Multi-Layer Sparse Bayesian Learning for mmWave Channel Estimation

Yaoyuan Zhang, Mohammed El-Hajjar, *Senior Member, IEEE,* and Lie-liang Yang, *Fellow, IEEE,*

***Abstract*—Millimeter wave (mmWave) communications has been considered one of the key techniques for the future generations of wireless systems due to the large mmWave bandwidth available. In mmWave systems, channel state information (CSI) is critical for the design of the precoder and combiner for operations respectively at transmitter and receiver. In this paper, we motivate to design the low-complexity and high-accuracy channel estimation methods for the mmWave systems employing orthogonal frequency division multiplexing (OFDM) signaling and hybrid transmitter/receiver beamforming. Specifically, a multi-layer sparse bayesian learning (SBL) channel estimator is proposed to both improve the performance of channel estimation and reduce the complexity of signal processing, when compared with a range of related channel estimators, including the orthogonal matching pursuit (OMP)-, approximate message passing (AMP)- and conventional SBL-assisted channel estimators. The proposed multi-layer SBL estimator is compared with these legacy channel estimators, when impacts from different perspectives are considered. Furthermore, the Bayesian Cramer-Rao Bound of channel estimation is analyzed and evaluated. Our studies and simulation results show that the proposed multi-layer SBL estimator is capable of achieving better performance than the benchmark estimators considered. Specifically, when compared with the traditional SBL estimator, the proposed multi-layer SBL estimator is capable of achieving a lower mean-square error (MSE), while simultaneously, requiring only about 1/10 of the computational complexity of the traditional SBL estimator.**

***Keywords*—mmWave, OFDM, Channel Estimation, Sparse Bayesian Learning, Adaptive Codebook, Multi-layer Structure.**

I. Introduction

Millimeter wave (mmWave) communications is expected to be one of the key techniques for the future generation wireless systems, due to the huge bandwidth resource available in the 3GHz to 300GHz frequency spectrum [1]. However, the mmWave propagation environment is more challenging than the sub-6GHz frequencies [1], [2]. Fortunately, owing to the short wavelength in mmWaves, a large number of antennas can be packed in a relatively small area and hence massive multiple-input multiple output (MIMO) systems can be implemented [3], [4]. Consequently, a high beamforming gain can be attained to overcome the high pathloss in mmWave communication.

Conventionally, there are two beamforming techniques widely used. The first one is the digital beamforming, which controls both the phase and amplitude of signal in the digital domain. However, it requires one distinct radio frequency (RF) chain for each antenna [5]–[8], resulting in dramatic power consumption and hardware cost, when a large antenna array is employed [9]. The second technique is the analog beamforming, which uses analog phase shifters to control only the phase of a signal. Analog beamforming is characterized by low hardware cost and power consumption [10]. However, its performance is usually worse than that of digital beamforming, because it is constrained by the constant amplitude. Additionally, analog beamforming results in degraded performance in multi-user systems [5], due to its low capability of multi-user interference mitigation. Considering the merits and shortcomings of the digital and analog beamforming, hybrid beamforming has been developed so as to approach the performance of digital beamforming, while taking the advantages of the reduced cost and power consumption of analog beamforming [11], [12].

The spectrum efficiency (SE) of hybrid beamforming is shown to be greatly constrained by the channel state information (CSI) [1]. Therefore, it is important to employ reliable channel estimation in the mmWave communications employing hybrid beamforming. Furthermore, orthogonal frequency division multiplexing (OFDM) has been widely employed in wireless communication systems for overcoming frequency selective fading, multipath effects [13], etc. Hence, channel estimation with OFDM is practically desirable. Therefore, *in this paper, we consider the challenge of channel estimation in the mmWave OFDM-aided MIMO systems.*

mmWave channels, as opposed to the sub-6GHz channels, are typically sparse channels [1], [14]–[16]. The conventional channel estimation techniques used in the microwave band communications like adaptive filtering algorithms [17] are usually not efficient for mmWave channel estimation, as they requires a high pilot overhead for accurate channel estimation, when applied for estimating the sparse mmWave channels [9]. Moreover, with the analog precoder and combiner employed in the mmWave, the estimated channel matrix can not be directly accessed because the measured channel in the digital domain is intertwined with the applied analog precoding and combining vectors [9]. For solving these problems, some channel estimators for mmWave channels have been proposed and studied, as seen, for example, as seen in [12], [18], [19].

One of the most popular methods applied for sparse channel estimation is compressed sensing (CS) [1]. For example, in [20], a CS-aided Beam Split Pattern Detection (BSPD)-based channel estimation scheme was proposed for the channel estimation in the Terahertz (THz) massive MIMO systems. In the context of CS, the orthogonal matching pursuit (OMP) algorithm is one of the popular algorithms applied for different

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom (email:{yz2m19,meh,lly}@ecs.soton.ac.uk).

purposes in wireless communications [21]. [22] presents a novel CS-based algorithm for jointly estimating channels and detecting signals. Unlike traditional methods, it combines channel estimation and signal recovery in an iterative process, using estimated data iteratively to improve performance. [23] addresses optimization problems in both uplink and downlink systems. It provides innovative approaches, such as the lternating direction method of multipliers (ADMM) method and alternating optimization (AO) method, to tackle the complexities of beamforming and power control, while also establishing their convergence properties. [24] focuses on leveraging the PARAllel FACtor (PARAFAC) decomposition along with the Alternating Least Squares (ALS) and Vector Approximate Message Passing (VAMP) algorithm to iteratively recover signals from noisy observations, while specifically addressing the uplink of a multi-user Multiple-Input Single-Output (MISO) communication system empowered by RIS. Specifically, related to channel estimation, the authors of [25] proposed an OMP-based open-loop channel estimation algorithm for the mmWave MIMO systems with hybrid beamforming. In this study, the channel model was assumed to be an angle-grid channel, representing a parametric channel with the quantized angle-of-departure (AoD) and angle-of-arrival (AoA). In [26], the authors proposed two OMP-based channel estimation algorithms for OFDM systems, which make use of the joint sparse recovery to estimate the channel information shared by different subcarriers. An adaptive codebook based channel estimator was developed in [11], where the channel information is inferred from a designed codebook. In this approach, an OMP-based method was used in the channel estimation and codebook design, which was shown to outperform the traditional OMP based channel estimation, while at a reduced complexity.

However, the channel estimation method proposed in [11] suffers from poor performance, when signal to noise ratio (SNR) is relatively low. On the other hand, a Sparse Bayesian Learning (SBL) algorithm was first proposed in [27] and [28] as a signal recovery method in CS. Then, for different application scenarios, various SBL-based algorithms were proposed. For example, the multiple measurement vector (MMV) SBL (MSBL) was proposed in [29] [30] to improve the performance of signal recovery of the single measurement vector (SMV) model. In [31], the authors proposed a cluster blocked SBL (CBSBL) based on blocked SBL (BSBL) for mmWave channel estimation, which is shown to achieve better performance than the BSBL. [32] introduces an RIS-aided Ultra-Reliable Access (URA) architecture, optimizing passive reflection for active device separation, which proposes a unique SBL-based method for constructing data symbols without pilot sequences, backed by a probabilistic model. [33] introduces the SBL-based algorithm for improving User Activity Detection (UAD) and Channel Estimation (CE) in grant-free Random Access (RA) scenarios in an LDS-OFDM system. It transforms the iterative MP-BSBL algorithm into a neural network, enhancing convergence in crowded RA scenarios. Importantly, the DNN-MP-BSBL algorithm's training is conducted offline, ensuring real-time implementation feasibility with low-latency requirements in crowded systems. However, *One of its main shortcomings is the high computational complexity, which is challenging for its application in channel estimation. This is because channel is time varying, making the algorithms with high complexity not feasible for estimating channels for instantaneous application, which is especially the case in high resolution situations.* Furthermore, in [34], an angle domain off-grid channel estimation algorithm for the uplink mmWave massive MIMO systems was proposed, which improves the accuracy of the angle estimation by operating the Expectation Maximization (EM) algorithm twice. With this method, more accurate angles are firstly estimated by the SBL algorithm and then the EM algorithm is operated to provide the finer angle estimation inside the angle ranges.

The SBL-based algorithms show better performance for channel estimation than the OMP-based algorithm [9]. In practice, when given the time varying nature of channels, it is essential to develop the high accuracy and low-complexity channel estimation techniques. From literature and the above discussion, we know that the SBL-based channel estimation is capable of achieving better performance than the OMP-based channel estimation, while at the trade-off of significantly higher complexity. Hence, one of the main challenges for designing the SBL-based channel estimation algorithms is reducing their computational complexity, while at the low cost of performance degradation. *To achieve this objective, in this paper, we propose the multi-layer SBL algorithm so as to maintain a good performance of channel estimation, while simultaneously reducing the computational complexity.*

To be more specific, a multi-layer SBL algorithm is proposed, which has lower complexity than the traditional SBL-based channel estimation. In our multi-layer SBL algorithm, the accuracy of channel estimated is increased via a multi-layer estimator by increasing from layer to layer the angle domain resolution layer by layer. At the same time, its computational complexity is reduced, as it only requires a small size dictionary in each layer. In summary, the novel contributions of this paper can be stated as follows.

- A multi-layer SBL channel estimation algorithm is proposed for the mmWave OFDM systems to increase the estimation accuracy and simultaneously reduce the computational complexity of the single-layer SBL-based channel estimation algorithm. This is achieved by selectively increasing the angle resolution for channel estimation layer by layer.
- The proposed multi-layer structure in the proposed algorithm is a flexible design, allowing to combine different channel estimation techniques in different layers. For example, the SBL-based algorithm can be operated in all layers to attain the lightest channel estimation. Alternatively, an OMP-based channel estimation can be used in the first layer, while the SBL-based algorithm is used in the following layers. Since the OMP algorithm has lower complexity than the SBL algorithm, it allows a flexible design to strike a good trade-off between accuracy and complexity.
- Our simulations compare the performance of different algorithms, showing that the proposed channel estimation algorithm can achieve similar accuracy but with just

1/10 of computational complexity when compared to the single-layer SBL-based channel estimation. Furthermore, the proposed channel estimation is capable of achieving significantly lower mean square error (MSE) than the traditional OMP-based channel estimation, approximate message passing (AMP)-based channel estimation, and the adaptive codebook-based channel estimation.

