Learning the Structured Sparsity: 3-D Massive MIMO Channel Estimation and Adaptive Spatial Interpolation

Yurong Wang, Aijun Liu, *Member, IEEE*, Xiaochen Xia, and Kui Xu, *Member, IEEE*

***Abstract*—This paper addresses the channel estimation problem for three-dimensional (3-D) massive multiple-input multiple-output (MIMO) systems, where the base station (BS) is equipped with a two-dimensional uniform planar array (UPA) to serve a number of user equipments (UEs). To implement with low hardware complexity, the number of available radio-frequency (RF) chains at BS is constrained to be much smaller than the number of antennas. The theoretical analysis of sparse property for 3-D massive MIMO channel reveals that there exists two kinds of sparse structures in beam-domain channel vector, namely the common sparsity structure among sub-arrays of UPA and block sparsity structure per sub-array. Based on this property, a novel structured sparse Bayesian learning framework is proposed to estimate the channel between BS and UE reliably. Moreover, an adaptive spatial interpolation scheme is proposed to further reduce the number of required RF chains at BS while maintaining the estimation performance. The simulation results show that the proposed scheme provides stable estimation performance for a variety of scenarios with different numbers of RF chains, transmit signal-to-noise ratios, Rician factors, and angular spreads, and outperforms the reference schemes significantly.**

***Index Terms*—Channel estimation, 3-D massive MIMO systems, channel sparsity structures, structured SBL, adaptive spatial interpolation.**

I. Introduction

THE substantial traffic growth driven by emerging applications, such as ultra-high definition video, virtual and augmented realities, have made system capacity enhancement one of the most important features in the fifth-generation (5G) and beyond mobile communication systems. Massive multiple-input multiple-output (MIMO) proposed by [1] is identified as a key technology for fulfilling the traffic requirement.

In the massive MIMO systems, the base station (BS) employs a large-scale antenna array with hundreds or more antenna elements to several tens of user equipments (UEs) using the same time-frequency resource. Through multi-UE beamforming, massive MIMO can produce tremendous multiplexing and array gains, which boost the system spectral efficiency significantly [2]–[4]. However, efficient beamforming depends on the reliable estimation of channel state information (CSI) between BS and UEs. In time-division duplex (TDD) system, by exploiting the channel reciprocity, it is generally accepted that the uplink and downlink CSI can be obtained by pilot-aided uplink training with training overhead scaling linearly with the number of UEs [4].

Despite of its great benefits, each antenna in the large-scale antenna array must be integrated with its own radio-frequency (RF) chain, which leads to high hardware complexity and circuit energy consumption. Many recent works seek to build low-complexity massive MIMO with limited number of RF chains (See [5], [6] and the references therein). In these schemes, the large-scale antenna array is connected to a number of RF chains much smaller than the number of antennas through a phase-shift network in analog-domain, which is known as hybrid analog-digital (HAD) architecture. Although reducing the hardware complexity greatly, such architecture poses a significant challenge for channel estimation. This is because the BS now cannot obtain the training signals from all antennas, but several combinations of them. For the network with L UEs and a BS equipped with N antennas and R RF chains, the analysis in [7] showed that the conventional least square (LS) channel estimation scheme in HAD architecture requires the pilot symbols not less than $\frac{LN}{R}$. Since usually $R \ll N$, the conventional LS scheme needs very long training phase which degrades the system spectral efficiency greatly.

In the practical system, the BS is usually elevated at a relatively high altitude, e.g., on the top of a high building or a dedicated tower or even in the air [8], [9], therefore there are few surrounding scatters. The received signal at BS is mainly caused by the scattering process in vicinity of UE, which results in very narrow signal angle of arrival (AoA) at BS. Based on a physical channel model, several works have shown that the signals received from different antennas of BS exhibit high correlation, and hence the channel between BS and UE can be represented by a sparse vector in some alternate domain (usually called beam-domain) [10]–[13]. By assuming the perfect knowledge of channel sparsity (e.g., the positions of significant elements in beam-domain channel vector), the length of required

Manuscript received November 27, 2018; revised February 21, 2019; accepted March 31, 2019. Date of publication April 4, 2019; date of current version November 12, 2019. The work was supported in part by the National Natural Science Foundation of China under Grants 61671472, 61671476, and 61501508, and in part by the Jiangsu Province Natural Science Foundation under Grant BK20160079. The review of this paper was coordinated by Prof. Y. Li. *(Corresponding author: Aijun Liu.)*

Y. Wang, A. Liu, and K. Xu are with the Army Engineering University of PLA, Nanjing 210007, China (e-mail: wyrbitzj@sina.com; liuaj.cn@163.com; lgdxxukui@sina.com).

X. Xia is with the Army Engineering University of PLA, Nanjing 210007, China, and 31682 Army of PLA, Lanzhou 730050, China (e-mail: tjuxxc@sina.com).

Digital Object Identifier 10.1109/TVT.2019.2909392

0018-9545 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

pilot sequence can be reduced by constructing equivalently low-dimension beam-domain channel estimation schemes [10], [13].

However, the prior of channel sparsity is not always available. In millimeter wave MIMO system, the beam cycling method can be employed to estimate the channel without the prior of channel sparsity [14]. But the resultant training overhead is proportional to the number of BS antennas. In [15], [16], the channel estimation was formulated as a sparse recovery problem, and addressed using compressive sensing algorithms such as orthogonal matching pursuit (OMP) [17]. If the UEs appear in cluster, different UEs may experience some common scatters, which makes the channels of different UEs partially share some common sparsity [18] (which we refer to as UE shared sparsity). By exploiting the UE shared sparsity, several schemes were proposed to improve the quality of channel estimation [18]–[20]. A modified OMP algorithm, called joint OMP (JOMP), was proposed in [18] to promote the UE shared sparsity in channel estimation. In [19], a reweighted block l_1/l_2 minimization algorithm was proposed and shown to outperform JOMP. Under the Bayesian framework, a novel Gaussian mixture prior for beam-domain channel was introduced to capture the UE shared sparsity [20], and the channel estimation was addressed by formulating a variational sparse Bayesian learning (SBL) problem. Note that the performances of these schemes [18]–[20] rely partly on the assumption of physical vicinity of UEs. When UEs are randomly located, the performance improvement may be reduced. On the other hand, due to nature of the physical scattering process, it is observed that the elements of beam-domain channel with significant amplitude appear in block [12], [13] which is called block or burst sparsity. By exploiting this structure, a novel channel estimation scheme based on burst LASSO algorithm was proposed in [21] to further improve estimation performance by assuming partial knowledge of channel sparsity, such as the length of the sparse block.

In this paper, we consider the channel estimation for TDD three-dimension (3D) massive MIMO system, where the BS is equipped with a two-dimension (2D) uniform planar array (UPA) to serve a number of UEs. It is assumed that the number of available RF chains at BS is constrained to be much smaller than the number of antennas. Comparing to ULA, UPA places the antennas in a 2D grid, and thus enables a large number of antennas to be deployed in a compact area. Note that all the aforementioned works assumed ULA at BS, which cannot be directly utilized in 3D massive MIMO systems. The contributions of the paper are summarized as follows.

- The sparse property of 3D massive MIMO channel in beam-domain is analyzed theoretically. The results show that the beam-domain channel vector exhibits two kinds of sparse structures, namely the common sparsity structure among sub-arrays of UPA and block sparsity structure per sub-array.
- By exploiting the sparse structures, a structured prior for beam-domain channel is proposed and a novel structured SBL framework is formulated for uplink channel estimation. The problem is solved iteratively using a block successive majorization-minimization (SMM) algorithm [22] with low complexity. Different from [18]–[20], [21], the structured SBL does not require the physical vicinity of UEs and additional knowledge of channel sparsity.
- An adaptive spatial interpolation scheme is proposed to further reduce the number of required RF chains at BS while maintaining the estimation performance. Contrary to the common belief that channel from every UPA's antenna to UE needs be estimated from the training signal, the scheme first samples a few sub-arrays from the UPA and performs structured SBL to estimate channel between the sampled sub-arrays and UEs. Then the complete channel vector is recovered by implementing an adaptive spatial interpolation algorithm.
- Extensive simulations are presented to validate the superiority of proposed scheme. The results show that the proposed scheme provides stable estimation performance for a variety of scenarios with different numbers of RF chains, transmit signal-to-noise ratios (SNRs), Rician factors and angular spreads, and outperforms the reference schemes significantly.

The rest of the paper is organized as follows. Section II presents the system model. Section III presents the channel sparsity analysis. Section IV presents the channel estimation scheme based on structured SBL. Section V presents the adaptive spatial interpolation scheme. Section VI presents simulation results and Section VII draws the conclusions.

Notations: $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\mathrm{tr}(\cdot)$, and $\mathrm{vec}(\cdot)$ denote conjugate, transpose, conjugate-transpose, trace and vectorization operators of matrix, respectively. $\mathbf{X}^\dagger$ denotes the right pseudo inverse of matrix $\mathbf{X}$, which is defined as $\mathbf{X}^\dagger = \mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}$. For an $m \times n$ matrix $\mathbf{X}$, if $\mathbf{x} = \mathrm{vec}\{\mathbf{X}\}$, then $\mathbf{X} = \mathrm{vec}_{m,n}^{-1}\{\mathbf{x}\}$.$\mathrm{cond}\{\mathbf{X}\}$ returns the condition number of $\mathbf{X}$. $\otimes$ denotes the Kronecker product. $\mathbf{e}_m$ denotes the mth unitary vector. $\mathbf{X}(:, A)$ ($\mathbf{X}(A, :)$) denotes the sub-matrix of $\mathbf{X}$ by selecting the columns (rows) indexed by set A. $|x|$ returns the amplitude of the complex variable x. $A \times B$ denotes the Cartesian product of two sets A and B. $|A|$ denotes the cardinality of set A. $\mathbb{E}[\cdot]$ denotes expectation. $\mathcal{N}_C(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\Gamma(x)$ denotes the gamma function [23, 8.310.1].

II. System Model

Consider a network with a BS and L single-antenna UEs, where the BS is equipped with a 2D UPA. The numbers of antennas in two dimensions of UPA (named as x-direction and y-direction) are N and M respectively as shown in Fig. 1. The corresponding antenna spacings are d_x and d_y. To reduce the hardware complexity, the BS employs a HAD architecture where each RF chain is connected to a subset of antennas with phase-shift network. The total number of available RF chains is R which is assumed to be much smaller than MN.

A. Channel Model

A general Rician model is employed for the channel vector between BS and UE l, that is

$$\mathbf{h}_l = \mathbf{h}_l^{\mathrm{LoS}} + \mathbf{h}_l^{\mathrm{NLoS}} \in \mathbb{C}^{MN \times 1}. \tag{1}$$

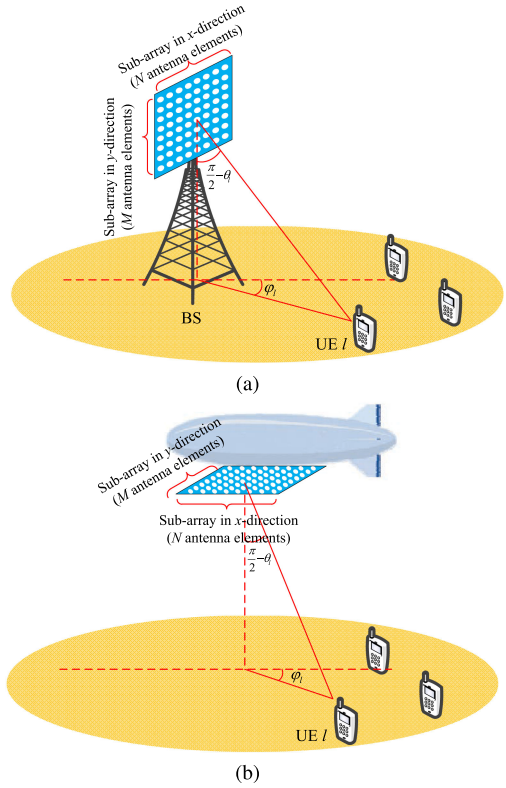


Fig. 1. Illustration of system model. (a) Terrestrial communication system. (b) Hybrid air-terrestrial communication system.

where $\mathbf{h}_l^{\mathrm{LoS}}$ denotes the deterministic line-of-sight (LoS) component and $\mathbf{h}_l^{\mathrm{NLoS}}$ denotes the random non-LoS (NLoS) component, which are modeled as

$$\mathbf{h}_l^{\mathrm{LoS}} = \sqrt{\frac{\beta_l K_l}{K_l+1}}\mathbf{a}_M\left(\rho_{l,y}\right)\otimes\mathbf{a}_N\left(\rho_{l,x}\right),$$

$$\mathbf{h}_k^{\mathrm{NLoS}} = \sqrt{\frac{\beta_l}{K_l+1}}\int_{\rho_{l,x}-\Delta_x}^{\rho_{l,x}+\Delta_x}\int_{\rho_{l,y}-\Delta_y}^{\rho_{l,y}+\Delta_y} r\left(\rho_x,\rho_y\right) \times \mathbf{a}_M\left(\rho_y\right)\otimes\mathbf{a}_N\left(\rho_x\right)d\rho_x d\rho_y. \quad (2)$$

In (2), $\mathbf{a}_n\left(\rho\right)=\left[1,\exp\left(\jmath 2\pi\rho\right),\ldots,\exp\left(\jmath 2\pi\left(n-1\right)\rho\right)\right]^T$ denotes the array response vector. β_l and K_l denote the large-scale fading and Rician factor, respectively. We call $\rho_{l,x}$ and $\rho_{l,y}$ as the virtual AoAs of UE l along x- and y-direction respectively with Δ_x and Δ_y denoting the corresponding angular spreads. $r_l(\rho_x,\rho_y)$ denotes the complex response gain for virtual AoA $\{\rho_x,\rho_y\}$. The values of $\rho_{l,x}$, $\rho_{l,y}$ and K_l depend on the placement of antenna array. For example, in the terrestrial communication systems (Fig. 1(a)), the UPA is usually deployed perpendicularly to ground. In this case, we have $\rho_{l,x}=\frac{d_x}{\lambda}\cos\theta_l\sin\varphi_l$ and $\rho_{l,y}=\frac{d_y}{\lambda}\sin\theta_l$ [25], [26], where θ_l and φ_l denote the physical AoAs of UE l in vertical and horizontal directions, and λ denotes the carrier wavelength. Moreover, the channel may have very weak LoS component ($K_l\to 0$) due to the ground obstacles. In the hybrid air-terrestrial communication systems [8], [9] (Fig. 1 (b)), the UPA can be deployed on an air platform and is parallel to ground. In this case, we have $\rho_{l,x}=\frac{d_x}{\lambda}\cos\theta_l\sin\varphi_l$ and $\rho_{l,y}=\frac{d_y}{\lambda}\cos\theta_l\cos\varphi_l$ [8], [9]. Moreover, the channel usually has non-negligible LoS component since the BS (air platform) is much higher than ground obstacles.

