhateeby, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *2015 IEEE ICASSP*, April 2015, pp. 2909–2913.

[5] 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 *2013 IEEE ICC*, June 2013, pp. 5163–5167.

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

[7] M. R. Akdeniz et. al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[8] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[9] R. T. Suryaprakash, M. Pajovic, K. J. Kim, and P. Orlik, "Millimeter wave communications channel estimation via Bayesian group sparse recovery," in *2016 IEEE ICASSP*, March 2016, pp. 3406–3410.

[10] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, 2006.

[11] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. on Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[12] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.

[13] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 4, pp. 259–269, 2012.

[14] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Trans. on Signal Processing*, vol. 60, no. 5, pp. 2286–2303, May 2012.

[15] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. on Signal Processing*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[16] H. Duan, L Zhang, J. Fang, L. Huang, and H. Li, "Pattern-coupled sparse Bayesian learning for inverse synthetic aperture imaging," *IEEE Signal Processing Letters*, vol. 22, pp. 1995–1999, Nov. 2015.

[17] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, August 2004.

[18] P. Wang, J. Fang, and P. Orlik et. al., "Fast approximate decoupled iteration schemes for coupled sparse Bayesian learning," *in preparation*, 2017.

[19] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.