The rest of the paper is organized as follows: mmWave OFDM system model and the mmWave channel model are provided in Section II. In Section III, the proposed multi-layer SBL algorithm is presented and correspondingly the Bayesian Cramer-Rao Bound is derived for channel estimation. In Section IV, we analyze the computational complexity, and demonstrate the MSE results of the various channel estimation algorithms. Finally, conclusions are provided in Section V.

The notations used in this paper are as follows: Lower-case and upper-case boldface letters $\boldsymbol{a}$ and $\boldsymbol{A}$ denote vectors and matrices, respectively; $(\cdot)^T$ and $(\cdot)^H$ express the transpose and conjugate transpose, respectively; $\mathbb{C}^{M\times N}$ is the set of $(M \times N)$-element in the complex field; $\mathbb{E}[\cdot]$ is the expectation operator; $\otimes$ is the matrix Kronecker product; Mode($\boldsymbol{a}$) is the mode of the elements in vector $\boldsymbol{a}$; $\lceil\cdot\rceil$ is the ceiling operator; $\|\cdot\|_F$ is the Frobenius norm; $vec\{\cdot\}$ denotes vectorization.

II. System Model

Figure 1 shows the block diagram of our OFDM-aided hybrid beamforming mmWave communications system employing K sub-carriers, N_t transmitter antennas and N_r receiver antennas. In the following, we present the processing stages in the transceiver followed by the description of the channel model, and then, the formulation of the channel estimation problem.

A. mmWave OFDM MIMO System Model

Consider an OFDM-based mmWave MIMO system with N_t transmit antennas and N_r receive antennas shown in Fig. 1, where K is the number of sub-carrier, $\mathbf{s}^{(k)} \in \mathbb{C}^{N_s\times 1}$ and $\mathbf{y}^{(k)} \in \mathbb{C}^{N_s\times 1}$ denote the transmitted signals and received signals on the k-th sub-carrier, respectively. N_s denotes the number of data streams. $\mathbf{F}_{BB}^{(k)}$ and $\mathbf{W}_{BB}^{(k)}$ are the baseband precoder and combiner on the k-th sub-carrier, respectively. The transmitter is equipped with N_{RF}^t number of radio frequency (RF) chains, and the receiver has N_{RF}^r RF chains. In addition, the number of RF chains is significantly smaller than the number of antennas, i.e. $N_{RF}^t \ll N_t$, and $N_{RF}^r \ll N_r$.

Based on the conventional transmission model in wireless communication $\mathbf{r}^{(k)} = \mathbf{H}^{(k)}\mathbf{x}^{(k)} + \mathbf{n}^{(k)}$, $\mathbf{r}^{(k)} \in \mathbb{C}^{N_r\times 1}$ is the received signal before the combiners, $\mathbf{x}^{(k)} \in \mathbb{C}^{N_t\times 1}$ is the transmitted signal, which can be given as $\mathbf{x}^{(k)} = \mathbf{F}_{RF}\mathbf{F}_{BB}^{(k)}\mathbf{s}^{(k)}$, with the signal $\mathbf{s}^{(k)}$ being first processed by the digital precoder $\mathbf{F}_{BB}^{(k)} \in \mathbb{C}^{N_{RF}^t\times N_s}$ using N_{RF}^t RF chains and then processed by the analog precoder $\mathbf{F}_{RF} \in \mathbb{C}^{N_t\times N_{RF}^t}$ before it is transmitted. Let $\mathbf{H} \in \mathbb{C}^{N_r\times N_t}$ denote a frequency domain mmWave channel, which will be described in the next section. As shown in Fig. 1, in the transmitter, the symbols in $\mathbf{s} = [(\mathbf{s}^{(1)})^T, (\mathbf{s}^{(2)})^T, \ldots, (\mathbf{s}^{(1)})^T)]^T$ are first processed in baseband by the digital precoder $\mathbf{F}_{BB}$ with respect to each of the carriers. Then, the signals are grouped in terms of sub-carriers. On each sub-carrier, after a K-point inverse fast Fourier transform (IFFT) and cyclic prefix (CP) adding process, the signal is transformed to N_{RF} RF chains. Finally, the analog precoder $\mathbf{F}_{RF}$ is applied on all sub-carrier signals before transmitted by the N_t antennas.

Following the reference [15], the received signal from the k-th sub-carrier in Fig. 1 can be expressed as

$$\begin{aligned}\mathbf{y}^{(k)} =&(\mathbf{W}_{BB}^{(k)})^H\mathbf{W}_{RF}^H\mathbf{H}^{(k)}\mathbf{F}_{RF}(\mathbf{F}_{BB}^{(k)})\mathbf{s}^{(k)}\\ &+(\mathbf{W}_{BB}^{(k)})^H\mathbf{W}_{RF}^H\mathbf{n}^{(k)},\end{aligned} \tag{1}$$

where $\mathbf{W}_{RF} \in \mathbb{C}^{N_r\times N_{RF}^r}$ and $\mathbf{W}_{BB}^{(k)} \in \mathbb{C}^{N_{RF}^r\times N_s}$ are the analog combiner and digital combiner, respectively. $\mathbf{n}^{(k)} \in \mathbb{C}^{N_r\times 1}$ denotes a complex, independent, and identically distributed (i.i.d.) zero-mean Gaussian noise vector with a covariance matrix $\sigma^2\mathbf{I}$, where σ^2 is the noise power. Hence, at the receiver side, an analog combiner $\mathbf{W}_{RF}$ is used for combining the received signal to transform it to N_{RF}^r chains. Then, the CP is removed, followed by applying K-point FFT to transform the signal to the frequency domain. Finally, K baseband digital combiners $\mathbf{W}_{BB}$ are operated with the K sub-carriers. Let us assume that N_m symbols are transmitted in N_m time slots, over which the channel is assumed to be the same. Then, we can extend (1) to the form

$$\begin{aligned}\mathbf{Y}^{(k)} =&(\mathbf{W}_{BB}^{(k)})^H\mathbf{W}_{RF}^H\mathbf{H}^{(k)}\mathbf{F}_{RF}\mathbf{F}_{BB}^{(k)}\mathbf{S}^{(k)}\\ &+(\mathbf{W}_{BB}^{(k)})^H\mathbf{W}_{RF}^H\mathbf{n}^{(k)},\end{aligned} \tag{2}$$

where $\mathbf{Y}^{(k)} = [\mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \cdots, \mathbf{y}_{N_m}^{(k)}]$ and $\mathbf{S}^{(k)} = [\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \cdots, \mathbf{s}_{N_m}^{(k)}]$ are measurements and correspondingly transmitted symbols, respectively. *Afterwards,* $\mathbf{S}^{(k)}$ *can be detected using the estimated channel* $\hat{\mathbf{H}}^{(k)}$. *Additionally, the design of the beamformers and combiners require the channel state information (CSI) of* $\hat{\mathbf{H}}^{(k)}$. Therefore, sufficiently accurate channel estimation is essential.

B. Channel Model

The frequency domain mmWave channel model for a uniform linear array (ULA)[1], with respect to the kth sub-carrier, can be expressed as [5], [15]:

$$\mathbf{H}^{(k)} = \sqrt{\frac{N_tN_r}{L}}\sum_{l=1}^{L}\alpha_l^{(k)}\mathbf{a}_r(\phi_l)\mathbf{a}_t^H(\theta_l), \tag{3}$$

where θ_l and ϕ_l are the azimuth angle of departure (AoD) and angle of arrival (AoA) in the l-th path, respectively, $\alpha_l^{(k)}$ is the channel gain of the l-th path on the k-th carrier. The ULA's response vectors $\mathbf{a}_t(\theta_l)$ and $\mathbf{a}_r(\phi_l)$ in angle θ_l and ϕ_l can be expressed as [5], [15]

$$\mathbf{a}_t(\theta_l) = \frac{1}{\sqrt{N}}[1, e^{j\frac{2\pi d}{\lambda}\sin(\theta_l)}, \ldots, e^{j(N-1)\frac{2\pi d}{\lambda}\sin(\theta_l)}]^T, \tag{4}$$

and

$$\mathbf{a}_r(\phi_l) = \frac{1}{\sqrt{N}}[1, e^{j\frac{2\pi d}{\lambda}\sin(\phi_l)}, \ldots, e^{j(N-1)\frac{2\pi d}{\lambda}\sin(\phi_l)}]^T, \tag{5}$$

[1]The designs and analysis in this paper can be similarly applied to any antenna structure and we use the ULA for simplicity.

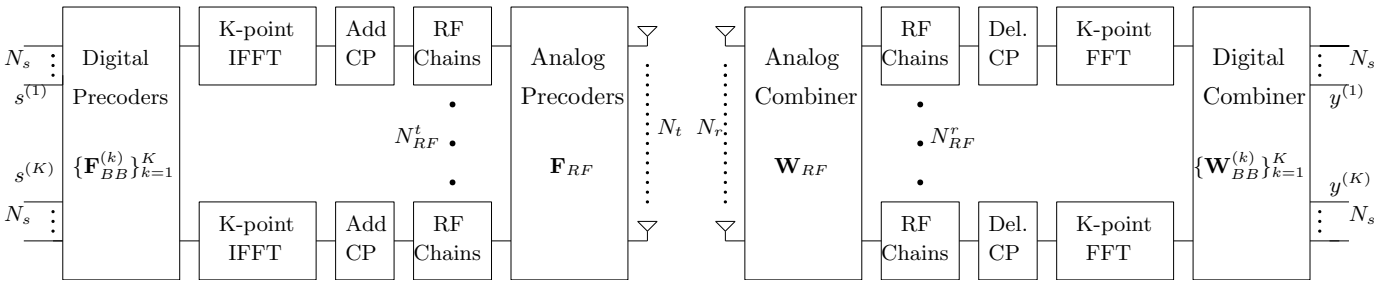

Fig. 1. Block diagram of mmWave OFDM MIMO system with RF and baseband beamformer at both ends.