B. Uplink Training

We consider uplink training which is usually adopted in TDD massive MIMO systems. In the training phase, all UEs transmit their pilots to BS simultaneously and BS estimates channels from received signals. Let $\mathbf{p}_l\in\mathbb{C}^{T_p\times 1}$ denote the pilot sequence sent by UE l which satisfies $\mathbf{p}_l^H\mathbf{p}_l=T_pP$ and $\mathbf{p}_l^H\mathbf{p}_{l'}=0$ ($l'\neq l$), where T_p denotes the length of pilot sequence and P denotes the power of each pilot symbol. The received training signal at BS can be expressed as

$$\tilde{\mathbf{Y}}=\mathbf{W}_{\mathrm{full}}^H\mathbf{H}\mathbf{P}^T+\tilde{\mathbf{N}}\in\mathbb{C}^{R\times T_p}, \quad (3)$$

where $\mathbf{P}=[\mathbf{p}_1,\ldots,\mathbf{p}_L]$ and $\mathbf{H}=[\mathbf{h}_1,\ldots,\mathbf{h}_L]$. $\mathbf{W}_{\mathrm{full}}\in\mathbb{C}^{MN\times R}$ denotes the RF receive processing matrix which satisfies $\mathbf{W}_{\mathrm{full}}^H\mathbf{W}_{\mathrm{full}}=\mathbf{I}_R$. $\tilde{\mathbf{N}}\in\mathbb{C}^{R\times T_p}$ denotes additive white Gaussian noise matrix with zero-mean and variance σ^2. By multiplying both sides of (3) with $\sqrt{\frac{1}{T_pP}}\mathbf{P}^*$, we have

$$\mathbf{Y}=\sqrt{\frac{1}{T_pP}}\tilde{\mathbf{Y}}\mathbf{P}^*=\sqrt{T_pP}\mathbf{W}_{\mathrm{full}}^H\mathbf{H}+\mathbf{N}\in\mathbb{C}^{R\times L}, \quad (4)$$

where $\mathbf{N}=\sqrt{\frac{1}{TP}}\tilde{\mathbf{N}}\mathbf{P}^*$. It is not difficult to see that $\mathbf{N}$ has the same distribution with $\tilde{\mathbf{N}}$. The channel vector between BS and UE l is estimated from the lth column of $\mathbf{Y}$, that is

$$\mathbf{y}_l=\sqrt{T_pP}\mathbf{W}_{\mathrm{full}}^H\mathbf{h}_l+\mathbf{n}_l\in\mathbb{C}^{R\times 1}. \quad (5)$$

III. Analysis of Channel Sparsity and Design of RF Receive Processing Matrix

In this and next sections, we will temporarily consider the situation with $N\gg M$, that is, the number of antennas in x-direction is much greater than that in y-direction. This may arise from the practical constraint on the size of antenna array. Another example of such configuration is the distributed massive MIMO systems where the antenna array of BS is constituted of M distributed ULAs with each containing N antennas. In this case, it is usually true that $N\gg M$. In Section V, we will extend the proposed scheme to the situation when both M and N are large.

A. Structured Channel Sparsity

Consider the matrix $\mathbf{B}=\sqrt{\frac{1}{MN}}\mathbf{I}_M\otimes\mathbf{A}_N\in\mathbb{C}^{MN\times MN}$, where $\mathbf{A}_N$ is a shift-version DFT matrix of dimension N, i.e.,

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

$$\mathbf{a}_N\left(\rho\right)=\left[1,\exp\left(\jmath 2\pi\rho\right),\ldots,\exp\left(\jmath 2\pi\left(N-1\right)\rho\right)\right]^T. \quad (6)$$

The columns of $\mathbf{B}$ form a set of orthogonal basis of dimension MN. The projection of channel vector $\mathbf{h}_l$ onto the basis can be

expressed as

$$\mathbf{z}_l = \mathbf{B}^H \mathbf{h}_l \tag{7}$$

In the large N regime, we have the following lemma on $\mathbf{z}_l$.

Lemma 1. Let $\mathbf{b}_{m,n}$ denote the $((m-1)N+n)$th basis in $\mathbf{B}$ (i.e., the $((m-1)N+n)$th column of $\mathbf{B}$), which can be written as $\mathbf{b}_{m,n} = \sqrt{\frac{1}{MN}}\mathbf{e}_m \otimes \mathbf{a}_N\left(-\frac{1}{2}+\frac{n-1}{N}\right)$. In the large N regime, the projection of $\mathbf{h}_l$ onto $\mathbf{b}_{m,n}$ (i.e., the $((m-1)N+n)$th element of $\mathbf{z}_l$) has significant amplitude only if $-\frac{1}{2}+\frac{n-1}{N} \in [\rho_{l,x}-\Delta_x, \rho_{l,x}+\Delta_x]$.

Note that this result can be proven using several methods developed in [9], [12], [13]. We omit the detailed proof due to space limitation.

In (7), $\mathbf{z}_l$ can be viewed as the representation of $\mathbf{h}_l$ in the domain spanned by basis $\mathbf{B}$. Following the convention of massive MIMO literature [10]–[13], we call each basis in $\mathbf{B}$ as a beam and call $\mathbf{z}_l$ as the beam-domain channel. Moreover, we call the index set of effective beams $\mathbf{b}_{m,n}$ with significant beam-domain channel amplitude $\left|z_{l,(m-1)N+n}\right|$ as beam support. From Lemma 1, the beam support can be expressed as $\Omega_l = \Omega_l^x \times \Omega_l^y$ with $\Omega_l^x = \left\{n \left| -\frac{1}{2}+\frac{n-1}{N} \in [\rho_{l,x}-\Delta_x, \rho_{l,x}+\Delta_x], n \in \mathbb{N}^+\right.\right\}$ and $\Omega_l^y = \{m | m \in \{1,\ldots,M\}\}$. The introduction of beam support will play an important role when extending to the situation with large M. As long as $\mathbf{z}_l$ is estimated, the channel $\mathbf{h}_l$ can be recovered by basis expansion model $\mathbf{h}_l = \mathbf{B}\mathbf{z}_l$.

To analyze the inherent structures of channel sparsity, we can express $\mathbf{h}_l$ and $\mathbf{z}_l$ as follows

$$\begin{aligned}
\mathbf{h}_l = \Bigg[&\underbrace{h_{l,1},\cdots h_{l,N}}_{\mathbf{h}_{l,1}},\ldots,\underbrace{h_{l,(m-1)N+1},\cdots h_{l,mN}}_{\mathbf{h}_{l,m}},\ldots, \\
&\underbrace{h_{l,(M-1)N+1},\cdots h_{l,MN}}_{\mathbf{h}_{l,M}}\Bigg]^T, \\
\mathbf{z}_l = \Bigg[&\underbrace{z_{l,1},\cdots z_{l,N}}_{\mathbf{z}_{l,1}},\ldots,\underbrace{z_{l,(m-1)N+1},\cdots z_{l,mN}}_{\mathbf{z}_{l,m}},\ldots, \\
&\underbrace{z_{l,(M-1)N+1},\cdots z_{l,MN}}_{\mathbf{z}_{l,M}}\Bigg]^T.
\end{aligned} \tag{8}$$

Note that the mth sub-vector $\mathbf{h}_{l,m}$ is the channel between UE l and the mth sub-array of BS in x-direction (See Fig. 1 for the definition of sub-array). Transforming $\mathbf{h}_l$ and $\mathbf{z}_l$ into matrix forms, that is, $\mathbf{H}_l = [\mathbf{h}_{l,1},\ldots,\mathbf{h}_{l,M}] \in \mathbb{C}^{N\times M}$ and $\mathbf{Z}_l = [\mathbf{z}_{l,1},\ldots,\mathbf{z}_{l,M}] \in \mathbb{C}^{N\times M}$, we have

$$\mathbf{Z}_l = \sqrt{\frac{1}{MN}}\mathbf{A}_N^H \mathbf{H}_l, \tag{9}$$

where the equality $(\mathbf{U}_2 \otimes \mathbf{U}_1)\,\mathrm{vec}(\mathbf{V}) = \mathrm{vec}\left(\mathbf{U}_1\mathbf{V}\mathbf{U}_2^T\right)$ is used. Therefore, $\mathbf{z}_{l,m}$ can be viewed as the beam-domain channel between UE l and the mth sub-array in x-direction under basis $\sqrt{\frac{1}{MN}}\mathbf{A}_N$. Note that nth element of $\mathbf{z}_{l,m}$ stores the $((m-1)N+n)$th element of $\mathbf{z}_l$.

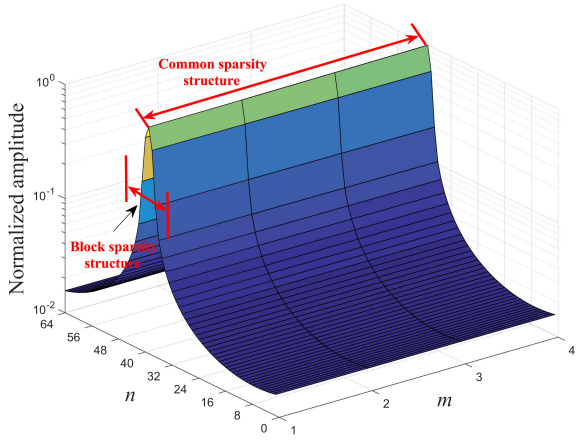


Fig. 2. Normalized amplitudes of the elements in $\mathbf{Z}_l$, where $N = 64$, $M = 4$, $K_l = -30$ dB, $\theta_l = 65°$, and $\varphi_l = 27°$. The angular spread is 5°. The figure is plotted by averaging the results for 10^2 independent channel realizations.

As discussed in Section I, the received signal at BS is mainly due to the scattering process in vicinity of UEs, which results in very narrow signal AoAs in vertical and horizontal directions (and hence very small Δ_x and Δ_y). Therefore, from Lemma 1, only a small fraction of elements in each sub-vector $\mathbf{z}_{l,m}$ have significant amplitudes. More precisely, $\mathbf{z}_l$ exhibits two kinds of sparse structures:

- Common sparsity structure among sub-arrays: The nth element of $\mathbf{z}_{l,m}$ has significant amplitude only when $n \in \Omega_l^x$. Since Ω_l^x is independent of m, the indexes of significant elements in each sub-vector $\mathbf{z}_{l,m}$ are identical.
- Block sparsity structure per sub-array: The indexes of significant elements in each sub-vector $\mathbf{z}_{l,m}$ appear in block.

The major difference between the two kinds of structures lies in that the block sparsity structure indicates the sparse structure in the beam-domain channel between a specific sub-array and UE, while the common sparsity structure emphasizes the similarity in the positions (i.e., index sets) of significant elements for the beam-domain channels associated to different sub-arrays. The two kinds of sparse structures are illustrated in Fig. 2 for an example with $N = 64$ and $M = 4$. In this example, we have $\Omega_l^x = \{35, 36, 37, 38, 39, 40, 41, 42, 43\}$. Here we shall point out that the nth element of beam-domain channel $\mathbf{z}_{l,m}$ with $n \notin \Omega_l^x$ is small but not zero. Therefore, $\mathbf{z}_{l,m}$ is only *approximately* sparse.

In a number of previous works [24]–[26], the orthogonal basis is commonly chosen as the Kronecker product of an $M \times M$ DFT matrix and an $N \times N$ DFT matrix (which will be discussed in detailed in Section V-A). The popularity of such a configuration is because, when M and N are large enough, each array response vector becomes a DFT vector due to the increasing of UPA's spatial resolution [24]–[26], and the channel in (1) and (2) approaches to the weighted sum of a number of Kronecker products between DFT vectors. Thus, the projection of channel onto this basis will give rise to the sparsest representation. However, as seen in the analysis above, the basis employed in our work is quite different. The reason of such a choice is that, we not only

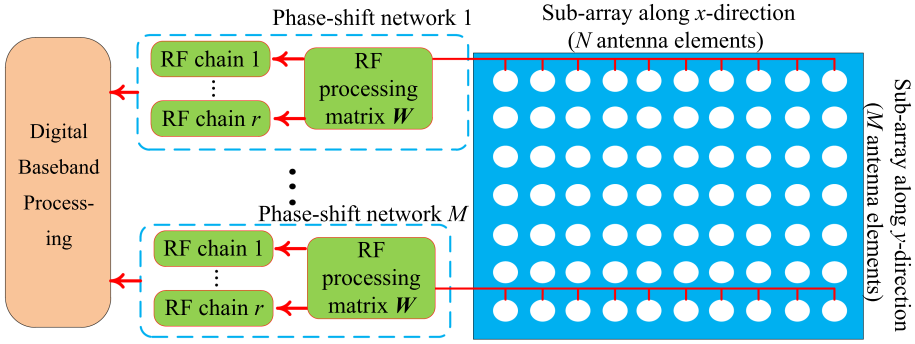


Fig. 3. Illustration of RF processing configuration.

expect a sparse beam-domain channel representation, but also a *structured* one. As will be seen in the following sections, the structures in sparsity play a key role in designing robust channel estimation schemes when $M \ll N$ as well as when both M and N are very large.

B. Design of RF Receive Processing Matrix

We assume that each sub-array in x-direction is connected to r RF chains through a separate phase-shift network as shown in Fig. 3, where r satisfies $Mr = R$. This is known as non-overlapped sub-array (NOSA) configuration [28]. NOSA is a practical architecture that can reduce the complexity significantly (e.g., in term of saving analog phase shifters) to realize HAD processing [27], [28]. Under NOSA configuration, the RF receive processing matrix takes a block-diagnal form

$$\mathbf{W}_{\text{full}} = \text{diag}\left\{\mathbf{W}, \ldots, \mathbf{W}\right\} = \mathbf{I}_M \otimes \mathbf{W}, \tag{10}$$

where $\mathbf{W} \in \mathbb{C}^{N\times r}$ denotes the RF receive processing matrix for each sub-array, and $\mathbf{W}^H\mathbf{W} = \mathbf{I}_r$.

Now we consider the requirements on RF receive processing matrix $\mathbf{W}_{\text{full}}$.

1) Constant Amplitude: Since each element of $\mathbf{W}$ corresponds to an analog phase shifter in the phase-shift network, $\mathbf{W}$ is a constant-amplitude matrix which satisfies $|[\mathbf{W}]_{i,j}| = \sqrt{\frac{1}{N}}$, $\forall i \in \{1, \ldots, N\}, j \in \{1, \ldots, r\}$. Equivalently, the jth column of $\mathbf{W}$ (denoted by $\mathbf{w}_j$) should satisfy

$$\text{diag}\left\{\mathbf{w}_j\mathbf{w}_j^H\right\} = \frac{1}{N}\mathbf{I}_N, \forall j \in \{1, \ldots, r\}. \tag{11}$$

2) Omnidirectional Reception: Using the basis expansion model $\mathbf{h}_l = \mathbf{B}\mathbf{z}_l$, the received training signal from UE l can be rewritten as

$$\begin{aligned}\mathbf{y}_l &= \sqrt{TP}\mathbf{W}_{\text{full}}^H \sum_{m=1}^{M}\sum_{n=1}^{N} \mathbf{b}_{m,n} z_{l,(m-1)N+n} + \mathbf{n}_l \\ &\approx \sqrt{TP}\mathbf{W}_{\text{full}}^H \sum_{\{m,n\}\in\Omega_l} \mathbf{b}_{m,n} z_{l,(m-1)N+n} + \mathbf{n}_l. \end{aligned} \tag{12}$$

From (12), the role of $\mathbf{W}_{\text{full}}$ is to coherently combine the signals on beams $\{\mathbf{b}_{m,n}\}_{\{m,n\}\in\Omega_l}$. However, since BS does not have the prior of Ω_l, it is natural to make the received signals on all MN beams $\{\mathbf{b}_{m,n}\}$ for $m \in \{1, \ldots, M\}$ and $n \in \{1, \ldots, N\}$ have the same combining gain. Therefore, $\mathbf{W}_{\text{full}}$ should satisfy

$$\begin{aligned}&\text{diag}\left\{\mathbf{B}^H\mathbf{W}_{\text{full}}\mathbf{W}_{\text{full}}^H\mathbf{B}\right\} \\ &= \frac{1}{MN}\text{diag}\left\{\left(\mathbf{I}_M \otimes \left(\mathbf{A}_N^H\mathbf{W}\right)\right)\left(\mathbf{I}_M \otimes \left(\mathbf{W}^H\mathbf{A}_N\right)\right)\right\} \\ &= \frac{1}{MN}\text{diag}\left\{\mathbf{I}_M \otimes \left(\mathbf{A}_N^H\mathbf{W}\mathbf{W}^H\mathbf{A}_N\right)\right\} \\ &= \frac{1}{MN}\text{diag}\left\{\mathbf{I}_M \otimes \text{diag}\left\{\sum_{j=1}^{r}\mathbf{A}_N^H\mathbf{w}_j\mathbf{w}_j^H\mathbf{A}_N\right\}\right\} \propto \mathbf{I}_{MN}. \end{aligned} \tag{13}$$

where the equalities $(\mathbf{U}_1 \otimes \mathbf{V}_1)(\mathbf{U}_2 \otimes \mathbf{V}_2) = (\mathbf{U}_1\mathbf{U}_2) \otimes (\mathbf{V}_1\mathbf{V}_2)$ and $\text{diag}\{\mathbf{I} \otimes \mathbf{U}\} = \text{diag}\{\mathbf{I} \otimes \text{diag}\{\mathbf{U}\}\}$ are used. A sufficient condition to ensure (13) can be given by

$$\text{diag}\left\{\mathbf{A}_N^H\mathbf{w}_j\mathbf{w}_j^H\mathbf{A}_N\right\} \propto \mathbf{I}_N, \forall j \in \{1, \ldots, r\}. \tag{14}$$

Combining (11), (14) and the normalization condition $\mathbf{W}^H\mathbf{W} = \mathbf{I}_r$, we can see that $\mathbf{w}_j$ should be a constant-amplitude zero auto-correlation (CAZAC) sequence of length N. A widely used CAZAC sequence is Zadoff-Chu sequence. The ith element of a length N Zadoff-Chu sequence $\mathbf{c}$ can be expressed as

$$c_i = \begin{cases} \sqrt{\frac{1}{N}}\exp\left(j\frac{p\pi(i-1)^2}{N}\right), N \text{ is even}, \\ \sqrt{\frac{1}{N}}\exp\left(j\frac{p\pi i(i-1)}{N}\right), N \text{ is odd}. \end{cases} \tag{15}$$

where p is an integer smaller than and relatively prime to N. Note that the cyclically shifted Zadoff-Chu sequences are orthogonal with each other, that is, $\mathbf{c}^H\mathbf{\Pi}_n\mathbf{c} = 0$ as long as $n \neq 0$, where $\mathbf{\Pi}_n$ is a cyclical-shift matrix of step n

$$\mathbf{\Pi}_n = \begin{bmatrix} \mathbf{0} & \mathbf{I}_n \\ \mathbf{I}_{N-n} & \mathbf{0} \end{bmatrix}. \tag{16}$$

Therefore, the columns of $\mathbf{W}$ can be designed as a set of cyclically shifted Zadoff-Chu sequences with shifting steps $n_1, \ldots, n_r \in \{1, \ldots, N\}$, that is

$$\mathbf{W} = \left[\mathbf{\Pi}_{n_1}\mathbf{c}, \ldots, \mathbf{\Pi}_{n_r}\mathbf{c}\right], n_1 \neq \cdots \neq n_r. \tag{17}$$

IV. Structured SBL Framework for Channel Estimation

By substituting (10) in (5), the received training signal from UE l can be rewritten as

$$\mathbf{y}_l = \sqrt{TP}\mathbf{W}_{\text{full}}^H \mathbf{B}\mathbf{z}_l + \mathbf{n}_l \triangleq \mathbf{\Phi}\mathbf{z}_l + \mathbf{n}_l, \tag{18}$$

where $\mathbf{\Phi} = \sqrt{TP}\mathbf{W}_{\text{full}}^H \mathbf{B}$. We consider the estimation of beam-domain channel $\mathbf{z}_l$. As long as $\mathbf{z}_l$ is estimated, $\mathbf{h}_l$ can be recovered from the basis expansion model.

A. Formulation of Structured SBL

The estimation of $\mathbf{z}_l$ based on (18) falls into the sparse recovery problem. Such problem can be addressed using the conventional SBL [29], [30], where a Gaussian prior for $\mathbf{z}_l$

$$p(\mathbf{z}_l) = \prod_{j=1}^{MN} \mathcal{N}_C\left(z_{l,j} \left| 0, \kappa_{l,j}^{-1}\right.\right) \tag{19}$$

is employed to promote the sparsity in $\mathbf{z}_l$. In (19), $\kappa_{l,j}$ denotes the precision (inverse of variance) of the jth element of $\mathbf{z}_l$. However, such prior fails to capture the structures in the channel sparsity shown in Section III-A. Thus the conventional SBL cannot provide satisfactory estimation performance when the number of RF chains is small, as will be seen in simulations.

To exploit the two kinds of sparse structures, in this work we propose a novel structured prior for $\mathbf{z}_l$, that is

$$\begin{aligned} p(\mathbf{z}_l|\,\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l) &= \mathcal{N}_C\left(\mathbf{z}_l \left| 0, \text{diag}^{-1}\{\boldsymbol{\gamma}_l\} \otimes \mathbf{D}_l^{-1}\right.\right) \\ &= \prod_{m=1}^{M} \mathcal{N}_C\left(\mathbf{z}_{l,m} \left| 0, \gamma_{l,m}^{-1}\mathbf{D}_l^{-1}\right.\right), \end{aligned} \tag{20}$$

where $\boldsymbol{\gamma}_l = [\gamma_{l,1}, \ldots, \gamma_{l,M}]^T$, $\boldsymbol{\alpha}_l = [\alpha_{l,1}, \ldots, \alpha_{l,N}]^T$ and $\mathbf{D}_l$ is a diagonal matrix whose nth diagonal element is given by

$$[\mathbf{D}_l]_{n,n} = \begin{cases} \alpha_{l,N} + \alpha_{l,1} + \alpha_{l,2}, n = 1 \\ \alpha_{l,n-1} + \alpha_{l,n} + \alpha_{l,n+1}, 2 \le n \le N-1 \\ \alpha_{l,N-1} + \alpha_{l,N} + \alpha_{l,1}, n = N \end{cases}. \tag{21}$$

Since the BS cannot obtain enough training samples with limited number of RF chains, the methods such as conventional SBL may suffer from many local optimal solutions with large estimation errors [31]. The proposed structured prior model in (20) and (21) has potential to mitigate this problem.

First, in the structured prior model of (20), the precision matrices of different sub-vectors $\mathbf{z}_{l,m}$ (i.e., $\gamma_{l,m}\mathbf{D}_l$) are different only with a scaling factor $\gamma_{l,m}$. Recall that $\mathbf{z}_{l,m}$ can be viewed as the beam-domain channel between UE l and the mth sub-array of BS in x-direction. Thus the above property just captures the common sparsity structure among sub-arrays. Here the role of $\gamma_{l,m}$ is to model the relative difference between amplitudes of different sub-vectors $\{\mathbf{z}_{l,m}\}_{m=1}^M$. By exploiting the common sparsity structure, the suboptimal solutions in the conventional methods which predict zero for one element of $\mathbf{z}_{l,m}$ and predict nonzero for the element of $\mathbf{z}_{l,m'}$ $(m \neq m')$ in the same position can be eliminated.

Moreover, with (21), the precision of the nth element of $\mathbf{z}_{l,m}$ can be expressed as $\gamma_{l,m}\left(\alpha_{l,n-1} + \alpha_{l,n} + \alpha_{l,n+1}\right)$ (without loss of generality, we consider the case of $2 \le n \le N-1$). In this case, the shared parameter $\alpha_{l,n}$ not only effects the estimation for the nth element of $\mathbf{z}_{l,m}$, but also its neighbors. Such a property can promote the estimation of $\mathbf{z}_{l,m}$ with block sparsity structure because, if the algorithm produces a very large $\alpha_{l,n}$, the estimations of the nth, $(n-1)$th and $(n+1)$th elements of $\mathbf{z}_{l,m}$ will be driven to zero simultaneously. This means that the zero elements (and hence nonzero elements) in the estimation of $\mathbf{z}_{l,m}$ will appear in block, which just captures the block sparsity structure. The solutions without such structure (which may be local optima in the conventional methods) are less likely to be selected in the algorithm.

Given the received signal $\mathbf{y}_l$ in (18), the posterior distribution of $\mathbf{z}_l$ can be expressed as

$$p(\mathbf{z}_l\,|\mathbf{y}_l, \boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l) \propto p(\mathbf{y}_l\,|\mathbf{z}_l)\, p(\mathbf{z}_l|\,\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l), \tag{22}$$

where $p(\mathbf{y}_l\,|\mathbf{z}_l)$ is the likelihood function of $\mathbf{y}_l$, that is

$$p(\mathbf{y}_l\,|\mathbf{z}_l) = \mathcal{N}_{\mathcal{C}}\left(\mathbf{y}_l \left| \mathbf{\Phi}\mathbf{z}_l, \sigma^2\mathbf{I}_R\right.\right). \tag{23}$$

By substituting (20) and (23) into (22), it can be readily to verify that $\mathbf{z}_l$ has Gaussian posterior $\mathcal{N}_C(\mathbf{z}_l\,|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ with mean and covariance matrix as follows

$$\begin{aligned} \boldsymbol{\mu}_l &= \frac{1}{\sigma^2}\boldsymbol{\Sigma}_l \mathbf{\Phi}^H \mathbf{y}_l, \\ \boldsymbol{\Sigma}_l &= \left(\text{diag}\{\boldsymbol{\gamma}_l\} \otimes \mathbf{D}_l + \frac{1}{\sigma^2}\mathbf{\Phi}^H\mathbf{\Phi}\right)^{-1}. \end{aligned} \tag{24}$$

To give a fully Bayesian treatment, similar to conventional SBL, we introduce a Gamma hyperprior for $\boldsymbol{\alpha}_l$

$$\begin{aligned} p(\boldsymbol{\alpha}_l) &= \prod_{n=1}^{N} \text{Gam}(\alpha_{l,n}\,|a, b) \\ &= \prod_{n=1}^{N} \frac{1}{\Gamma(a)} b^a \alpha_{l,n}^a \exp(-b\alpha_{l,n}), \end{aligned} \tag{25}$$

where a and b are fixed hyperparameters. In this paper, we follow the convention in [29], [30] to set a and b as small positive numbers (e.g., $a = b = 10^{-4}$). Such a setup promotes a non-informative (i.e., flat) prior for $\boldsymbol{\alpha}_l$. As discussed in [32, Chapter. 2.3.7], in this case the marginal distribution of $\mathbf{z}_l$, i.e., $p(\mathbf{z}_l|\,\boldsymbol{\gamma}_l) = \int p(\mathbf{z}_l|\,\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l)\, p(\boldsymbol{\alpha}_l)\, d\boldsymbol{\alpha}_l$, has a longer tail than Gaussian distribution, which can make the recovery algorithm more robust to the outliers in the training data.

Moreover, we introduce a Dirichlet hyperprior for $\boldsymbol{\gamma}_l$

$$p(\boldsymbol{\gamma}_l) = C(\mathbf{u}) \prod_{m=1}^{M} \gamma_{l,m}^{u_m}, \tag{26}$$

where $\mathbf{u} = [u_1, \ldots, u_M]^T$ is fixed parameter. $C(\mathbf{u}) = \frac{\Gamma\left(\sum_{m=1}^M u_m\right)}{\Gamma(u_1)\cdots\Gamma(u_M)}$ is the normalization constant. Since the expectation of $\gamma_{l,m}$ with respect to the distribution (26) is given by $\mathbb{E}[\gamma_{l,m}] = \frac{u_m}{\sum_{m'=1}^M u_{m'}}$, we can interpret $\mathbf{u}$ as the parameter which gives a initial guess on the relative difference between the amplitudes of sub-vectors $\{\mathbf{z}_{l,m}\}_{m=1}^M$. Note that a default setup for

$\mathbf{u}$ can be $u_m = 1$ for $m = 1, \ldots, M$ which indicates an initial assumption of equal amplitude for beam-domain channels associated to different sub-arrays. This is a reasonable assumption since $\{\mathbf{z}_{l,m}\}_{m=1}^{M}$ will have similar expected amplitude in far-field propagation scenario. However, we will also compare the performances under other choices of $\mathbf{u}$ in the simulations.