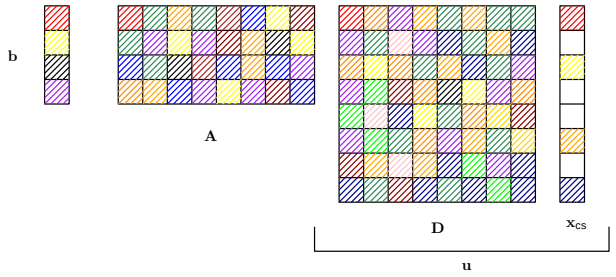

Fig. 2. Illustration of compressed sensing.

where λ is the signal's wavelength, and d denotes the inter-element spacing. Explicitly, the channel matrix (3) can also be expressed in a more compact form as:

$$\mathbf{H}^{(k)} = \mathbf{A}_r \boldsymbol{\alpha}^{(k)} \mathbf{A}_t^H \tag{6}$$

where

$$\mathbf{A}_r = [\mathbf{a}_r(\phi_1), \mathbf{a}_r(\phi_2), \cdots, \mathbf{a}_r(\phi_L)], \tag{7}$$

$$\mathbf{A}_t = [\mathbf{a}_t(\theta_1), \mathbf{a}_t(\theta_2), \cdots, \mathbf{a}_t(\theta_L)], \tag{8}$$

$$\boldsymbol{\alpha}^{(k)} = \gamma \cdot \mathrm{diag}([\alpha_1^{(k)}, \alpha_2^{(k)}, \cdots, \alpha_L^{(k)}]), \tag{9}$$

where $\mathbf{A}_r \in \mathbb{C}^{N_r \times L}$ and $\mathbf{A}_t \in \mathbb{C}^{N_t \times L}$ include all array manifold vectors of the receiver and transmitter, respectively, $\boldsymbol{\alpha}^{(k)} \in \mathbb{C}^{L \times L}$ is the diagonal matrix containing the complex gains of the L paths, $\gamma = \sqrt{\frac{N_t N_r}{L}}$ denotes a normalization factor.

The CSI is required in the beamforming design, where the AoAs, AoDs and $\boldsymbol{\alpha}^{(k)}$ of all the paths are required. It is well-known that mmWave channel is a sparse channel [5], where signals are only spread in certain angles in space. In general, the energy conveyed by the line-of-sight (LoS) path and a few of non-line-of-sight (NLoS) paths, due to the poor propagation characteristics like high pathloss and attenuation [1], [15], [35]. Owing to this, the CS methods are suitable for solving the mmWave channel estimation problem [1]. In general, the CS model can be expressed as [36], [37]

$$\mathbf{b} = \mathbf{A}\mathbf{D}\mathbf{x}_{cs} \tag{10}$$

where $\mathbf{b} \in \mathbb{C}^{m \times 1} (m < n)$ is the encoded vector, $\mathbf{A} \in \mathbb{C}^{m \times n}$ is an encoding matrix independent of the original signal $\mathbf{u} = \mathbf{D}\mathbf{x}_{cs}$, $\mathbf{D}$ denotes the dictionary matrix, and $\mathbf{x}_{cs}$ is the sparse vector [38].

The process of (10) is illustrated in Fig. 2, The object of CS is to recover signal $\mathbf{x}_{cs}$ from $\mathbf{b}$ with known $\mathbf{A}$ and $\mathbf{D}$, which can be formulate as the optimization problem of

$$\min \|\mathbf{x}_{cs}\|_1 \quad s.t.\, \mathbf{A}\mathbf{D}\mathbf{x}_{cs} = \mathbf{b}, \tag{11}$$

where the $1-$norm of $\|\mathbf{x}_{cs}\|$ is used for reducing the computational complexity. The problem of (11) can be solved by the convex optimization algorithms [39].

Since the number of columns (codewords) of the dictionary matrix $\mathbf{D}$ should be larger than the number of elements of the encoded vector $\mathbf{b}$, and the number of codewords of the dictionary matrix depends on the resolution, both the AoD and AoA space $[0, \pi)$ should be divided into $G_T \geq max\{N_t, N_r\}$ and $G_R \geq max\{N_t, N_r\}$ portions, respectively, by the angle space partition method [25]. For example, in a 64×32 MIMO system, we should choose $G_T \geq 64$ and $G_R \geq 64$. Using this method, the quantized angle sets $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ are called the grids. The grids form the set with the condition that $\cos(\theta_g^t)$ and $\cos(\phi_g^t)$ should follow a uniform distribution, which can be expressed as [25]

$$\cos(\theta_g^t) = \frac{2(g-1)}{G_T} - 1, 1 \leq g \leq G_T, \tag{12}$$

$$\cos(\phi_g^r) = \frac{2(g-1)}{G_R} - 1, 1 \leq g \leq G_R, \tag{13}$$

and, the sets $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ can be expressed as

$$\boldsymbol{\Theta} = \{\theta_g^t : \theta_g^t \in [0, \pi), 1 \leq g \leq G_T\}, \tag{14}$$

$$\boldsymbol{\Phi} = \{\phi_g^r : \phi_g^r \in [0, \pi), 1 \leq g \leq G_R\}. \tag{15}$$

For example, if $G_T = 3$, we obtain the grid $\boldsymbol{\Theta} = [\arccos(-1), \arccos(-\frac{1}{3}), \arccos(\frac{1}{3})]$.

Based on the quantized angle sets $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, the transmit and receive array response dictionary matrices can be represented as

$$\mathbf{A}_T(\boldsymbol{\Theta}) = [\mathbf{a}_t(\theta_1^t), \mathbf{a}_t(\theta_2^t), \cdots, \mathbf{a}_t(\theta_{G_T}^t)], \tag{16}$$

$$\mathbf{A}_R(\boldsymbol{\Phi}) = [\mathbf{a}_t(\phi_1^r), \mathbf{a}_t(\phi_2^r), \cdots, \mathbf{a}_t(\phi_{G_R}^r)], \tag{17}$$

which are usually choose to satisfy

$$\mathbf{A}_T(\boldsymbol{\Theta})\mathbf{A}_T(\boldsymbol{\Theta})^H = \frac{G_T}{N_t}\boldsymbol{I}_{N_t}, \tag{18}$$

$$\mathbf{A}_R(\boldsymbol{\Phi})\mathbf{A}_R(\boldsymbol{\Phi})^H = \frac{G_R}{N_r}\boldsymbol{I}_{N_r}. \tag{19}$$

Hence, by replacing $\mathbf{A}_t$ and $\mathbf{A}_r$ in (6) in $\mathbf{A}_T$ and $\mathbf{A}_R$, as defined in (16) and (17), the frequency domain channel can

be expressed based on the beamspace channel $\mathbf{H}_b^{(k)}$ as

$$\mathbf{H}^{(k)} = \mathbf{A}_R \mathbf{H}_b^{(k)} \mathbf{A}_T^H. \tag{20}$$

Compared with the accurate $\mathbf{A}_t$ and $\mathbf{A}_r$ in (8) and (7), which contain the terms matching to the real rays, $\mathbf{A}_T$ and $\mathbf{A}_R$ contain all the possible array responses on the grid covering $[0, \pi]$. Hence, in a mmWave communication environment having limited scatter, $\mathbf{H}_b^{(k)}$ is a sparse matrix with many zero elements, which is beneficial to the channel estimation using CS methods.

In order to invoke CS method to estimate the beamspace channel, the channel matrices in (20) should be vectorized, which can be obtained as

$$\mathbf{h}^{(k)} = (\mathbf{A}_T^* \otimes \mathbf{A}_R)\mathbf{h}_b^{(k)}, \tag{21}$$

where $\mathbf{h}_b^{(k)} = vec\{\mathbf{H}_b^{(k)}\}$, $\mathbf{h}^{(k)} = vec\{\mathbf{H}^{(k)}\}$.

C. Observations for Channel Estimation

In order to facilitate the channel estimation, the observed equation (2) can be vectorized to obtain

$$\mathbf{y}_v^{(k)} = \big(\Big(\mathbf{F}^{(k)}\mathbf{S}^{(k)}\Big)^T \otimes \Big(\mathbf{W}^{(k)}\Big)^H \big)\mathbf{h}^{(k)} + \mathbf{n}_v^{(k)}, \tag{22}$$

where $\mathbf{y}_v^{(k)} = vec\{\mathbf{Y}^{(k)}\}$ and $\mathbf{n}_v^{(k)} = vec\{\mathbf{W}^{(k)}\mathbf{n}^{(k)}\}$ are the vectorized measurement signal and noise, respectively, $\mathbf{F}^{(k)} = \mathbf{F}_{RF}\mathbf{F}_{BB}^{(k)}$, $\mathbf{W}^{(k)} = \mathbf{W}_{RF}\mathbf{W}_{BB}^{(k)}$. Note that, for channel estimation, $\mathbf{S}^{(k)}$ contains pilot symbols. Substituting (21) into (22) yields

$$\mathbf{y}_v^{(k)} = \big(\Big(\mathbf{F}^{(k)}\mathbf{S}^{(k)}\Big)^T \otimes \Big(\mathbf{W}^{(k)}\Big)^H \big)\big(\mathbf{A}_T^* \otimes \mathbf{A}_R\big)\mathbf{h}_b^{(k)} + \mathbf{n}_v^{(k)}. \tag{23}$$

The objective of channel estimation is to estimate $\mathbf{h}_b^{(k)}$ based on (23), for which many techniques, such as OMP-based channel estimation [11], have been proposed. However, the OMP and AMP-based channel estimation techniques [11], [40] usually have poor performance. The SBL-based channel estimation techniques are capable of attaining an improved performance, when compared with the OMP-based techniques, but at the expense of a significantly increased complexity. Therefore, in the following section, we propose a reduced complexity SBL-based multi-layer channel estimation algorithm, which is also capable of achieving an improved performance, when compared with the OMP and AMP-based methods.