As long as $\boldsymbol{\alpha}_l$ and $\boldsymbol{\gamma}_l$ are obtained, the maximum posterior (MAP) estimation of $\mathbf{z}_l$ can be given by its posterior mean in (24). Therefore, in the following, we focus on finding the optimal $\boldsymbol{\alpha}_l$ and $\boldsymbol{\gamma}_l$ by solving the MAP problem

$$\begin{aligned}\left\{\boldsymbol{\alpha}_l^{\text{opt}}, \boldsymbol{\gamma}_l^{\text{opt}}\right\} &= \arg\max_{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l} p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l | \mathbf{y}_l\right) \\ &= \arg\min_{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l} -\log p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l\right). \end{aligned} \tag{27}$$

Different from the conventional SBL, directly solving the above problem is quite difficult due to employment of structured prior. Although we can build a valid algorithm using gradient descent based approach, the computational complexity will be very high since the gradient of objective function does not have closed-form expression. To address this challenge, we employ a block SMM algorithm to find the solution efficiently.

B. Solving the Structured SBL Using Block SMM Algorithm

The block SMM algorithm is closely related to expectation maximization (EM) algorithm intensively used for conventional SBL [22]. It generalizes EM by replacing the expectation step with a majorization step (MA-step) which finds a surrogate function (local approximation) of objective function. In minimization step (MI-step), the surrogate function is minimized with respect to the parameters. The surrogate function is chosen to ensure closed-form/simple solution is available in MI-step. Different from EM where all parameters are updated simultaneously, the block SMM allows parameters to be divided into several blocks and optimized alternately. This is useful when the minimization of surrogate function with respect to all parameters is still intractable, which is just the problem in solving structured SBL.

Mathematically, let $\mathcal{U}\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right)$ be the surrogate function selected in MA-step, where $\{\hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\}$ denotes arbitrary estimation of $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\}$. According to [33, Theorem 1], if $\mathcal{U}\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right)$ satisfies the following conditions

$$\begin{aligned}&\mathcal{U}\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right) \geq -\log p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l\right), \forall \boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \\ &\mathcal{U}\left(\hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right) = -\log p\left(\hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l, \mathbf{y}_l\right), \\ &\left.\frac{\partial \mathcal{U}\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right)}{\partial \boldsymbol{\alpha}_l}\right|_{\boldsymbol{\alpha}_l = \hat{\boldsymbol{\alpha}}_l} \\ &\qquad = \left.\frac{\partial\left\{-\log p\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l, \mathbf{y}_l\right)\right\}}{\partial \boldsymbol{\alpha}_l}\right|_{\boldsymbol{\alpha}_l = \hat{\boldsymbol{\alpha}}_l}, \\ &\left.\frac{\partial \mathcal{U}\left(\hat{\boldsymbol{\alpha}}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l | \hat{\boldsymbol{\alpha}}_l, \hat{\boldsymbol{\gamma}}_l\right)}{\partial \boldsymbol{\gamma}_l}\right|_{\boldsymbol{\gamma}_l = \hat{\boldsymbol{\gamma}}_l} \\ &\qquad = \left.\frac{\partial\left\{-\log p\left(\hat{\boldsymbol{\alpha}}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l\right)\right\}}{\partial \boldsymbol{\gamma}_l}\right|_{\boldsymbol{\gamma}_l = \hat{\boldsymbol{\gamma}}_l}, \end{aligned} \tag{28}$$

and meanwhile, $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\}$ is updated using the following rule

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_l^{(i+1)} &= \arg\max_{\boldsymbol{\alpha}_l} \mathcal{U}\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l^{(i)}, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \\ \hat{\boldsymbol{\gamma}}_l^{(i+1)} &= \arg\max_{\boldsymbol{\gamma}_l} \mathcal{U}\left(\hat{\boldsymbol{\alpha}}_l^{(i+1)}, \boldsymbol{\gamma}_l, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \end{aligned} \tag{29}$$

where $\left\{\hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right\}$ denotes the estimation of $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\}$ in the ith iteration, then the block SMM algorithm can increase the original target function (27) after each iteration and every limit point of the iterations is a stationary point of (27).

The detailed algorithm is described in the following.

1) Majorization Step: In MA-step of the $(i+1)$th iteration, a surrogate function of objective function $-\log p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l\right)$ is needed. In this work, we utilize the surrogate function suggested by [22] which satisfies the conditions specified in (28), that is

$$\begin{aligned}&\mathcal{U}\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \\ &= -\int p\left(\mathbf{z}_l \middle| \mathbf{y}_l, \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \log \frac{p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{z}_l, \mathbf{y}_l\right)}{p\left(\mathbf{z}_l \middle| \mathbf{y}_l, \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right)} d\mathbf{z}_l \\ &\geq -\log p\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l\right) \end{aligned} \tag{30}$$

where the last step is based on the Jensen's inequality, and the equality holds when $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\} = \left\{\hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right\}$. A short proof for (30) is presented in Appendix A. Note that the upper bound in (30) is just the negative of Q-function in the EM algorithm [34]. By substituting (22) into (30), we can obtain

$$\begin{aligned}\mathcal{U}\left(\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) &= -\log p\left(\boldsymbol{\alpha}_l\right) - \log p\left(\boldsymbol{\gamma}_l\right) \\ &\quad - \mathbb{E}_{\mathbf{z}_l \middle| \mathbf{y}_l, \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}} \\ &\quad \times \left[\log p\left(\mathbf{z}_l | \boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\right)\right] + const. \end{aligned} \tag{31}$$

where the expectation is with respect to $p\left(\mathbf{z}_l \middle| \mathbf{y}_l, \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right)$.

2) Minimization Step: In MI-step of the $(i+1)$th iteration, the parameters $\boldsymbol{\alpha}_l$ and $\boldsymbol{\gamma}_l$ are updated sequentially to minimize the surrogate function.

Update of $\boldsymbol{\alpha}_l$: When updating $\boldsymbol{\alpha}_l$, $\boldsymbol{\gamma}_l$ is fixed at its latest estimation $\boldsymbol{\gamma}_l = \hat{\boldsymbol{\gamma}}_l^{(i)}$. By substituting (20) and (25) into (31) and neglecting the terms irrelevant to $\boldsymbol{\alpha}_l$, we have

$$\begin{aligned}&\mathcal{U}\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l^{(i)}, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \\ &= -\sum_{n=1}^{N} \left(a \log \alpha_{l,n} - b\alpha_{l,n}\right) - \log \left|\text{diag}\left\{\hat{\boldsymbol{\gamma}}_l^{(i)}\right\} \otimes \mathbf{D}_l\right| \\ &\quad + \text{tr}\left(\left(\text{diag}\left\{\hat{\boldsymbol{\gamma}}_l^{(i)}\right\} \otimes \mathbf{D}_l\right)\left(\hat{\boldsymbol{\Sigma}}_l + \hat{\boldsymbol{\mu}}_l \hat{\boldsymbol{\mu}}_l^H\right)\right) + const., \end{aligned} \tag{32}$$

In (32), $\hat{\boldsymbol{\Sigma}}_l$ and $\hat{\boldsymbol{\mu}}_l$ are computed using (24) by replacing $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\}$ with $\left\{\hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right\}$.

To examine the optimality condition, we consider the derivative of (32) with respect to $\alpha_{l,n}$

$$\frac{\partial \mathcal{U}\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l^{(i)}, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right)}{\partial \alpha_{l,n}} = \frac{a}{\alpha_{l,n}} - b - \sum_{m=1}^{M} \hat{\gamma}_{l,m}^{(i)} g_{l,m,n} + M\left(\omega_{l,n-1} + \omega_{l,n} + \omega_{l,n+1}\right), \tag{33}$$

where $\omega_{l,n}$ and $g_{l,m,n}$ are defined as

$$\omega_{l,n} = \frac{1}{\alpha_{l,n-1} + \alpha_{l,n} + \alpha_{l,n+1}}, \tag{34}$$

$$\begin{aligned} g_{l,m,n} = &\left[\hat{\boldsymbol{\Sigma}}_l\right]_{(m-1)N+n-1,(m-1)N+n-1} + \left|\hat{\mu}_{l,(m-1)N+n-1}\right|^2 \\ &+ \left[\hat{\boldsymbol{\Sigma}}_l\right]_{(m-1)N+n,(m-1)N+n} + \left|\hat{\mu}_{l,(m-1)N+n}\right|^2 \\ &+ \left[\hat{\boldsymbol{\Sigma}}_l\right]_{(m-1)N+n+1,(m-1)N+n+1} + \left|\hat{\mu}_{l,(m-1)N+n+1}\right|^2. \end{aligned} \tag{35}$$

Note that the above definition is valid when $2 \le n \le N-1$. When $n = 1$, all subscripts $n-1$ on right-hand sides of (34) and (35) should be replaced with N, and when $n = N$, all subscripts $n+1$ on right-hand sides of (34) and (35) should be replaced with 1.

Unfortunately, we cannot obtain the closed-form solution of $\alpha_{l,n}$ by setting (33) to zero, since $\alpha_{l,n}$ is coupled with the adjacent $\alpha_{l,n+1}$ and $\alpha_{l,n-1}$ in the term $\omega_{l,n}$. To derive a computationally efficient solution, when updating $\alpha_{l,n}$ we temporarily assume that $z_{l,(m-1)N+n}$ has the same variance with the adjacent $z_{l,(m-1)N+n-1}$ and $z_{l,(m-1)N+n+1}$. In this case, we will have $\alpha_{l,n-2} = \alpha_{l,n-1} = \alpha_{l,n} = \alpha_{l,n+1} = \alpha_{l,n+2}$. Under this assumption, (33) can be simplified as

$$\frac{\partial \mathcal{U}\left(\boldsymbol{\alpha}_l, \hat{\boldsymbol{\gamma}}_l^{(i)}, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right)}{\partial \alpha_{l,n}} = \frac{a}{\alpha_{l,n}} - b + \frac{M}{\alpha_{l,n}} - \sum_{m=1}^{M} \hat{\gamma}_{l,m}^{(i)} g_{l,m,n}. \tag{36}$$

By setting (36) to zero, a suboptimal closed-form solution for $\alpha_{l,n}$ can be obtained as

$$\hat{\alpha}_{l,n}^{(i+1)} = \frac{a + M}{\sum_{m=1}^{M} \hat{\gamma}_{l,m}^{(i)} g_{l,m,n} + b}. \tag{37}$$

The above approximate solution gives good estimation performance as will be seen in simulations. Moreover, the solution provides a clear insight on the difference between structured SBL and conventional SBL. In conventional SBL, the update of precision $\kappa_{l,j}$ ($j \in \{1, \ldots, MN\}$) is only related with posterior mean and variance of $z_{l,j}$. While in structured SBL, the update of $\alpha_{l,n}$ ($n \in \{1, \ldots, N\}$) is related with posterior means and variances of $\left\{z_{l,(m-1)N+n-1}, z_{l,(m-1)N+n}, \ z_{l,(m-1)N+n+1}\right\}$ for all $m = 1, \ldots, M$, which is just the consequence of exploiting the common and block sparsity structures. Thus the structured SBL is expected to achieve better performance.

Algorithm 1: Block SMM Algorithm For Structured SBL.

1: **Initialization:**
- Set a tolerance ε and maximum iteration times $I_{\max}$.
- Set $\boldsymbol{\alpha}_l$ and $\boldsymbol{\gamma}_l$ to some initial values $\hat{\boldsymbol{\alpha}}_l^{(0)}$ and $\hat{\boldsymbol{\gamma}}_l^{(0)}$.

2: **MA-step in the $(i+1)$th iteration:**
- Fix $\{\boldsymbol{\alpha}_l, \boldsymbol{\gamma}_l\}$ at $\left\{\hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right\}$. Update the posterior mean and covariance matrix of $\mathbf{z}_l$ using (24).

3: **MI-step in the $(i+1)$th iteration:**
- Fix $\boldsymbol{\gamma}_l$ at $\hat{\boldsymbol{\gamma}}_l^{(i)}$. Update $\boldsymbol{\alpha}_l$ using (37).
- Fix $\boldsymbol{\alpha}_l$ at $\hat{\boldsymbol{\alpha}}_l^{(i+1)}$. Update $\boldsymbol{\gamma}_l$ using (40).

4: **Test of convergence:**
- Let $\hat{\mathbf{z}}_l^{(i+1)}$ denote the estimation of $\mathbf{z}_l$ after the $(i+1)$th iteration. If $\|\hat{\mathbf{z}}_l^{(i+1)} - \hat{\mathbf{z}}_l^{(i)}\|^2 \le \varepsilon$ or $i+1 = I_{\max}$, output $\hat{\mathbf{z}}_l = \hat{\mathbf{z}}_l^{(i+1)}$. Otherwise, set $i = i+1$ and go back to step 2.