III. Sparse Bayesian Learning Channel Estimation

As mentioned above and shown in [28], [31], [41], [42], the SBL-based channel estimation techniques are capable of improving the channel estimation performance in low SNR, but at the expense of a high computational complexity. In this section, we present a SBL-based multi-layer channel estimation algorithm, which improves the performance of channel estimation by utilizing the SBL-based algorithm and simultaneously reduces the computational complexity by invoking multiple layers to obtain the higher resolutions for AoA and AoD. To serve as a benchmark and also the basis, in this section, the conventional SBL algorithm for channel estimation in single-carrier systems is first presented. Then, the SBL-based multi-layer channel estimation algorithm is proposed for the channel estimation in multiple carrier systems.

A. Single-Layer SBL Channel Estimation in Single-Carrier Communications

As discussed in Section II, in channel estimation, the transmitter sends pilot data, and then the channel estimator in the receiver side estimates the CSI, which is essential for the precoder and combiner design. First, the precoder $\mathbf{F}_0$ and combiner $\mathbf{W}_0$ are set randomly, when the transmitter sends pilot data to start the channel estimation process. The grid points needed to establish the beamspace channel model are obtained from the angle space partition method in (12) and (13), where both the AoD and AoA spaces $[0, \pi)$ are divided according to the angle-dividing method described in Section II.

Considering a single-carrier system, as modeled in (23), we can express it as

$$\mathbf{y}_v = \mathbf{P}\mathbf{h}_b + \mathbf{n}_v, \tag{24}$$

where

$$\mathbf{P} = (\mathbf{S}^T\mathbf{F}^T \otimes \mathbf{W}^H)(\mathbf{A}_T^* \otimes \mathbf{A}_R). \tag{25}$$

In (25), $\mathbf{S}$ is the pilot data that is known to both transmitter and receiver. The estimator aims to estimate the beamspace channel $\mathbf{h}_b$ based on the observations $\mathbf{y}_v$ for a given $\mathbf{P}$. First, the parameterized Gaussian prior can be assigned to the unknown channel vector as [30]

$$p(\mathbf{h}_b; \Gamma) = \prod_{i=1}^{G_r G_t} (\pi\gamma_i)^{-1} \exp\left(- \frac{|\mathbf{h}_b(i)|^2}{\gamma_i} \right), \tag{26}$$

where γ_i is the hyper-parameter, and $\mathbf{\Gamma} = diag\{\gamma_1, \gamma_2, \ldots, \gamma_{G_r G_t}\}$. The noise covariance matrix is given by $\mathbf{R}_q = \mathbf{E}[\mathbf{n}_v \mathbf{n}_v^H] = \sigma^2 \mathbf{W}^H \mathbf{W}$ [29]. The posteriori probability density function (PDF) of the beamspace channel vector $\mathbf{h}_b$ can be expressed as $p(\mathbf{h}_b | y; \mathbf{\Gamma}) \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ [30], where

$$\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{P}^H \mathbf{R}_q^{-1} \mathbf{y}_v, \tag{27}$$

and

$$\mathbf{\Sigma} = (\mathbf{P}^H \mathbf{R}_q^{-1} \mathbf{P} + \mathbf{\Gamma}^{-1})^{-1}. \tag{28}$$

Based on these modelings, then the SBL algorithm can be implemented to estimate the channel. The SBL algorithm is a kind of expectation maximization (EM) algorithm, which includes an E-step and a M-step in each iteration [43]. In the E-step, the hyper-parameters are assumed to be known, and the expectation can be obtained for the known hyper-parameters. Then, the maximum likelihood estimation (MLE) is utilized to estimate the new parameters to be used in the next iteration. By repeating the two steps, the parameters will approach their accurate and steady solutions. Hence, in the SBL algorithm, the Bayesian evidence $p(\mathbf{y}; \Gamma)$ can be maximized using the EM algorithm.

In detail, in the E-step, the log-likelihood function can be expressed as [29]

$$\mathcal{L}(\mathbf{\Gamma}|\hat{\mathbf{\Gamma}}) = \mathbf{E}[\log(p(\mathbf{y}, \mathbf{h}_b; \mathbf{\Gamma}))], \tag{29}$$

which can be factorized to

$$\mathcal{L}(\mathbf{\Gamma}|\hat{\mathbf{\Gamma}}) = \mathbf{E}[\log(p(\mathbf{y}|\mathbf{h}_b)) + \log(p(\mathbf{h}_b; \mathbf{\Gamma}))]. \tag{30}$$

The maximization step (M-step), which maximizes $\mathcal{L}(\mathbf{\Gamma}|\hat{\mathbf{\Gamma}})$ with respect to the hyper-parameter γ using the MLE, can be shown as [41]

$$\gamma(n_{iter}) = \arg\max_{\gamma} \mathcal{L}(\mathbf{\Gamma}|\hat{\mathbf{\Gamma}}), \tag{31}$$

where n_{iter} is the iteration number. By solving (31), we can get the hyper-parameter matrix as [30]

$$\mathbf{\Gamma} = \mathbf{\Sigma} + \|diag\{\boldsymbol{\mu}\}\|^2, \tag{32}$$

and the beamspace channel is given as $\hat{\mathbf{h}}_b = \boldsymbol{\mu}$. Then, the frequency domain channel can be obtained based on $\hat{\mathbf{H}}_b = vec^{-1}\{\hat{\mathbf{h}}_b\}$ as

$$\hat{\mathbf{H}} = \mathbf{A}_R \hat{\mathbf{H}}_b \mathbf{A}_T^H. \tag{33}$$

As mentioned before, the single layer SBL-based algorithm requires high computational complexity, since the size of the dictionary matrix for CS depends on the resolution used. In other words, in the case when high resolution is needed for attaining high accuracy, the size of dictionary matrix is large, which makes the complexity of the single layer SBL-based algorithm grow exponentially. Hence, in the next section, we propose the multi-layer SBL algorithm to reduce the complexity, while aiming at maintaining the estimation accuracy.

B. Proposed Multi-layer SBL Channel Estimation

Let us now extend the single-layer SBL channel estimation to the multi-layer structure in order to reduce the complexity. With the multi-layer structure, the AoA and AoD ranges can be first divided to form low resolution dictionaries and then increase the resolution layer after layer, until finally reach G_R and G_T codewords, respectively. In this way, the size of sensing matrix is reduced and correspondingly, the implementation complexity can be reduced. To operate the multi-layer SBL channel estimation, similar to the case of single-layer SBL, the random precoder $\mathbf{F}_0$ and combiner $\mathbf{W}_0$ are used in the first layer.

First, the beamspace channel $\hat{\mathbf{H}}_{b,1}$ is estimated in the first layer by the SBL algorithm with the resolutions $G_{T,1} > N_t$ and $G_{R,1} > N_r$, which can be set to the values that are significantly smaller than that in the single-layer SBL algorithm described in Section III-A. In an idealized case, after estimating the sparse beamspace channel in the first SBL layer, we obtain a channel matrix $\hat{\mathbf{H}}_{b,1}$, which has a 0-value element corresponding to the beam having no or very small energy, and a $h_l \neq$ 0-value element corresponding to a beam in this direction, where l denotes the l-th path (beam). Based on $\hat{\mathbf{H}}_{b,1}$, the directions can also be estimated with the aid of the codebook used in the first layer. For example, let the idealized beamspace channel estimated be expressed as

$$\hat{\mathbf{H}}_{b,1} = \begin{bmatrix} 0 & h_1 & \cdots & 0 \\ 0 & 0 & \cdots & h_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{34}$$

where except h_1 and h_2, all other elements are zero. Then, from $\mathbf{H}_{b,1}$ we can know that the beams are located in the second angle and the $G_{T,1}$-th angle in the transmit array response dictionary $\mathbf{A}_{T,1}$, and in the first angle and the second angle in the receive array response dictionary $\mathbf{A}_{R,1}$.

Afterwards, the estimated angles are further divided into higher resolutions of $G_{T,2}$ and $G_{R,2}$, from which the new array response dictionary matrices $\mathbf{A}_{T,2}$ and $\mathbf{A}_{R,2}$ can be obtained. Furthermore, the new precoder $\mathbf{F}_1$ and combiner $\mathbf{W}_1$ are calculated based on the singular value decomposition (SVD) of the estimated channel $\hat{\mathbf{H}}_1$. With this updated information, the SBL algorithm estimates the channel in the second layer, giving as $\hat{\mathbf{H}}_2$. Similarly, the resolutions and the precoder/combiner can be updated for the third layer of the channel estimation. This process repeats until the estimated channel converges or when the maximum number of iterations is reached. From the above description we can know that with the multi-layer method, the size of codebooks used in each layer is relatively small. Hence, the size of sensing matrix can be significantly reduced in comparison with the single-layer case, which results in significant reduction of complexity.

However, we should note that the estimated channel matrix $\hat{\mathbf{H}}_{b,1}$ in beamspace is not exactly that as presented in (34). Instead of the non-zero h_l-value elements and 0-value elements, $\mathbf{H}_{b,1}$ usually contains the relatively large values representing the h_l of activated beams and the relatively small values representing the 0 of inactive beams. Besides, due to the beam matching issue and the channel estimation errors, the positions of the largest values are not necessary in the exact direction of the active beams. Based on the above consideration, below we introduce three methods to select the active beams (paths) and analyze the channel estimation performance.

Method 1: Fixed number of paths: With this method, the number of active paths is assumed to be L_d. Then, after channel estimation, the L_d largest values in $\hat{\mathbf{H}}_{b,1}$ are selected, while all the other elements are set to '0'.

Method 2: Threshold based selection: In this method, the elements in $\hat{\mathbf{H}}_{b,1}$ with the highest power are selected until the sum of their power reaches a pre-set threshold.

Method 3: Joint threshold and number of paths. With this method, the maximum number of paths is assumed to be L_d. Then, the threshold based method is executed until either the total power as in Method 2 is reached or the number of selected elements reaches L_d.