Update of $\boldsymbol{\gamma}_l$: By substituting (20), (26) into (31) and neglecting the terms irrelevant to $\boldsymbol{\gamma}_l$, we can obtain

$$\begin{aligned} &\mathcal{U}\left(\hat{\boldsymbol{\alpha}}_l^{(i+1)}, \boldsymbol{\gamma}_l, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right) \\ &= -\sum_{m=1}^{M} u_m \log \boldsymbol{\gamma}_{l,m} - \log\left|\operatorname{diag}\{\boldsymbol{\gamma}_l\} \otimes \hat{\mathbf{D}}_l\right| \\ &\quad + \operatorname{tr}\left(\left(\operatorname{diag}\{\boldsymbol{\gamma}_l\} \otimes \hat{\mathbf{D}}_l\right)\left(\hat{\boldsymbol{\Sigma}}_l + \hat{\boldsymbol{\mu}}_l \hat{\boldsymbol{\mu}}_l^H\right)\right) + const., \end{aligned} \tag{38}$$

where $\boldsymbol{\alpha}_l$ is fixed at its latest estimation $\boldsymbol{\alpha}_l = \hat{\boldsymbol{\alpha}}_l^{(i+1)}$. $\hat{\mathbf{D}}_l$ can be computed using (21) by replacing $\boldsymbol{\alpha}_l$ with $\hat{\boldsymbol{\alpha}}_l^{(i+1)}$. The derivative of (38) with respect to $\gamma_{l,m}$ can be expressed as

$$\begin{aligned} &\frac{\partial \mathcal{U}\left(\hat{\boldsymbol{\alpha}}_l^{(i+1)}, \boldsymbol{\gamma}_l, \mathbf{y}_l \middle| \hat{\boldsymbol{\alpha}}_l^{(i)}, \hat{\boldsymbol{\gamma}}_l^{(i)}\right)}{\partial \gamma_{l,m}} \\ &= \frac{N + u_m}{\gamma_{l,m}} - \sum_{n=1}^{N}\left(\hat{\alpha}_{l,n}^{(i+1)} + \hat{\alpha}_{l,n-1}^{(i+1)} + \hat{\alpha}_{l,n+1}^{(i+1)}\right) \\ &\quad \times \left(\left[\hat{\boldsymbol{\Sigma}}_l\right]_{(m-1)N+n,(m-1)N+n} + \left|\hat{\mu}_{l,(m-1)N+n}\right|^2\right). \end{aligned} \tag{39}$$

By setting (39) to zero, the optimal $\gamma_{l,m}$ can be obtained as

$$\begin{aligned} \hat{\gamma}_{l,m}^{(i+1)} = &\frac{N + u_m}{\sum_{n=1}^{N}\left(\hat{\alpha}_{l,n}^{(i+1)} + \hat{\alpha}_{l,n-1}^{(i+1)} + \hat{\alpha}_{l,n+1}^{(i+1)}\right)} \\ &\times \left([\hat{\boldsymbol{\Sigma}}_l]_{(m-1)N+n,(m-1)N+n} + \left|\hat{\mu}_{l,(m-1)N+n}\right|^2\right)^{-1}. \end{aligned} \tag{40}$$

The MA-step and MI-step are performed iteratively until convergence. A summary of block SMM algorithm for structured SBL is presented in Algorithm 1.

C. Analysis of Computational Complexity

In Algorithm 1, the main computation load in each iteration comes from the calculation of posterior mean and covariance matrix of $\mathbf{z}_l$ in MA-step. During updating $\mathbf{\Sigma}_l$ and $\boldsymbol{\mu}_l$, we need to compute matrix inversion of dimension $MN \times MN$, which involves computational complexity of $\mathcal{O}(M^3N^3)$. Fortunately, by exploiting the special structure of the matrix, the matrix inversion can be computed with complexity $\mathcal{O}(Mr^3)$. To see this, rewrite $\mathbf{\Sigma}_l$ in (24) as follows

$$\begin{aligned}\mathbf{\Sigma}_l &= \left(\operatorname{diag}\{\boldsymbol{\gamma}_l\} \otimes \mathbf{D}_l + \frac{TP}{MN\sigma^2}\mathbf{I}_M \otimes \left(\mathbf{A}_N^H \mathbf{W}\mathbf{W}^H \mathbf{A}_N\right)\right)^{-1} \\ &= \operatorname{diag}\{\mathbf{\Sigma}_{l,1}, \ldots, \mathbf{\Sigma}_{l,M}\}, \end{aligned} \tag{41}$$

where $\mathbf{\Sigma}_{l,m} = \left(\gamma_{l,m}\mathbf{D}_l + \frac{TP}{MN\sigma^2}\mathbf{A}_N^H\mathbf{W}\mathbf{W}^H\mathbf{A}_N\right)^{-1} \in \mathbb{C}^{N\times N}$. The first step follows from the equality $\mathbf{\Phi} = \sqrt{TP}\mathbf{W}_{\text{full}}^H\mathbf{B} = \sqrt{\frac{TP}{MN}}\mathbf{I}_M \otimes \left(\mathbf{W}^H\mathbf{A}_N\right)$. The second step follows from the block diagonal nature of the matrix in bracket. With the Woodbury identity, the computation of $\mathbf{\Sigma}_{l,m}$ can be reduced to $\mathcal{O}(r^3)$. Thus, the overall computational complexity of Algorithm 1 is $\mathcal{O}(IMr^3)$, where I is the number of iterations to converge. Since r is much smaller than N, the computational complexity scales well with the system parameters.

V. Adaptive Spatial Interpolation for UPA With Large M

From Section IV, the number of required RF chains at BS to implement the proposed structured SBL is equal to $R = Mr$. If the number of antennas in y-direction becomes large and comparable to N, direct utilization of the proposed scheme still results in high hardware complexity. Recalling the segmentation of $\mathbf{h}_l$ in (8), we can transform $\mathbf{h}_l$ into the matrix form $\mathbf{H}_l = [\mathbf{h}_{l,1}, \ldots, \mathbf{h}_{l,M}] \in \mathbb{C}^{N\times M}$, where the mth column denotes the channel between UE l and the mth sub-array of BS in x-direction, and the nth row denotes the channel between UE l and the nth sub-array of BS in y-direction. The corresponding matrix form for beam-domain channel is given by $\mathbf{Z}_l = [\mathbf{z}_{l,1}, \ldots, \mathbf{z}_{l,M}] \in \mathbb{C}^{N\times M}$. When both M and N are large, the inefficiency of structured SBL is due to the fact that it exploits only the correlation in column direction of $\mathbf{H}_l$ (or equivalently the sparsity in the column direction of $\mathbf{Z}_l$). When M becomes large and comparable to N, the correlation in the row direction of $\mathbf{H}_l$ should also be utilized in order to design efficient channel estimation scheme. In this section, we propose an adaptive spatial interpolation based channel estimation scheme to achieve this goal.

A. Channel Sparsity When Both M and N Are Large

The reduce the difficulty in channel estimation, the conventional method is to find more sparse representation for $\mathbf{H}_l$ by exploiting the correlations in both column and row directions. According to the discussion in Section III-A, this can be achieved by using projection matrix $\dot{\mathbf{B}} = \sqrt{1/MN}\mathbf{A}_M \otimes \mathbf{A}_N$ whose columns form a set of orthogonal basis of dimension MN. Let $\dot{\mathbf{z}}_l = \dot{\mathbf{B}}^H\mathbf{h}_l$ denote the beam-domain channel of $\mathbf{h}_l$ under basis $\dot{\mathbf{B}}$. Following the analytic methods in [9], [12], [13], when both M and N grows very large, the sparse property of channel is given by the lemma below.

Lemma 2: Let $\dot{\mathbf{b}}_{m,n}$ denote the $((m-1)N+n)$th basis in $\dot{\mathbf{B}}$ (i.e., the $((m-1)N+n)$th column of $\dot{\mathbf{B}}$), which can be expressed as $\dot{\mathbf{b}}_{m,n} = \sqrt{\frac{1}{MN}}\mathbf{a}_M\left(-\frac{1}{2}+\frac{m-1}{M}\right) \otimes \mathbf{a}_N\left(-\frac{1}{2}+\frac{n-1}{N}\right)$. In the large $\{N, M\}$ regime, the projection of $\mathbf{h}_l$ onto $\dot{\mathbf{b}}_{m,n}$ (i.e., the $((m-1)N+n)$th element of $\dot{\mathbf{z}}_l$) has significant amplitude only if $-\frac{1}{2}+\frac{n-1}{N} \in [\rho_{l,x}-\Delta_x, \rho_{l,x}+\Delta_x]$ and $-\frac{1}{2}+\frac{m-1}{M} \in [\rho_{l,y}-\Delta_y, \rho_{l,y}+\Delta_y]$.

Based on Lemma 2, the beam support can be expressed as $\Omega_l = \Omega_l^x \times \Omega_l^y$ with

$$\begin{aligned}\Omega_l^x &= \left\{n \left| -\frac{1}{2}+\frac{n-1}{N} \in [\rho_{l,x}-\Delta_x, \rho_{l,x}+\Delta_x], n \in \mathbb{N}^+\right.\right\}, \\ \Omega_l^y &= \left\{m \left| -\frac{1}{2}+\frac{m-1}{M} \in [\rho_{l,y}-\Delta_y, \rho_{l,y}+\Delta_y], m \in \mathbb{N}^+\right.\right\}. \end{aligned} \tag{42}$$

In this case, the received training signal from UE l can be expressed as

$$\mathbf{y}_l = \sqrt{T_pP}\mathbf{W}_{\text{full}}^H\dot{\mathbf{B}}\dot{\mathbf{z}}_l + \mathbf{n}_l. \tag{43}$$

Unfortunately, there is no obvious structure on the sparsity of $\dot{\mathbf{z}}_l$ which can be exploited to design robust recovery scheme. We can surely utilize conventional SBL or OMP to recover $\dot{\mathbf{z}}_l$ by treating $\sqrt{T_pP}\mathbf{W}_{\text{full}}^H\dot{\mathbf{B}}$ as measurement matrix. However, the performance is poor as will be seen in simulations.

To overcome this challenge, in the following, we propose a novel scheme which first estimates the channel between a few sampled sub-arrays and UE and then recovers the whole channel vector by proposing an adaptive spatial interpolation scheme. The benefit of the scheme is two-fold: 1) First it can exploit the correlations in both column and row directions of $\mathbf{H}_l$ efficiently and thus can reduce the number of required RF chains at BS significantly. 2) Moreover, it retains the two kinds of sparse structures in beam-domain channel. This allows the structured SBL framework to be utilized to ensure reliable channel estimation.

B. Channel Estimation and Beam Support Recovery With Sampled Sub-Arrays

In the proposed scheme, the first step is to estimate the channel between UE and a few sampled sub-arrays of UPA, and then recover the beam support defined in (42).

1) Sampling Along y-Direction: During the training phase, the sub-arrays in x-direction indexed by $\mathcal{S}^y \triangleq \{s_1^y, \ldots, s_{|\mathcal{S}^y|}^y\}$ are selected out. This is equivalent to perform spatial sampling on antennas along y-direction as shown in Fig. 4. The sampled sub-arrays are connected to RF chains using NOSA configuration, that is, each selected sub-array is connected to r RF chains through a separate phase-shift network. The channel between UE l and $|\mathcal{S}^y|$ sampled sub-arrays is given by $\mathbf{h}_l^x = \operatorname{vec}\left(\mathbf{H}_l\left(:, \mathcal{S}^y\right)\right)$. The received training signal from the sampled sub-arrays can be expressed

$$\mathbf{y}_l^x = \sqrt{T_pP}(\mathbf{W}^x)^H\mathbf{B}^x\mathbf{z}_l^x + \mathbf{n}_l^x \triangleq \mathbf{\Phi}^x\mathbf{z}_l^x + \mathbf{n}_l^x. \tag{44}$$

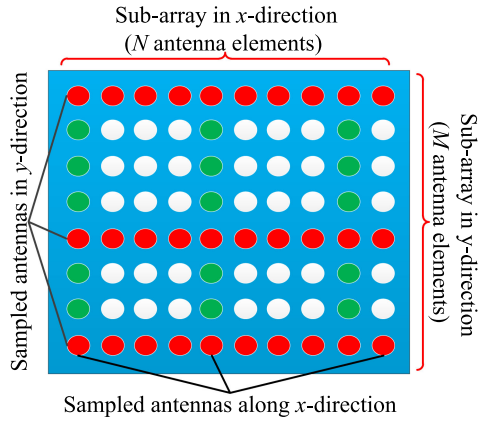


Fig. 4. Illustration of antenna sampling.

In (44), $\mathbf{W}^x = \mathbf{I}_{|\mathcal{S}^y|} \otimes \mathbf{W}$ is the RF receive processing matrix, where $\mathbf{W}$ is defined in Section III-B. $\mathbf{B}^x = \sqrt{\frac{1}{|\mathcal{S}^y|N}}\mathbf{I}_{|\mathcal{S}^y|} \otimes \mathbf{A}_N$ is the orthogonal basis set. $\mathbf{z}_l^x = (\mathbf{B}^x)^H \mathbf{h}_l^x$ is the beam-domain representation of $\mathbf{h}_l^x$ under basis $\mathbf{B}^x$ and $\mathbf{n}_l^x$ is the additive noise vector. $\mathbf{\Phi}^x = \sqrt{T_p P}(\mathbf{W}^x)^H \mathbf{B}^x$ is the measurement matrix. Based on Lemma 1, it can be seen that $\mathbf{z}_l^x$ exhibits the two kinds of sparse structures discussed in III-A, and $|z^x_{l,(m-1)N+n}|$ has significant value only when $n \in \Omega_l^x$. Therefore, a reliable estimation of $\mathbf{z}_l^x$ can be obtained with the structured SBL in Section IV. Moreover, the index set Ω_l^x can be estimated by solving the problem as follows

$$\min |\Omega_l^x| \ \text{s.t.} \frac{1}{\|\hat{\mathbf{z}}_l^x\|^2} \sum_{n\in\Omega_l^x} \sum_{m=1}^{|\mathcal{S}^y|} \left|\hat{z}^x_{l,(m-1)N+n}\right|^2 \geq \eta, \quad (45)$$

where η denotes the threshold close to 1, and $\hat{\mathbf{z}}_l^x$ denotes the estimation of $\mathbf{z}_l^x$. Although the problem (45) is combinatorial, it can be simply solved by adding the index n with largest beam-domain channel power $\sum_{m=1}^{|\mathcal{S}^y|} |\hat{z}^x_{l,(m-1)N+n}|^2$ into Ω_l^x one by one until the constraint in (45) is satisfied.

2) Sampling Along x-Direction: Similarly, the sub-arrays in y-direction indexed by $\mathcal{S}^x \triangleq \{s_1^x, \ldots, s_{|\mathcal{S}^x|}^x\}$ are selected out, and each selected sub-array is connected to r RF chains through a separate phase-shift network. This is equivalent to spatial antenna sampling along x-direction as shown in Fig. 4. The channel between UE l and the sampled sub-arrays is given by $\mathbf{h}_l^y = \text{vec}(\mathbf{H}_l(\mathcal{S}^x, :)^T)$. The received training signal from the sampled sub-arrays can be expressed as

$$\mathbf{y}_l^y = \sqrt{T_p P}(\mathbf{W}^y)^H \mathbf{B}^y \mathbf{z}_l^y + \mathbf{n}_l^y \triangleq \mathbf{\Phi}^y \mathbf{z}_l^y + \mathbf{n}_l^y, \quad (46)$$

where $\mathbf{W}^y, \mathbf{B}^y, \mathbf{z}_l^y, \mathbf{n}_l^y$ and $\mathbf{\Phi}^y$ are defined similar to that in (44). With Lemma 1, it is seen that $\mathbf{z}_l^y$ exhibits the two kinds of sparse structures, and $|z^y_{l,(n-1)M+m}|$ has significant value only when $m \in \Omega_l^y$. Therefore, a reliable estimation of $\mathbf{z}_l^y$ can be obtained using structured SBL. The index set Ω_l^y can be obtained by solving the following problem

$$\min |\Omega_l^y| \ \text{s.t.} \frac{1}{\|\hat{\mathbf{z}}_l^y\|^2} \sum_{m\in\Omega_l^y} \sum_{n=1}^{|\mathcal{S}^x|} \left|\hat{z}^y_{l,(n-1)M+m}\right|^2 \geq \eta, \quad (47)$$

where $\hat{\mathbf{z}}_l^y$ denotes the estimation of $\mathbf{z}_l^y$.