Let us illustrate the multi-layer SBL technique using a 2-layer structure example, in order to further explain the proposed technique. Assume that $S = 2$, $N_1 = 8$, $N_2 = 16$ and $K = 2$, which are used to divide the AoAs and AoDs as shown in Fig. 3, i.e., the first layer has a resolution of $\frac{2\pi}{8}$, and the second layer $\frac{2\pi}{16}$. In the first layer, the beam

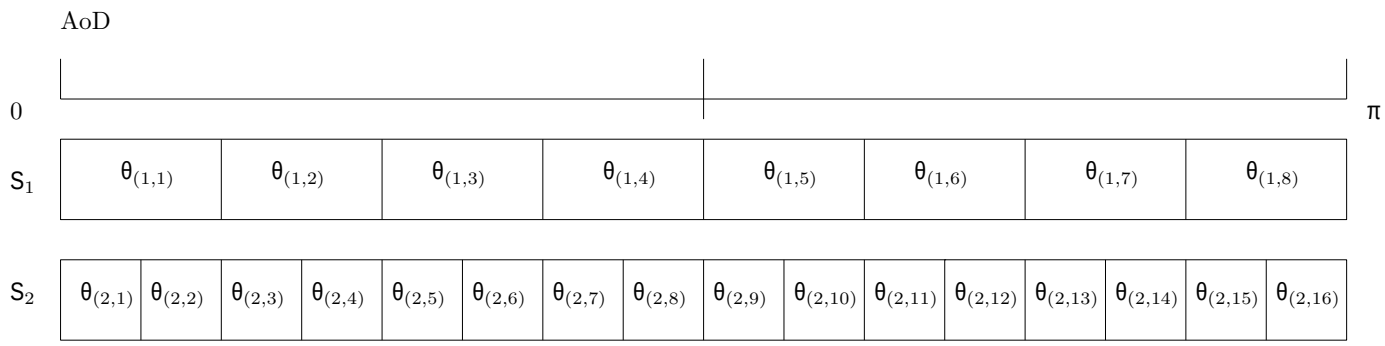

Fig. 3. Illustration of the angle region division in the proposed algorithm using 2 layers, 8 codewords for the first layer and 16 codewords for the second layer.

search considers the angles in the set S_1 shown in Fig. 3. Afterwards, the estimated beamspace channel $\hat{\mathbf{h}}_{b,1}$ in the first layer can be obtained using the single-layer SBL channel estimation algorithm as described in the previous section. Then, the positions of the L highest values in $\hat{\mathbf{h}}_{b,1}$ can be determined, which correspond to the L beams having the relatively higher power in the beam domain. For example, let us assume that the target beams at transmitter side are located in $\{\theta_{(2,1)}, \theta_{(2,7)}, \theta_{(2,15)}\}$, and that at receiver side are located in $\{\phi_{(2,5)}, \phi_{(2,9)}, \phi_{(2,12)}\}$ [2]. As shown in Fig.3, the angular ranges in the first layer matching the target angles should be $\{\theta_{(1,1)}, \theta_{(1,4)}, \theta_{(1,8)}\}$ and $\{\phi_{(1,3)}, \phi_{(1,5)}, \phi_{(1,6)}\}$ for transmitter side and receiver side, respectively. When assuming an idealized case that the estimated beamspace channel are accurate, the vectorized estimated sparse beamspace channel $\hat{\mathbf{h}}_{b,1}$ in the first layer can be represented in the form of

$$\hat{\mathbf{h}}_{b,1} = \{0, \dots, \hat{h}_1, 0, \dots, \hat{h}_2, 0, \dots, \hat{h}_3, 0, \dots\}, \tag{35}$$

where $\hat{h}_1$, $\hat{h}_2$, and $\hat{h}_3$ are three non-zero values corresponding to the three beam angles, as above-mentioned. Furthermore, the positions of $\hat{h}_1$, $\hat{h}_2$ and $\hat{h}_3$ correspond to the columns $\mathbf{A}_T$ or $\mathbf{A}_R$ in the first layer dictionary matrix, which are determined by the angles in $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$. Hence, when we select these angles to generate $\mathbf{A}_T$ and $\mathbf{A}_R$ and remove the small values to '0' in the beamspace channel, the quantized-channel will not change. Consequently, the frequency-domain channel matrix can be expressed as

$$\begin{aligned}\mathbf{H} =& \mathbf{A}_R(\boldsymbol{\Phi})\mathbf{H}_b\mathbf{A}_T^H(\boldsymbol{\Theta}) \\ =& \mathbf{A}_R(\phi_{1,3}, \phi_{1,5}, \phi_{1,6})\bar{\mathbf{H}}_b\mathbf{A}_T^H(\theta_{1,1}, \theta_{1,4}, \theta_{1,8}),\end{aligned} \tag{36}$$

where $\bar{\mathbf{H}}_b \in \mathbb{C}^{3\times 3}$ is the beamspace channel matrix estimated in the first layer.

Then, following Fig. 3, $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Phi}_2$ are updated to the higher resolution for the second layer within the new ranges of the estimated angles as follow:

$$\boldsymbol{\Theta}_2 = \left\{\theta_{(2,1)}, \theta_{(2,2)}; \theta_{(2,7)}, \theta_{(2,8)}; \theta_{(2,15)}, \theta_{(2,16)}\right\}, \tag{37}$$

$$\boldsymbol{\Phi}_2 = \left\{\phi_{(2,5)}, \phi_{(2,6)}; \phi_{(2,9)}, \phi_{(2,10)}; \phi_{(2,11)}, \phi_{(2,12)}\right\}. \tag{38}$$

Based on $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Phi}_2$, the process of layer 1 is repeated to obtain the updated $\mathbf{A}_T$, $\mathbf{A}_R$, precoder and combiner. Furthermore, a new sensing matrix $\mathbf{P}_2$ as shown in (26) can be obtained, based on which the SBL algorithm can update the estimation of $\mathbf{h}_b$. Overall, the multi-layer SBL algorithm is presented in Algorithm 1, where Step 19 can be achieved by implementing the SVD on the estimated channel matrix in a layer.

The complexity of the proposed multi-layer SBL algorithm can be further reduced at the expense of some channel estimation accuracy. More specifically, the first layer of the multi-layer algorithm can be replace by the adaptive codebook-based channel estimation [11], which uses the OMP-based channel

Algorithm 1: Multi-Layer SBL Channel Estimation

Data: AoDs are initialized to $\boldsymbol{\Theta}_0$, AoAs are initialized to $\boldsymbol{\Phi}_0$, maximum iteration number k_{max}, threshold ϵ

Input: $\boldsymbol{\gamma}_0 = \mathbf{I}$

1 Generate random $\mathbf{F}_0$ from (49)
2 Generate random $\mathbf{W}_0$ from (50)
3 **for** $s = 1 : S_{max}$ **do**
4 **for** $i = n : N_{max}$ **do**
5 Calculate sensing matrix
6 $\mathbf{P}_s = (\mathbf{s}^T\mathbf{F}_{s-1}^T \otimes \mathbf{W}_{s-1}^H)(\mathbf{A}_T^*(\boldsymbol{\Theta}_{s-1}) \otimes \mathbf{A}_R(\boldsymbol{\Phi}_{s-1}))$
7 Compute $\boldsymbol{\sigma}$ from (28)
8 Compute $\boldsymbol{\mu}$ from (27)
9 **for** $i = 1 : G_T G_R$ **do**
10 $\boldsymbol{\gamma}_n(i) = \|\boldsymbol{\mu}(i)\|^2 + \boldsymbol{\Sigma}(i,i)$
11 **end**
12 $\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma}_n)$
13 **if** $\|\boldsymbol{\gamma}_n - \boldsymbol{\gamma}_{n-1}\|^2 \leq \epsilon$, **then**
14 break
15 **end**
16 $\hat{\mathbf{h}}_b = \boldsymbol{\mu}$
17 **end**
18 Update $\boldsymbol{\Theta}_s$ and $\boldsymbol{\Phi}_s$
19 Update $\mathbf{F}_s$ and $\mathbf{W}_s$ via SVD of the estimated channel matrix
20 Update $\mathbf{A}_T$ and $\mathbf{A}_R$ based on $\boldsymbol{\Theta}_s$ and $\boldsymbol{\Phi}_s$
21 **end**

[2]We use θ to represent the angle at transmitter side and ϕ to represent that at receiver side.

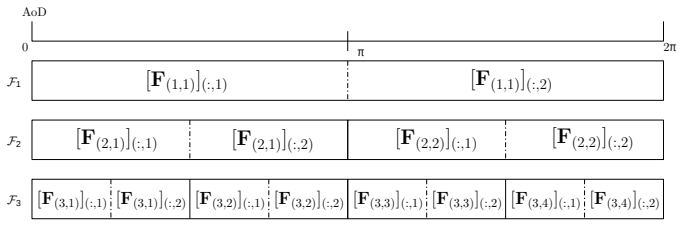

Fig. 4. Multi-resolution codebook structure with $S = 3$ stages and each subset having $K = 2$ codewords [11].

estimation in codebook structure to estimate the rough channel matrix. Then, this estimated channel matrix by OMP can be used as the initial conditions for the following layers' estimation using the SBL algorithm. For example, let us consider the adaptive codebook in Fig. 4, where a beamspace channel with a resolution of $N = 8$ can be obtained after applying the adaptive codebook channel estimation [12]. The results can be considered as the first layer estimation in the context of the multi-layer structure shown in Fig.3. Then, starting from the second layer, the SBL algorithm is implemented to improve the accuracy, as described in the multi-layer SBL channel estimation algorithm. In this case, the complexity can be significantly reduced because the complexity of the OMP-based algorithm is significantly lower than that of the SBL-based algorithm. However, the MSE performance may not be as good as that of the full multi-layer SBL channel estimation, because the angle estimation of the adaptive codebook based algorithm is not as accurate as that obtained by the SBL channel estimation. The details of the OMP assisted multi-layer SBL algorithm is summarized in Algorithm 2. The MSE and angle estimation performance of these algorithms will be presented and compared in Section IV.