Combining (45) and (47), the beam support can be obtained as $\Omega_l = \Omega_l^x \times \Omega_l^y$. To implement with low hardware complexity, the numbers of sampled antennas along x- and y-direction are made much smaller than N and M, that is, $|\mathcal{S}^x| \ll N$ and $|\mathcal{S}^y| \ll M$. The total number of required RF chains is equal to $(|\mathcal{S}^x| + |\mathcal{S}^y|)r$, which is much smaller than MN.

During the channel estimation phase, it is seen that only a small fraction of BS antennas are connected to RF chains. While in the data transmission phase, to fully exploit the potential of large scale antenna array, it is better to connect all antennas to RF chains through NOSA [28]. This can be simply achieved by RF switching technology [35] which has been widely utilized for antenna selection in MIMO systems. It is important to note that this *does not* mean more RF chains are needed in data transmission phase. The reason is that as the channel between UE and BS is estimated, the transmit/receive signal can be processed in the beam-domain with very low effective dimension (See [12], [13] for more details). Therefore, the number of required RF chains is small.

C. Adaptive Spatial Interpolation

Using the basis expansion model $\mathbf{h}_l = \dot{\mathbf{B}}\dot{\mathbf{z}}_l = \sqrt{\frac{1}{MN}}(\mathbf{A}_M \otimes \mathbf{A}_N)\dot{\mathbf{z}}_l$, we can rewrite $\mathbf{H}_l$ as

$$\begin{aligned}\mathbf{H}_l &= \text{vec}_{N,M}^{-1}\left\{\sqrt{\frac{1}{MN}}(\mathbf{A}_M \otimes \mathbf{A}_N)\dot{\mathbf{z}}_l\right\} \\ &\approx \sqrt{\frac{1}{MN}}\text{vec}_{N,M}^{-1}\left\{(\mathbf{A}_M(:,\Omega_l^y,) \otimes \mathbf{A}_N(:,\Omega_l^x))\ddot{\mathbf{z}}_l\right\} \\ &= \sqrt{\frac{1}{MN}}\mathbf{A}_N(:,\Omega_l^x,)\,\text{vec}^{-1}_{|\Omega_l^x|,|\Omega_l^y|}\{\ddot{\mathbf{z}}_l\}\,\mathbf{A}_M(:,\Omega_l^y)^T,\end{aligned} \quad (48)$$

where $\ddot{\mathbf{z}}_l$ is obtained by removing all elements $\{\dot{z}_{l,(m-1)N+n}\}_{\{m,n\}\notin\Omega_l}$ in $\dot{\mathbf{z}}_l$. The second step follows from Lemma 2. The last step follows from the equality $\text{vec}(\mathbf{U}_1\mathbf{V}\mathbf{U}_2^T) = (\mathbf{U}_2 \otimes \mathbf{U}_1)\,\text{vec}(\mathbf{V})$.

As long as $\text{vec}^{-1}_{|\Omega_l^x|,|\Omega_l^y|}\{\ddot{\mathbf{z}}_l\}$ is known, the channel $\mathbf{h}_l$ can be recovered from (48) approximately. However, as discussed in Section V-A, direct estimation of $\dot{\mathbf{z}}_l$ (and hence $\ddot{\mathbf{z}}_l$) is difficult. To address this problem, we propose a channel estimation scheme based on adaptive spatial interpolation using the estimations of $\mathbf{z}_l^x$, $\mathbf{z}_l^y$ and Ω_l obtained in the last subsection.

Defining $\ddot{\mathbf{Z}}_l = \text{vec}^{-1}_{|\Omega_l^x|,|\Omega_l^y|}\{\ddot{\mathbf{z}}_l\}$ and using a similar procedure as that in (48), we can obtain

$$\mathbf{H}_l(:,\mathcal{S}^y) = \sqrt{\frac{1}{MN}}\mathbf{A}_N(:,\Omega_l^x)\,\ddot{\mathbf{Z}}_l\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T. \quad (49)$$

Moreover, based on the basis expansion model, we have $\mathbf{H}_l(:,\mathcal{S}^y) = \text{vec}^{-1}_{N,|\mathcal{S}^y|}(\mathbf{B}^x\mathbf{z}_l^x)$. Therefore, with $\hat{\mathbf{z}}_l^x$, an estimation of $\ddot{\mathbf{Z}}_l$ can be obtained as

$$\ddot{\mathbf{Z}}_l = \sqrt{MN}\mathbf{A}_N(:,\Omega_l^x)^H \mathbf{H}_l(:,\mathcal{S}^y)\left(\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\right)^\dagger = \sqrt{MN}\mathbf{A}_N(:,\Omega_l^x)^H \mathrm{vec}^{-1}_{N,|\mathcal{S}^y|}(\mathbf{B}^x\hat{\mathbf{z}}_l^x)\left(\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\right)^\dagger. \quad (50)$$

By substituting (50) into (48), $\mathbf{H}_l$ can be estimated as

$$\hat{\mathbf{H}}_l^{(1)} = \mathbf{A}_N(:,\Omega_l^x,)\,\mathbf{A}_N(:,\Omega_l^x)^H \mathrm{vec}^{-1}_{N,|\mathcal{S}^y|}(\mathbf{B}^x\hat{\mathbf{z}}_l^x) \times \left(\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\right)^\dagger \mathbf{A}_M(:,\Omega_l^y)^T, \quad (51)$$

or equivalently, an estimation of $\mathbf{h}_l$ can be obtained as

$$\hat{\mathbf{h}}_l^{(1)} = \mathrm{vec}\left\{\hat{\mathbf{H}}_l^{(1)}\right\} = \mathbf{T}^{(1)}\hat{\mathbf{h}}_l^{(x)}, \quad (52)$$

where $\hat{\mathbf{h}}_l^{(x)} = \mathbf{B}^x\hat{\mathbf{z}}_l^x$ denotes the estimation of channel between UE l and sampled sub-arrays, and $\mathbf{T}^{(1)} = ((\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T)^\dagger \mathbf{A}_M(:,\Omega_l^y)^T)^T \otimes (\mathbf{A}_N(:,\Omega_l^x,)\,\mathbf{A}_N(:,\Omega_l^x)^H)$ is the interpolation matrix.

To further improve the estimation quality, we can obtain another copy of estimation for $\mathbf{h}_l$ through spatial interpolation on $\hat{\mathbf{h}}_l^y = \mathbf{B}^y\hat{\mathbf{z}}_l^y$, that is

$$\hat{\mathbf{h}}_l^{(2)} = \mathbf{T}^{(2)}\mathrm{vec}\left\{\left(\mathrm{vec}^{-1}_{M,|\mathcal{S}^x|}\left\{\hat{\mathbf{h}}_l^{(y)}\right\}\right)^T\right\}, \quad (53)$$

where $\mathbf{T}^{(2)} = (\mathbf{A}_M(:,\Omega_l^y)\mathbf{A}_M(:,\Omega_l^y)^H) \otimes (\mathbf{A}_N(:,\Omega_l^x)(\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T)^{\dagger T})$ is the interpolation matrix. The detailed derivation is similar to (52), thus it is omitted.

Finally, the estimation of $\mathbf{h}_l$ can be refined by combining (52) and (53), that is

$$\hat{\mathbf{h}}_l = \frac{1}{2}\left(\hat{\mathbf{h}}_l^{(1)} + \hat{\mathbf{h}}_l^{(2)}\right). \quad (54)$$

Note that in (52) and (53), to ensure the existence of $(\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T)^\dagger$ and $(\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T)^\dagger$, Ω_l^x and Ω_l^y must satisfy $|\Omega_l^x| \le |\mathcal{S}^x|$ and $|\Omega_l^y| \le |\mathcal{S}^x|$. This indicates that when the channel becomes less sparse (i.e., $|\Omega_l^x|$ and $|\Omega_l^y|$ increase), we should give a dense sampling for the antennas along x- and y-direction in order to recover $\mathbf{h}_l$ reliably.

However, the knowledge of Ω_l^x and Ω_l^y is not available before we decide the set of sampled antennas and then estimate the beam support with the method of the last subsection. A more significant problem is that, even $|\Omega_l^x| \le |\mathcal{S}^x|$ and $|\Omega_l^y| \le |\mathcal{S}^y|$ are satisfied, the matrices $\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)$ and $\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)$ may be ill-conditioned, that is, have large condition numbers. This will result in significant noise amplification in spatial interpolation when multiplying $(\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T)^\dagger$ and $(\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T)^\dagger$, which degrades the estimation performance greatly.

To address these problems, we propose an adaptive scheme to adjust the estimated beam support, which can make the spatial interpolation always give stable recovery performance. The strategy is as follows. After determining Ω_l^x using (45), $|\Omega_l^x|$ is compared with $|\mathcal{S}^x|$. 1) If the condition $|\Omega_l^x| \le |\mathcal{S}^x|$ is violated, Ω_l^x is updated by removing the $|\Omega_l^x| - |\mathcal{S}^x|$ indexes $\{n\}$ with smallest beam-domain channel power $\sum_{m=1}^{|\mathcal{S}^y|}|\hat{z}^x_{l,(m-1)N+n}|^2$. 2) Based on the updated Ω_l^x, the condition number of matrix $\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)$ is computed and compared with a predefined threshold $C_{\max}$. If $\mathrm{cond}\{\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)\} \le C_{\max}$, output Ω_l^x as the final estimation. Otherwise, update Ω_l^x by removing the index n with smallest beam-domain channel power $\sum_{m=1}^{|\mathcal{S}^y|}|\hat{z}^x_{l,(m-1)N+n}|^2$. Step 2) is repeated until the condition number of $\mathbf{A}_N(\mathcal{S}^x,\Omega_{l,x})^T\mathbf{A}_N(\mathcal{S}^x,\Omega_{l,x})$ is smaller or equal to $C_{\max}$. Ω_l^y can be adjusted using the same approach.

We name the spatial interpolation with adaptive beam support adjusting as adaptive spatial interpolation scheme which is summarized in Algorithm 2. The alogrithm will always converge if $C_{\max} \ge 1$ because $\mathrm{cond}\{\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)\}$ and $\mathrm{cond}\{\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)\}$ have an achievable lower bound equal to 1.

Algorithm 2: Adaptive Spatial Interpolation Scheme.

1: **Initialization:**
- Set a threshold $C_{\max} \ge 1$.
- Select $\mathcal{S}^x$ and $\mathcal{S}^y$ based on the available RF chains.
- Compute the initial estimations of Ω_l^x and Ω_l^y using (45) and (47).

2: **Adaptive adjusting of Ω_l^x:**
- If $|\Omega_l^x| > |\mathcal{S}^x|$, remove the $|\Omega_l^x| - |\mathcal{S}^x|$ indexes in Ω_l^x with smallest $\sum_{m=1}^{|\mathcal{S}^y|}|\hat{z}^x_{l,(m-1)N+n}|^2$.
- If $\mathrm{cond}\{\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)^T\mathbf{A}_N(\mathcal{S}^x,\Omega_l^x)\} > C_{\max}$, remove the index in Ω_l^x with smallest $\sum_{m=1}^{|\mathcal{S}^y|}|\hat{z}^x_{l,(m-1)N+n}|^2$ and repeat this step. Otherwise, output Ω_l^x.

3: **Adaptive adjusting of Ω_l^y:**
- If $|\Omega_l^y| > |\mathcal{S}^y|$, remove the $|\Omega_l^y| - |\mathcal{S}^y|$ indexes in Ω_l^y with smallest $\sum_{n=1}^{|\mathcal{S}^x|}|\hat{z}^y_{l,(n-1)M+m}|^2$.
- If $\mathrm{cond}\{\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)^T\mathbf{A}_M(\mathcal{S}^y,\Omega_l^y)\} > C_{\max}$, remove the index in Ω_l^y with smallest $\sum_{n=1}^{|\mathcal{S}^x|}|\hat{z}^y_{l,(n-1)M+m}|^2$ and repeat this step. Otherwise, output Ω_l^y.

4: **Perform spatial interpolation and combining using (52), (53) and (54).**

D. Discussion on Channel Estimation With Nonorthogonal Pilot Allocation

Note that the channel estimation schemes in Section IV and Section V are established on the assumption of orthogonal pilot allocation. For nonorthogonal pilot allocation (i.e., $T_p < L$), the channel estimation problems for different UEs cannot be decoupled any more. In this case, a possible solution is to estimate the channels of all UEs jointly. In particular, vectorizing the received training matrix given by (3), we can obtain

$$\tilde{\mathbf{y}} = \mathrm{vec}\left(\tilde{\mathbf{Y}}\right) = \left(\mathbf{P}\otimes\mathbf{W}_{\mathrm{full}}^H\right)\mathbf{h} + \mathrm{vec}\left(\tilde{\mathbf{N}}\right) \in \mathbb{C}^{RT_p\times 1} \quad (55)$$

where $\mathbf{h} = \mathrm{vec}(\mathbf{H}) = \left[\mathbf{h}_1^T,\ldots,\mathbf{h}_L^T\right]^T \in \mathbb{C}^{MNL\times 1}$ is the concatenation of all UEs' channel vectors. In (55), we encounter

a more complex channel estimation problem with very high dimension. In order to estimate the channel reliably, innovative solutions must be developed from the aspects of sparse and structured channel representation methods, efficient estimation algorithms and etc. Although this problem is interesting and meaningful (in term of reducing pilot resource consumption), we will assume orthogonal pilot allocation in this work and leave the nonorthogonal case as future research.

VI. Simulation Results and Discussion

This section presents the simulation results to validate the effectiveness of proposed scheme. For convenience, the large-scale fading and noise variance are normalized to 1. The number of UEs is set to $L = 10$. Note that since UEs' pilot sequences are mutually orthogonal, the estimation performance is independent of L. Since the orthogonal pilot allocation requires $T_p \geq L$, we use the setup of $T_p = L$ to minimize the pilot resource consumption. Without loss of generality, we let $\rho_{l,x} = \frac{d_x}{\lambda} \cos\theta_l \sin\varphi_l$ and $\rho_{l,y} = \frac{d_y}{\lambda} \cos\theta_l \cos\varphi_l$ which corresponds to the setup of hybrid air-terrestrial communication systems. The vertical AoA θ_l and horizontal AoA φ_l are generated randomly from $[20°, 160°]$ and $[0°, 360°]$, respectively. The complex response gains $r\,(\rho_x, \rho_y)$ for different $\{\rho_x, \rho_y\}$ are uncorrelated and follow complex Gaussian distribution with zero-mean [4]. The variance of $r\,(\rho_x, \rho_y)$ for given sample of $\{\rho_x, \rho_y\}$ can be calculated using the method in [9, Sec. VI]. Unless otherwise specified, the hyperparamaters in (25) and (26) will be set as $a = b = 10^{-4}$ and $u_m = 1$ for $m = 1, \ldots, M$.

The performance metric considered is the normalized MSE, which is defined as $\text{MSE} = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{\|\mathbf{h}_l\|^2} \|\hat{\mathbf{h}}_l - \mathbf{h}_l\|^2$. All figures are obtained by averaging the results over 10^3 independent channel realizations.

To show the superiority of proposed scheme, the following reference schemes are considered.

- Conventional SBL [29]: $\mathbf{z}_l$ is given a Gaussian prior of form (19). The conventional SBL problem is solved using the relevance vector machine algorithm in [30].
- OMP [17]: The standard OMP algorithm that does not exploit any structure of sparsity.
- Reweighted block l_1/l_2 minimization [19]: Considering the common sparsity among sub-arrays and UE shared sparsity, it recovers $\mathbf{z}_l$ by building a reweighted mixed l_1/l_2 minimization problem. The problem is solved using the alternating direction approach [36].
- Variational SBL [20]: $\mathbf{z}_l$ is given a Gaussian mixture prior to capture the common sparsity and UE shared sparsity. The estimation of $\mathbf{z}_l$ is formulated as variational SBL.
- Genie-aided LS [10], [13]: It assumes that the beam support is perfectly known at BS, and the beam-domain channel can be estimated using the standard LS approach.

A. Case of Small M

Fig. 5 simulates the normalized MSE performance for different numbers of RF chains per sub-array. It is seen that the conventional SBL and OMP give poor performance when the number of RF chains is small, since they cannot utilize the structures inherent in channel sparsity. In contrast, by exploiting the common sparsity structure and UE shared sparsity, the reweighted block l_1/l_2 minimization and variational SBL can provide better performance. When the number of RF chains increases, the performance of proposed structured SBL approaches to the genie-aided LS. Moreover, the proposed scheme achieves the best performance among all schemes which do not rely on the prior of beam support. The reason is that, despite of the common sparsity structure, the structured SBL also utilizes the block sparsity structure per sub-array.

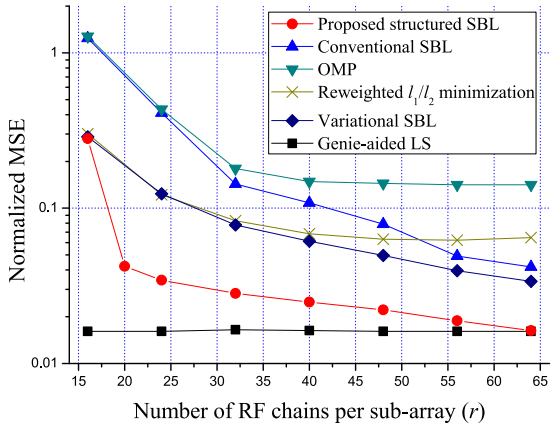


Fig. 5. Normalized MSE versus the number of RF chains per sub-array, where $N = 128$ and $M = 4$. The transmit SNR of UE is $P = 5$ dB. The Rician factor is $K_l = -30$ dB. The angular spreads in vertical and horizontal directions are 3°.

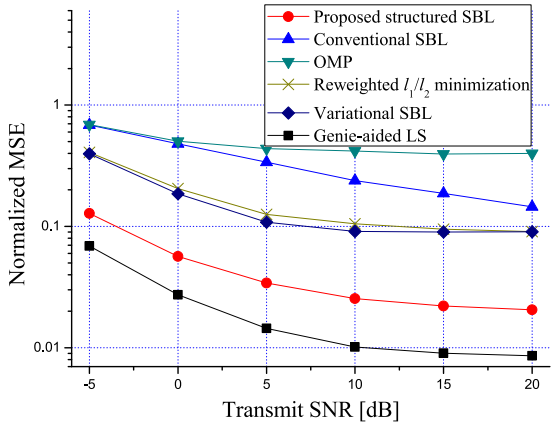


Fig. 6. Normalized MSE performance versus transmit SNR, where $N = 128$ and $M = 4$. The number of RF chains per sub-array $r = 24$. The Rician factor is $K_l = -30$ dB. The angular spreads in vertical and horizontal directions are 3°.

Fig. 6 shows the normalized MSE performance for different transmit SNRs. Again, the proposed scheme achieves the best performance among the schemes which do not rely on the prior of beam support. It is seen that the performances of all schemes converge to MSE floors for large SNR. The reason is that the beam-domain channel is approximately sparse as discussed in Section III-A. That is, the nth element of $\mathbf{z}_{l,m}$ with $n \notin \Omega_l^x$ is small but not zero. With limited number of RF chains, it is generally difficult to recover all these small elements of $\mathbf{z}_{l,m}$ precisely. As a result, with the increasing of the SNR, the effect of the mismatch between the true values of these small elements and their

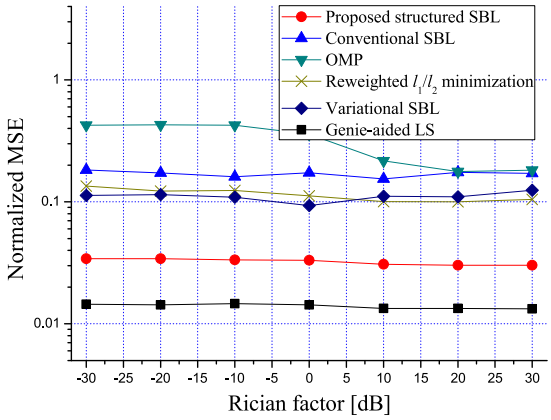


Fig. 7. Normalized MSE performance versus Rician factor K_l, where $N = 128$ and $M = 4$. The transmit SNR is $P = 5$ dB. The number of RF chains per sub-array $r = 24$. The angular spreads in vertical and horizontal directions are 3°.

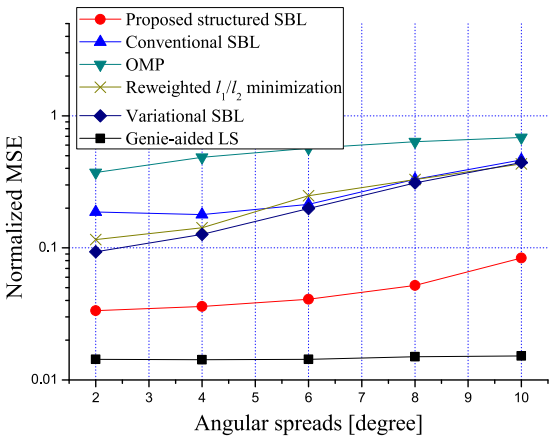


Fig. 8. Normalized MSE versus angular spreads in vertical and horizontal directions, where $N = 128$ and $M = 4$. The transmit SNR of UE is set to $P = 5$ dB. The number of RF chains per sub-array $r = 24$. The Rician factor is $K_l = -30$ dB.

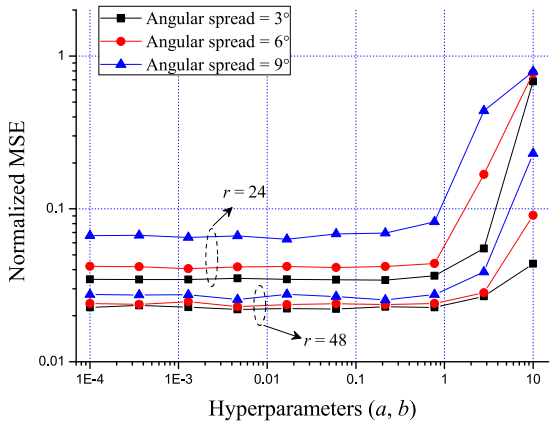


Fig. 9. Normalized MSE of structured SBL versus different a and b, where $N = 128$ and $M = 4$ and $a = b$. The transmit SNR of UE is set to $P = 5$ dB. The Rician factor is $K_l = -30$ dB.

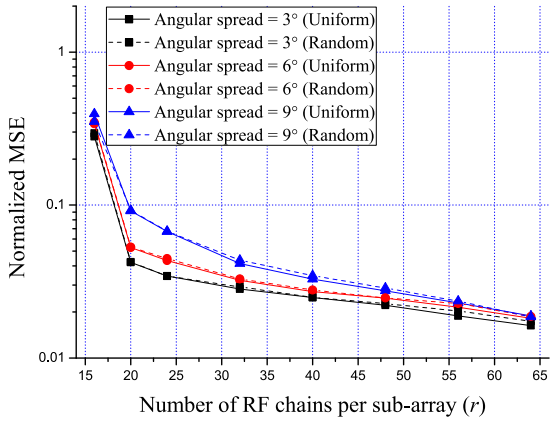


Fig. 10. Comparison of normalized MSE performances under uniform and random setups for $\{u_m\}_{m=1}^M$, where $N = 128$ and $M = 4$. The transmit SNR of UE is set to $P = 5$ dB. The Rician factor is $K_l = -30$ dB.

estimates (which we called modeling error) becomes dominant over the noise, and finally causes performance floor in the high SNR region. We mention that the floor effect cannot be observed from the simulations of [18]–[20] since these works assume that there is no channel power leakage outside Ω_l^x. Although this assumption simplifies the model, it makes the simulations fail to capture the true performance in high SNR region.

Fig. 7 shows the normalized MSE performance for different Rician factors K_l. For convenience, we assume that the Rician factors for all UEs are the same and increase from -30 dB to 30 dB. This corresponds to the situations vary from NLoS dominant scenario to LoS dominant scenario. It is seen that the proposed scheme provides stable estimation performance for both NLoS dominant and LoS dominant scenarios. However, the OMP is sensitive to the strong NLoS channel.

Fig. 8 simulates the normalized MSE performance with different angular spreads in vertical and horizontal directions. As shown in the figure, the performances of all schemes degrade as the angular spreads increase as expected. This is because, with the increase of angular spreads, the cardinality of beam support (that is, the sparsity of channel) becomes larger. According to the sparse recovery theory, more RF chains are needed in order to maintain the same estimation performance.

Then we consider the estimation performance of structured SBL for different hyperparamters in (25) and (26). Fig. 9 shows the effect of a and b on the normalized MSE performance, where we let a and b vary from 10^{-4} to 10. It is seen that, for the selected system parameters, the structured SBL performs well when a and b are smaller than 0.1. This coincides with the analysis in Section IV-A. Fig. 10 compares the normalized MSE performances under uniform and random setups for $\{u_m\}_{m=1}^M$, where $\{u_m\}_{m=1}^M$ are all set to 1 in uniform setup and $\{u_m\}_{m=1}^M$ are randomly generated with constraints $u_m > 0$ and $\sum_{m=1}^M u_m = M$. It is seen that the uniform setup gives slightly better normalized MSE than the random one. Generally, the performance is not quite sensitive to the initial value of $\{u_m\}_{m=1}^M$.

B. *Case of Large M*

This subsection presents the simulations for large M where the UPA with $M = N = 64$ is considered. The threshold η

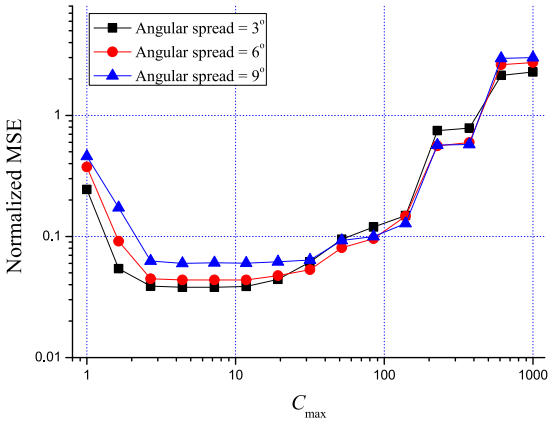


Fig. 11. Normalized MSE performance versus $C_{\max}$. The transmit SNR of UE is $P = 10$ dB. The number of RF chains per sub-array is $r = 16$. The number of sampled antennas is $|\mathcal{S}^x| = |\mathcal{S}^y| = 12$. The Rician factor is $K_l = -30$ dB.

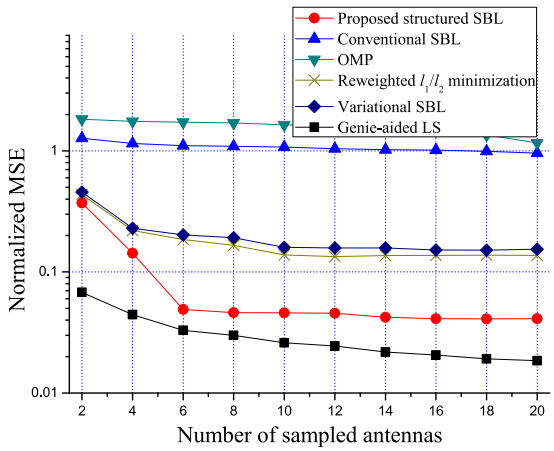


Fig. 12. Normalized MSE performance versus the number of sampled antennas, where $|\mathcal{S}^x| = |\mathcal{S}^y|$. The transmit SNR is $P = 10$ dB. The number of RF chains per sub-array is $r = 12$. The Rician factor is $K_l = -30$ dB. The angular spreads in vertical and horizontal directions are 3°.