C. Multi-layer SBL Channel Estimation in OFDM Systems

After considering the multi-layer SBL algorithm in the single-carrier scenario, in this section, we extend it to the OFDM systems, where the signals conveyed by multiple subcarriers can be exploited to enhance the performance of channel estimation. Assume an OFDM system employing K sub-carriers, with $\mathbf{s}^{(k)}$ and $\mathbf{y}^{(k)}$ denoting respectively the transmitted and received signals on the k-th sub-carrier. $\mathbf{F}_{BB}^{(k)}$ and $\mathbf{W}_{BB}^{(k)}$ are the baseband precoder and combiner for the k-th sub-carrier. We assume that a transmitter with N_t antennas and N_{RF}^t RF chains communicates with a receiver employing N_r antennas and N_{RF}^r RF chains. The mmWave channel model in the OFDM system is the same as that represented by (3). Similarly, the vectorized channel of a subcarrier can be expressed as (21). Overall, the mmWave MIMO OFDM channel can be represented as

$$\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_K^T]^T, \tag{39}$$

which can be expressed using the beamspace channel as

$$\mathbf{h} = (\mathbf{I}_K \otimes \mathbf{A}_T^* \otimes \mathbf{A}_R)\mathbf{h}_b, \tag{40}$$

where

$$\mathbf{h}_b = [\mathbf{h}_{b,1}^T, \mathbf{h}_{b,2}^T, \ldots, \mathbf{h}_{b,K}^T]^T. \tag{41}$$

Algorithm 2: Adaptive Codebook OMP-assisted Multi-Layer SBL Channel Estimation

Data: Both transmitter and receiver know N, K, L_d, codebook $\mathcal{F}$ and $\mathcal{W}$, angle grid $\boldsymbol{A}_F$, $\boldsymbol{A}_W$, maximum iteration number k_{max}, threshold ϵ

Input: $S = \log_K N$, $\hat{\boldsymbol{H}} = 0$

$l = 1$
if $l \leq L_{omp}$ **then**
  $\hat{\boldsymbol{h}} = vec\{\hat{\boldsymbol{H}}\}$
  $k_1^{(t)} = 1$; $k_1^{(r)} = 1$; $s = 1$
  if $s \leq S$ **then**
    Transmitter applies $[\boldsymbol{F}_{(s,k_s^{(t)})}]$ for precoding
    Receiver uses $[\boldsymbol{W}_{(s,k_s^{(r)})}]$ for combining
    $\boldsymbol{y} = \boldsymbol{F}_{(s,k_s^{(t)})} \otimes \boldsymbol{W}^H_{(s,k_s^{(t)})}(\boldsymbol{h} - \hat{\boldsymbol{h}}) + \boldsymbol{e}$
    $\hat{m} = \arg\max_{\forall m} \|\boldsymbol{y}(m)\|$
    $p = \text{Mode}(\hat{m})$
    $\hat{m}_t = \lceil p/K \rceil$
    $\hat{m}_r = p - (\hat{m}_t - 1)K$
    $k_{s+1}^{(t)} = (\hat{m}_t - 1)K + m_t$
    $k_{s+1}^{(r)} = (\hat{m}_r - 1)K + m_r$
    $s = s + 1$
  end
  $\hat{\boldsymbol{h}}_b = \frac{L_e}{\sqrt{N_t N_r}}[\boldsymbol{y}(p,:)]$
  $\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}} + \frac{\sqrt{N_t N_r}}{L_e}\hat{\boldsymbol{h}}_b \otimes [\boldsymbol{A}_F(:, k_{s+1}^{(t)})\boldsymbol{A}_W^H(:, k_{s+1}^{(t)})]$
  $\boldsymbol{\theta}_l = \boldsymbol{\Theta}(k_{s+1}^{(t)})$
  $\boldsymbol{\phi}_l = \boldsymbol{\Phi}(k_{s+1}^{(t)})$
  $l = l + 1$
end
$\boldsymbol{\Theta}_{ac} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{L_{omp}}]$
$\boldsymbol{\Phi}_{ac} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{L_{omp}}]$
Generate random $\mathbf{F}_0$ from (49)
Generate random $\mathbf{W}_0$ from (50)
for $n = 1 : N_{max}$ **do**
  Calculate sensing matrix:
  $\mathbf{P}_s = (\mathbf{s}^T\mathbf{F}_0^T \otimes \mathbf{W}_0^H)(\mathbf{A}_T^*(\boldsymbol{\Theta}_{ac}) \otimes \mathbf{A}_R(\boldsymbol{\Phi}_{ac}))$
  Compute $\boldsymbol{\sigma}$ from (28)
  Compute $\boldsymbol{\mu}$ from (27)
  for $i = 1 : G_T G_R$ **do**
    $\boldsymbol{\gamma}_n(i) = \|\boldsymbol{\mu}(i)\|^2 + \boldsymbol{\Sigma}(i,i)$
  end
  $\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma}_n)$
  if $\|\boldsymbol{\gamma}_n - \boldsymbol{\gamma}_{n-1}\|^2 \leq \epsilon$, **then**
    break
  end
  $\hat{\mathbf{h}}_b = \boldsymbol{\mu}$
end

In an OFDM system, physically, the AoAs and AoDs of different sub-carrier signals are the same, but the complex gains of different sub-carriers may be different [44]. Therefore, the positions of the beams in the beamspace channels of different sub-carriers are the same. In this case, for beam estimation, the most commonly estimated angles from different carriers can be selected as the most reliably estimated angle for the next layer of estimation. Consequently, when the multi-layer SBL algorithm is employed, we can not only reduce the complexity of channel estimation, but also improve the accuracy of estimated channels.

The details of the multi-layer SBL algorithm for OFDM systems are as follows. First, the channels of different sub-carriers are estimated individually in the first layer using Algorithm 1 or using the adaptive codebook based OMP algorithm [12]. After obtaining the estimated angles from different sub-carriers, a specific angle is selected based on majority votes of the individual sub-carrier estimates, which gives the final angle estimation of a layer. With the aid of this finally estimated angle, then, the dictionary in the second layer is calculated and used for generating the precoder and combiner for the second layer of estimation. These processes are repeat until sufficiently accurate channel estimation is achieved. To sumarize, the above described channel estimation is stated as Algorithm 3. Note again, the first layer estimation can be implemented by the adaptive codebook based OMP algorithm, as in Algorithm 2, in order to further reduce complexity.

Algorithm 3: Multi-Layer SBL Channel Estimation in OFDM system

Data: AoDs initialized $\mathbf{\Theta}_0$, AoAs initialized $\mathbf{\Phi}_0$, maximum iteration number k_{max}, threshold ϵ
Input: $\boldsymbol{\gamma}_0 = \mathbf{I}$
Generate random $\mathbf{F}_0$ from (49)
Generate random $\mathbf{W}_0$ from (50)
for $s = 1 : S_{max}$ **do**
  for $k = 1 : K$ **do**
    for $n = 1 : N_{max}$ **do**
      Calculate sensing matrix $\mathbf{P}_{s,k} = (\mathbf{s}_k^T \mathbf{F}_{s-1,k}^T \otimes \mathbf{W}_{s-1,k}^H)(\mathbf{A}_T^*(\mathbf{\Theta}_{s-1}) \otimes \mathbf{A}_R(\mathbf{\Phi}_{s-1}))$
      Compute $\boldsymbol{\sigma}_k$ from (28)
      Compute $\boldsymbol{\mu}_k$ from (27)
      for $i = 1 : G_T G_R$ **do**
        $\boldsymbol{\gamma}_{n,k}(i) = \|\boldsymbol{\mu}_k(i)\|^2 + \mathbf{\Sigma}_k(i,i)$
      end
      $\mathbf{\Gamma}_k = diag(\boldsymbol{\gamma}_{n,k})$
      if $\|\boldsymbol{\gamma}_{n,k} - \boldsymbol{\gamma}_{n-1,k}\|^2 \leq \epsilon$ **then**
        break
      end
      $\hat{\mathbf{h}}_{k,b} = \boldsymbol{\mu}_k$
    end
    Generate $\mathbf{\Theta}_{s,k}$ and $\mathbf{\Phi}_{s,k}$ based on $\hat{\mathbf{h}}_{k,b}$
  end
  Update $\mathbf{\Theta}_s$ and $\mathbf{\Phi}_s$ via identifying the most frequency estimated angles from by the K sub-carriers
  Update $\mathbf{A}_T$ and $\mathbf{A}_R$ using $\mathbf{\Theta}_s$ and $\mathbf{\Phi}_s$
  Update $\mathbf{F}_s$ and $\mathbf{W}_s$ based on $\mathbf{\Theta}_s$ and $\mathbf{\Phi}_s$
end

D. Bayesian Cramer-Rao Bound

Cramer-Rao Bound (CRB) has been widely used to gain the information about the fundamental limits of estimation problems [41]. For a Bayesian estimator, the Bayesian Cramer-Rao Bound (BCRB) can be used to find a lower bound of the MSE distortion [45], [46]. In particular, BCRB can be calculated by inverting the Bayesian Fisher Information Matrix (BFIM), which consists of two parts. The first part is the Fisher Information Matrix (FIM) with respect to the observed signal, and the second part is the FIM with the prior density of the unknown vector to be estimated [47]. Specifically for the BFIM for the proposed algorithm based on (24), the BFIM can be written as

$$\mathbf{J}_B = \mathbf{J}_D + \mathbf{J}_P, \tag{42}$$

where

$$\mathbf{J}_D = -\mathbb{E}_{\mathbf{y},\mathbf{h}_b}\left\{\frac{\partial^2 \mathcal{L}(\mathbf{y}|\mathbf{h}_b)}{\partial \mathbf{h}_b \partial \mathbf{h}_b^H}\right\}, \tag{43}$$

and

$$\mathbf{J}_P = -\mathbb{E}_{\mathbf{h}_b}\left\{\frac{\partial^2 \mathcal{L}(\mathbf{h}_b;\mathbf{\Gamma})}{\partial \mathbf{h}_b \partial \mathbf{h}_b^H}\right\}. \tag{44}$$

By employing the results in [48] and ignoring the constant part in $\mathbf{J}_D$ and $\mathbf{J}_P$, we can derive that

$$\mathbf{J}_D = \mathbf{P}^H \mathbf{R}_q^{-1} \mathbf{P}, \tag{45}$$

and

$$\mathbf{J}_P = \mathbf{\Gamma}^{-1}, \tag{46}$$

where $\mathbf{\Gamma}$ is given by (32). Consequently, the BCRB for the MSE of the beamspace channel estimation can be evaluated as

$$\text{MSE}(\hat{\mathbf{h}}_b) \geq Tr\{\mathbf{J}_B^{-1}\} = Tr\{(\mathbf{P}^H \mathbf{R}_q^{-1} \mathbf{P} + \mathbf{\Gamma}^{-1})^{-1}\}. \tag{47}$$

Furthermore, upon substituting (33) into (47), we have

$$\text{MSE}(\hat{\mathbf{h}}) \geq Tr\{\mathbf{Q}(\mathbf{P}^H \mathbf{R}_q^{-1} \mathbb{P} + \Gamma^{-1})^{-1} \mathbf{Q}^H\}, \tag{48}$$

where $\mathbf{Q} = \mathbf{A}_T^* \otimes \mathbf{A}_R$.

IV. Simulation Results and Analysis

In this section, we present and analyze the performance results of the channel estimation schemes considered for the system model shown in Fig. 1 and the channel model described in (3). For comparison, the traditional single-layer SBL channel estimation [30], adaptive codebook (AC) channel estimation [12], AMP channel estimation [49], OMP channel estimation [40] and the BCRB of the multi-layer SBL [41] are used as benchmarkers. Additionally, in our simulations, we assume that there are 2 layers in the proposed multi-layer SBL algorithm, although more layers can be freely employed.

TABLE I
SIMULATION PARAMETERS

Parameters	Simulation value
Number of Paths	$L = 4$
Number of Streams	$N_s = 4$
Number of Transmit Antennas	$N_t = 8$
Number of Receive Antennas	$N_r = 8$
Number of transmitter RF chains	$N_{RF}^t = 4$
Number of Receiver RF chains	$N_{RF}^r = 4$

Furthermore, whenever needed by an algorithm, the initial precoder $\mathbf{F}_0$ and combiner $\mathbf{W}_0$ are constructed as:

$$\mathbf{F}_0(m,n) = \frac{1}{\sqrt{N_t}} e^{j\phi_{m,n}}, \tag{49}$$

$$\mathbf{W}_0(m,n) = \frac{1}{\sqrt{N_t}} e^{j\varphi_{m,n}}, \tag{50}$$

where $\phi_{m,n}, \varphi_{m,n} \in \mathcal{A}$, and $\mathcal{A}$ denotes the quantized angle set, which is assumed to have N_q-quantization bits, and can be expressed as

$$\mathcal{A} = \{0, \frac{\pi}{2^{N_q}}, \ldots, (2^{N_q}-1)\frac{\pi}{2^{N_q}}\}. \tag{51}$$

The details of simulation parameters are listed as Table I for convenience. Let us first consider the complexity.

A. Computational Complexity Analysis

The majority of calculations of the SBL algorithm is the processing shown in (27), (28), and (32). Therefore, the computational complexity of the above three steps can be used as a standard to measure to the complexity of the SBL algorithm. For the purpose of illustration, we assume that $G_T = 2G_{T,1}$ and $G_R = 2G_{R,1}$. Since the resolution should be larger than the antenna size, we set $G_{T,1} \geq \max\{N_t, N_r\}$ and $G_{R,1} \geq \max\{N_t, N_r\}$. The complexities of the single-layer and multi-layer SBL algorithms are listed in Table II, where $\mathbf{G}_s = [1, G_T G_R, (G_T G_R)^2, (G_T G_R)^3]^T$, $\mathbf{G}_1 = [1, G_{T,1}G_{R,1}, (G_{T,1}G_{R,1})^2, (G_{T,1}G_{R,1})^3]^T$, $\mathbf{G}_2 = [1, G_{T,2}G_{R,2}, (G_{T,2}G_{R,2})^2, (G_{T,2}G_{R,2})^3]^T$, and

$$\mathbf{P}_m = \begin{bmatrix} N_s N_m (N_t + N_s N_r + N_t N_r (2 + N_s N_m)) \\ N_t N_r (1 + N_s N_m) + N_s N_m N_i (1 + 3 N_s N_m) \\ (2 N_s N_m + 1) N_i \\ 2 N_i \end{bmatrix}^T, \tag{52}$$

$$\mathbf{P}_a = \begin{bmatrix} N_t N_m (N_s - 1) + N_s^2 N_m (N_r - 1) \\ N_s N_t N_r N_m + 2 N_s N_m N_i (N_s N_m - 1) \\ 2 N_s N_m N_i \\ 2 N_i \end{bmatrix}^T. \tag{53}$$

In the above matrices, N_i is the number of iterations used in the SBL algorithm, while $G_{T,2}$ and $G_{R,2}$ depend on the number of selected angles. Since mmWave channel is a sparse channel, the number of paths is usually smaller than that of the antennas of a massive MIMO system. Therefore, we can assume that $G_{T,2}(/G_{R,2}) \leq \min\{N_t, N_r\}$. As shown in Table II, $\mathbf{P}_m$ and $\mathbf{P}_a$ are the same for both the single-layer SBL and multi-layer SBL algorithm. Therefore, we can compare their complexity by comparing $\mathbf{G}_s$ and $\mathbf{G}_1 + \mathbf{G}_2$. Applying $G_T = 2G_{T,1}$ and $G_R = 2G_{R,1}$, we obtain

$$\mathbf{G}_s = \begin{bmatrix} 1 \\ 4G_{T,1}G_{R,1} \\ 16G_{T,1}^2 G_{R,1}^2 \\ 64G_{T,1}^3 G_{R,1}^3 \end{bmatrix}. \tag{54}$$

Therefore,

$$\mathbf{G}_s - (\mathbf{G}_1 + \mathbf{G}_2) = \begin{bmatrix} -1 \\ 3G_{T,1}G_{R,1} - G_{T,2}G_{R,2} \\ 15G_{T,1}^2 G_{R,1}^2 - (G_{T,2}G_{R,2})^2 \\ 63G_{T,1}^3 G_{R,1}^3 - (G_{T,2}G_{R,2})^3 \end{bmatrix}. \tag{55}$$

Since $G_{T,2}(/G_{R,2}) \leq \min\{N_t, N_r\} \leq \max\{N_t, N_r\} \leq \min\{G_{T,1}, G_{R,1}\}$, we can know that from [30], the second elements in the parameter matrix is larger than the first element. This implies that our proposed multi-layer SBL channel estimation algorithm has a significantly lower complexity than the single-layer SBL channel estimation algorithms.

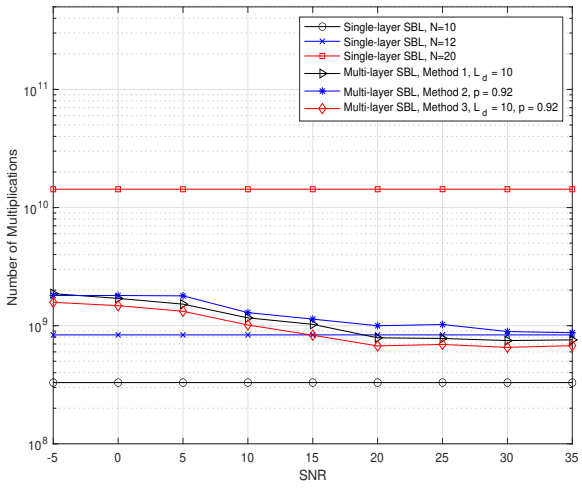

Fig. 5. The numbers of multiplications required by the single-layer SBL and multi-layer SBL for different methods.

Furthermore, since the multiplication dominates the computational complexity of the algorithms, to compare the computational complexity of our proposed algorithm with that of their benchmark algorithms, the number of multiplications versus SNR with respect to different sizes of the channel matrix is shown in Fig. 5, where the system parameters are set as follow: $G_T = G_R = 20$, $G_{T,1} = G_{R,1} = 10$, $N_s = 4$. Specifically, we compare the complexity of the multi-layer SBL algorithm associated with the Method 1, Method 2 and Method 3 in Section III-B after the first layer SBL. In Method 1, $L_d = 10$ is used in our simulation. In Method 2, we use 92% of the power as the threshold in our simulation. In Method 3, we combine the parameters of Method 1 and Method 2, where the selected beam should satisfy both the thresholds. As shown in Fig. 5, Method 3 shows the lowest complexity, when $N = 20$. The single-layer SBL algorithm has a complexity

TABLE II
COMPUTATIONAL COMPLEXITY ANALYSIS

Algorithms	Number of Multiplication	Number of Addition
Single-layer SBL	$\mathbf{P}_m\mathbf{G}_s$	$\mathbf{P}_a\mathbf{G}_s$
Multi-layer SBL	$\mathbf{P}_m(\mathbf{G}_1+\mathbf{G}_2)$	$\mathbf{P}_a(\mathbf{G}_1+\mathbf{G}_2)$

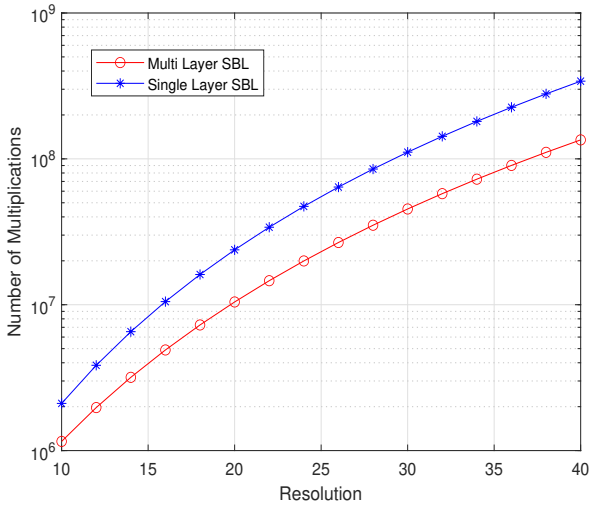

Fig. 6. The numbers of multiplications with different resolutions

that is independent of SNR. This is because the maximum number of iterations for EM algorithm is set to a fixed value. Fig. 5 also shows that the computational complexity of the proposed multi-layer SBL channel estimation algorithm is much lower than that of the single-layer SBL algorithm, when the resolutions N are the same. The complexity of the proposed algorithm with $N = 20$ is similar to that of the single-layer SBL with $N = 12$. In addition, the amount of computation of the proposed multi-layer SBL algorithm also reduces as SNR increases, as the result that the size of the sensing matrix reduces with the improvement of the estimation accuracy. Please note that in Fig. 5 we show the total number of multiplications needed for channel estimation, where this is accumulated over several processing steps as described in Section III.