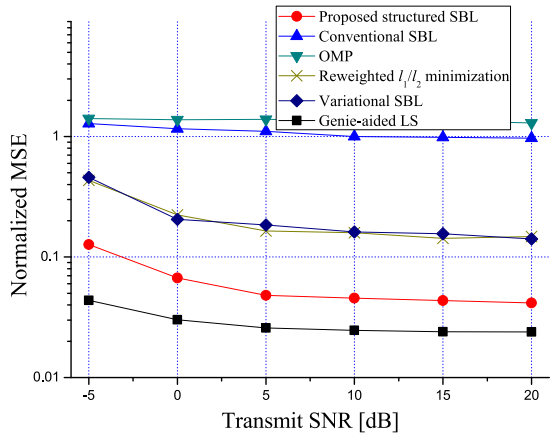


Fig. 13. Normalized MSE performance versus transmit SNR. The number of RF chains per sub-array is $r = 12$. $|\mathcal{S}^x| = |\mathcal{S}^y| = 12$. The Rician factor is $K_l = -30$ dB. The angular spreads in vertical and horizontal directions are 3°.

for initial beam support estimation is set to $\eta = 0.95$. The indexes in $\mathcal{S}^x$ and $\mathcal{S}^y$ are uniformly sampled from $\{1, \ldots, N\}$ and $\{1, \ldots, M\}$, respectively.

Fig. 11 shows the effect of threshold $C_{\max}$ on adaptive spatial interpolation scheme, where we let $C_{\max}$ vary from 1 to 10^3. It is seen that the estimation performance is poor when $C_{\max}$ is too small or too large. A vary small $C_{\max}$ poses strict constraint on the condition numbers of $\mathbf{A}_N(\mathcal{S}^x, \Omega_l^x)^T \mathbf{A}_N(\mathcal{S}^x, \Omega_l^x)$ and $\mathbf{A}_M(\mathcal{S}^y, \Omega_l^y)^T \mathbf{A}_M(\mathcal{S}^y, \Omega_l^y)$. In this case, the adaptive scheme will kick out most of the beams in Ω_l, which results in large modeling error. On the contrary, a large $C_{\max}$ allows matrices $\mathbf{A}_N(\mathcal{S}^x, \Omega_l^x)^T \mathbf{A}_N(\mathcal{S}^x, \Omega_l^x)$ and $\mathbf{A}_M(\mathcal{S}^y, \Omega_l^y)^T \mathbf{A}_M(\mathcal{S}^y, \Omega_l^y)$ to have very large condition numbers, which may make the problem ill-conditioned and result in severe noise amplification. This figure also demonstrates the necessity for the adaptive adjustment of the estimated beam support. Note that although different $C_{\max}$ produce great difference on the estimation performance, the selection of $C_{\max}$ is not difficult in the practical application because the adaptive spatial interpolation scheme provides a good performance for a broad range of $C_{\max}$ (from 2.5 to 13).

Then we compare the proposed scheme with reference schemes. To implement reweighted block l_1/l_2 minimization and variational SBL based schemes in the considered scenario, we modify them to first estimate the beam support using the antenna sampling based approach in Section V-B, and then perform adaptive spatial interpolation algorithm in Section V-C to obtain the complete channel estimation. The conventional SBL and OMP are implemented directly based on the signal model (43), where the total number of RF chains is set to $R = (|\mathcal{S}^x| + |\mathcal{S}^y|)r$ which is the same with other schemes.

Fig. 12 simulates the normalized MSE performance for different numbers of sampled antennas, where we let $|\mathcal{S}^x| = |\mathcal{S}^y|$ and vary from 2 to 20. As expected, the normalized MSE is improved by increasing the number of sampled antennas for all schemes. Moreover, the performance of proposed scheme saturates to a lower bound when $|\mathcal{S}^x|$ and $|\mathcal{S}^y|$ exceed 12. Therefore, $|\mathcal{S}^x| = |\mathcal{S}^y| = 12$ is a suitable operation point which gives a good tradeoff between performance and hardware complexity under the considered simulation parameters.

Fig. 13 simulates the normalized MSE performance for different transmit SNRs. Following the result of Fig. 12, we set the number of sampled antennas as $|\mathcal{S}^x| = |\mathcal{S}^y| = 12$. In this case, the ratio between the number of RF chains and the number of BS antennas is $\frac{(|\mathcal{S}^x|+|\mathcal{S}^y|)r}{MN} \approx 0.07$, which is very small. Again, the proposed scheme achieves the best performance among the schemes which do not rely on the prior of beam support. Moreover, Fig. 12 and Fig. 13 show that directly solving the sparse recovery problem (43) using OMP and conventional SBL fails to provide a reasonable estimation.

Fig. 14 shows the normalized MSE performance for different numbers of iterations. It is observed that the number of iterations required for the proposed scheme is greater than that of the conventional SBL and variational SBL, and less than that of the reweighted block l_1/l_2 minimization. Moreover, the computational complexities in each iteration for different schemes are shown in Table I. The OMP is not considered in the comparison since it consists of fix number of operations with complexity

TABLE I
COMPARISON OF COMPUTATION COMPLEXITIES IN EACH ITERATION FOR DIFFERENT SCHEMES

Scheme	Structured SBL	Conventional SBL	Variational SBL	Reweighted l_1/l_2 minimization
Computational complexity in each iteration	$\mathcal{O}\left(\left\|\mathcal{S}^x\right\| r^3+\left\|\mathcal{S}^y\right\| r^3\right)$	$\mathcal{O}\left(\left(\left\|\mathcal{S}^x\right\|+\left\|\mathcal{S}^y\right\|\right)^3 r^3\right)$ [29]	$\mathcal{O}\left(\left\|\mathcal{S}^x\right\| r^3+\left\|\mathcal{S}^y\right\| r^3\right)$ [20]	$\mathcal{O}\left(\left\|\mathcal{S}^y\right\| Nr+\left\|\mathcal{S}^x\right\| Mr\right)$ [36]

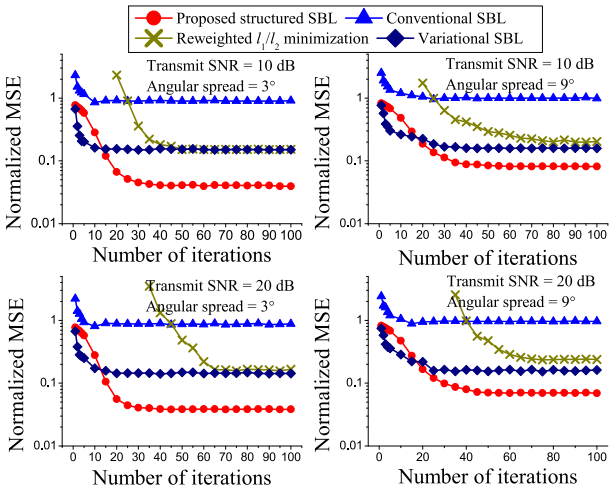


Fig. 14. Normalized MSE performance versus the number of iterations. The number of RF chains per sub-array is $r = 12$. $|\mathcal{S}^x| = |\mathcal{S}^y| = 12$. The Rician factor is $K_l = -30$ dB.

$\mathcal{O}\left(\delta MN\left(|\mathcal{S}^x|+|\mathcal{S}^y|\right)r\right)$ [37], where δ denotes the number of significant elements in the beam-domain channel vector defined in Section V-A. From Fig. 14 and Table I, we can conclude that the overall computational complexity of the proposed scheme is higher than OMP, variational SBL and reweighted block l_1/l_2 minimization, and lower than conventional SBL. Note that, if we just expect some performance improvement over the reference schemes, the computational complexity of the proposed scheme can still be reduced. For example, in the upper left subfigure of Fig. 14, the proposed scheme has already achieved a considerable improvement if we run only 20 iterations (which is the number required for the convergence of variational SBL).

VII. CONCLUSION

The channel estimation for 3D massive MIMO systems with limited RF chains is investigated. By exploiting the sparse structures in the beam-domain channel, a novel structured SBL framework is proposed for uplink channel estimation. An adaptive spatial interpolation scheme is proposed to further reduce the hardware complexity of BS while maintaining the estimation performance. The simulation results show that the proposed scheme outperforms the reference schemes significantly for a variety of scenarios with different numbers of RF chains, transmit SNRs, Rician factors and angular spreads. Moreover, the adaptive adjustment of the estimated beam support is crucial for spatial interpolation in order to obtain good estimation performance.

APPENDIX A
PROOF OF (30)

Define $f(\phi) = -\log(\phi)$ and $\phi = \frac{p(\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l,\mathbf{z}_l,\mathbf{y}_l)}{p\left(\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.\right)}$. Since $f(\phi)$ is a convex function of ϕ, using Jensen's inequality we can obtain

$$
\begin{aligned}
&-\int p\left(\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.\right)\log\frac{p\left(\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l,\mathbf{z}_l,\mathbf{y}_l\right)}{p\left(\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.\right)}d\mathbf{z}_l\\
&=\mathbb{E}_{\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.}\left[f(\phi)\right]\\
&\geq f\left(\mathbb{E}_{\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.}\left[\phi\right]\right)\\
&=-\log\left(\int p\left(\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.\right)\frac{p\left(\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l,\mathbf{z}_l,\mathbf{y}_l\right)}{p\left(\mathbf{z}_l\left|\mathbf{y}_l,\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right.\right)}d\mathbf{z}_l\right)\\
&=-\log p\left(\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l,\mathbf{y}_l\right),
\end{aligned}
\tag{56}
$$

which is just the result in (30). Moreover, letting $\{\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l\} = \left\{\hat{\boldsymbol{\alpha}}_l^{(i)},\hat{\boldsymbol{\gamma}}_l^{(i)}\right\}$ and applying the equality $p\left(\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l,\mathbf{z}_l,\mathbf{y}_l\right) = p\left(\mathbf{z}_l\left|\mathbf{y}_l,\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l\right.\right)p\left(\mathbf{y}_l,\boldsymbol{\alpha}_l,\boldsymbol{\gamma}_l\right)$, we can easily seen that (30) holds with equality. This completes the proof.

REFERENCES

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

[2] J. Hoydis, S. Brink, and M. Debbah, "Massive MIMO in UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[3] S. Jin, X. Liang, K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.

[4] L. You, X. Gao, X. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

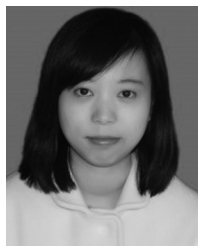
[6] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] Y. Ding, S. E. Chiu, and B. Rao, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 499–513, Jun. 2018.

[8] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, pp. 2898–2904.

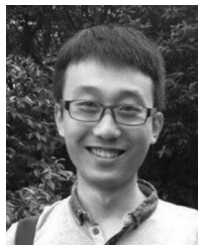
[9] Y. Xu, X. Xia, K. Xu, and Y. Wang, "Three-dimension massive MIMO for air-to-ground transmission: location-assisted precoding and impact of AoD uncertainty," *IEEE Access*, vol. 5, pp. 15582–15596, 2017.

[10] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Jun. 2013.
[11] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.
[12] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.
[13] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.
[14] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
[15] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[16] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "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.
[17] J. Lee, G. Gil, and Y. 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.
[18] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
[19] C. C. Tseng, J. Y. Wu, and T. S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block ℓ_1-minimization," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1055–1067, Mar. 2016.
[20] X. Cheng, J. Sun, and S. Li, "Channel estimation for FDD multi-user massive MIMO: A variational Bayesian inference-based approach," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7590–7602, Nov. 2017.
[21] A. Liu, V. K. N. Lau, and W. Dai, "Exploiting burst-sparsity in massive MIMO with partial channel support information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7820–7830, Nov. 2016.
[22] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
[23] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th edition. New York, NY, USA: Academic, 2007.
[24] Liu, W. Wang, Y. Li, M. Hu, and H. Zhang, "Dual layer beamforming with limited feedback for full-dimension MIMO systems," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Hong Kong, 2015, pp. 2329–2333.
[25] X. Li, S. Jin, H. Suraweera, J. Hou, and X. Gao, "Statistical 3-D beamforming for large-scale MIMO downlink systems over Rician channels," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1529–1543, Apr. 2016.
[26] X. Li, S. Jin, X. Gao, and R. W. Heath, "Three-dimensional beamforming for large-scale FD-MIMO systems exploiting statistical channel state information," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 8992–9005, Nov. 2016.
[27] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.
[28] N. Song, T. Yang, and H. Sun, "Overlapped subarray based hybrid beamforming for millimeter wave multiuser massive MIMO," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 550–554, May 2017.
[29] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
[30] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
[31] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[32] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[33] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[34] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
[35] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.
[36] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," Dept. Comput. Appl. Math., Rice University, Houston, TX, USA, Tech. Rep. TR11-06, 2011.
[37] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.



Yurong Wang was born in China, 1990. She received the B.S. degree in computer science and technology in 2013 from the Beijing Institute of Technology, Beijing, China, and the M.S. degree in communication and information system in 2016 from the Army Engineering University of PLA, Nanjing, China, where she is currently working toward the Ph.D. degree with the Institution of Communications Engineering. Her research interests include full-duplex communication, massive MIMO system, and broadband wireless communications.

Aijun Liu (M'15) received the B.S. degree in microwave communications in 1990, and the M.S. and Ph.D. degrees in communications engineering and information systems in 1994 and 1997, respectively, from the College of Communications Engineering, Army Engineering University of PLA, Nanjing, China, where he is currently a Full Professor. His research interests include satellite communication system theory, signal processing, space heterogeneous networks, channel coding, and information theory.



Xiaochen Xia was born in China, 1987. He received the B.S. degree in electronic science and technology from Tianjin University, Tianjin, China, in 2010, and the M.S. and Ph.D degrees in communication and information system from the Army Engineering University of PLA, Nanjing, China, in 2013 and 2017, respectively. His research interests include signal processing in MIMO systems, machine learning for communications, cooperative networks, and full-duplex communications. He received the 2013 Excellent Master Degree Dissertation Award of Jiangsu Province (China) and Army Engineering University of PLA. He has also received the 2018 Excellent Doctor Degree Dissertation Award of Chinese Institute of Command and Control and Army Engineering University of PLA. He is the Reviewer for the IEEE WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEE ACCESS, and etc.

Kui Xu was born in 1982. He received the B.S. degree in wireless communications in 2004 and the Ph.D. degree in software-defined radio in 2009 from the Army Engineering University of PLA, Nanjing, China, where he is currently a Lecturer in the College of Communications Engineering. Since December 2013, he has been an Associate Professor. He is the author of about 50 papers in refereed journals and conference proceedings and holds five patents in China. His research interests include broadband wireless communications, signal processing for communications, network coding, and wireless communication networks. He was on the Technical Program Committee of the IEEE Wireless Communications and Signal Processing 2014. He received the URSI Young Scientists Award in 2014 and the 2010 ten Excellent Doctor Degree Dissertation Award of Army Engineering University of PLA. He is the Reviewer for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTER, and IEEE SIGNAL PROCESSING LETTERS.