On the other hand, based on [50], a multiply operation would require around 0.049 Watt. Hence, at SNR=20 dB, the "Multi-Layer SBL Method 1 with $L_d = 10$" requires 16% of the power consumption of the single-layer SBL with N=20, while the "Multi-Layer SBL Method 2 with p=0.92" requires 20% of the power consumption of the single-layer SBL with N=20. Additionally, the "Multi-Layer SBL Method 3 with $L_d = 10$ and p=0.92" requires 13% of the power consumption of the single-layer SBL with N=20.

In Fig. 6, we show the complexity comparison between the traditional single-layer SBL and the proposed multi-layer SBL algorithms, when the system parameters are set to $G_T = G_R = 32$, $G_{T,1} = G_{R,1} = 16$, $N_s = 8$. Explicitly, with the increase of resolution, the complexity of both algorithms increases but the gap between the two algorithms also increases. From this we can be implied that when a moderate to high resolution is employed, the computational complexity of the multi-layer SBL algorithm can be significantly lower than that of the single layer SBL algorithm.

B. MSE Performance

In the multi-layer SBL algorithm, the accuracy of the angles estimated in the first layer has a big impact on the performance of the algorithm, since the angle ranges used in the following layers depend on the estimated angles in the first layer. To reflect this, we define the angle coverage rate (CR) for the first layer angle estimation, which is expressed as

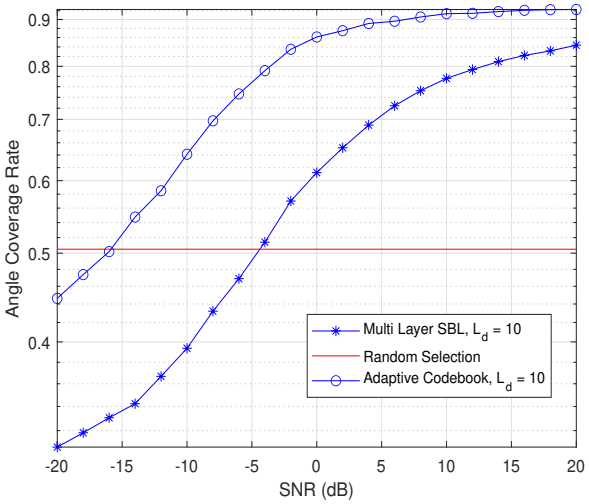

Fig. 7. Comparison of the angle coverage rate of the SBL algorithm, adaptive codebook, and random selection.

$$\text{CR} = \frac{size\{\mathbf{\Theta}_e \cap \mathbf{\Theta}_r\}}{size\{\mathbf{\Theta}_r\}}, \tag{56}$$

where $\mathbf{\Theta}_e$ and $\mathbf{\Theta}_r$ denote the estimated (selected) and actual angle array vectors. Fig. 7 shows the CR achieved by the proposed algorithm, adaptive codebook channel estimation and the random selection with respect to different SNR. As shown in Fig. 7, the CR of random selection does not change as the SNR increases. By contrast, the estimated angles by the SBL and adaptive codebook estimations become more accurate as SNR increases, reflected by that the CR increases as SNR increases. Furthermore, the accuracy of the estimated angles by the proposed algorithm is higher than that by the adaptive codebook channel estimation, when the same number of paths L_d are assumed. The adaptive codebook channel estimation is more accurate than the random selection, when SNR is

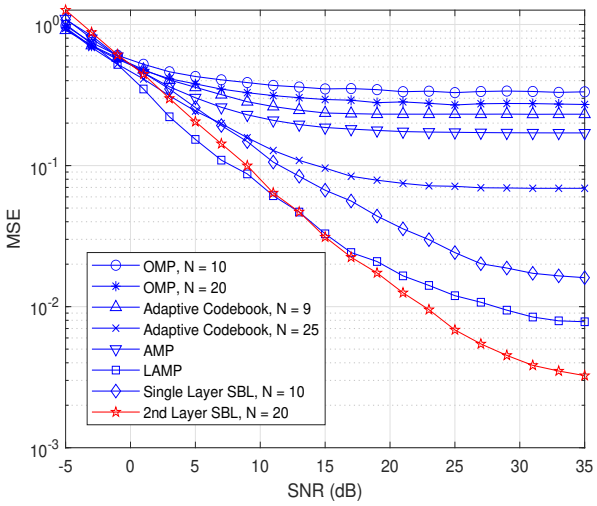

Fig. 8. MSE performance of different channel estimation algorithms with resolution $N = 10$ or 20 for single-layer SBL and OMP, $N = 20$ for multi-layer SBL ($N = 10$ in the first layer), and $N = 9$ or 25 for adaptive codebook channel estimation.

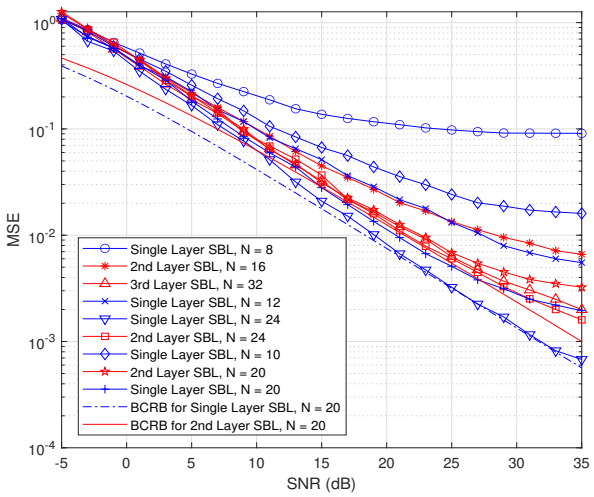

Fig. 9. MSE performance of multi-layer SBL channel estimation with different resolutions and number of layers.

higher than -6dB, while the proposed algorithm provides more accurate estimation than the random selection, provided that SNR is higher than -17dB. When SNR$>$ 0dB, the CR of the proposed algorithm is more than 85%. Hence, the angle regions provided to the next layer is highly reliable, which should improve the overall performance of the estimation.

Fig. 8 compares the MSE performance of different channel estimation techniques, showing that MSE reduces with the increase of resolution. The adaptive codebook channel estimation algorithm attains better MSE performance than the traditional OMP algorithm, AMP algorithm [40] and also the learned AMP (LAMP) algorithm [51]. The single-layer SBL algorithm performs about 10 times better than the adaptive codebook channel estimation with the same resolution, when SNR is large. By contrast, when SNR is very low, the single-layer SBL algorithm can be outperformed by the adaptive codebook channel estimation. As shown in Fig. 8, the proposed channel estimation algorithm is capable of achieving a lower MSE than the single-layer SBL algorithm, which also has a lower computational complexity than the single-layer SBL algorithm, as shown in Fig. 6. There are two reasons for this performance advantage. Firstly, the angle range used in the second layer is more accurate than that used in the single-layer SBL. Secondly, instead of the random precoder and combiner used in single-layer SBL algorithm, the precoder and combiner employed in the second layer are obtained based on the estimated channel in the first layer, which makes the sensing matrix more efficient.

In Fig. 9, we compare the results of the single-layer and multi-layer SBL algorithms with different number of layers. From Fig. 9 we can observe that as the number of layers increases, the estimation accuracy decreases. For example, the 3-layer SBL channel estimation with the resolutions of $N = 8,\ 16,\ 32$ shows a similar MSE performance as the 2-layer SBL channel estimation with the resolutions of $N = 10,\ 20,$ despite a significant increase in resolution. Furthermore, when given the number of layers, the performance gap between the multi-layer SBL channel estimation and the single-layer SBL channel estimation also widens, as the resolution increases. For instance, for the 2-layer SBL estimator, the performance gap between it and the single-layer SBL estimator at $N = 12,\ 24$ is larger than that at $N = 10,\ 20$. However, it is important to note that the MSE attained by the algorithm with a higher resolution is still lower than that with a lower resolution. When comparing the results with their BCRBs, the gap between MSE and BCRB is small at first until around SNR=25dB, which then becomes larger. This is because the MSE of channel estimation tends to converge, while the BCRB keeps decreasing. Additionally, as shown in Fig. 5, the computational complexity of the single-layer SBL algorithm with $N = 12$ and that of the proposed multi-layer SBL estimator with $N = 20$ are similar. Correspondingly, as shown in Fig. 8, when given $N = 12$, the MSE achieved by the proposed multi-layer estimator is only about 1/3 of that of the single-layer SBL. Therefore, the proposed algorithm is capable of achieving a better MSE performance than the single-layer SBL when both of them have a similar computational complexity.

If we define an adaptive codebook channel estimation combined SBL (AC-SBL) by using the adaptive codebook channel estimation in the first layer to obtain a tentative estimation and then using the SBL as the second layer, the AC-SBL can further reduce the complexity of the multi-layer SBL algorithm. This is because the SBL-assisted channel estimation is much more complex than the adaptive codebook channel estimation. Fig. 10 compares the MSE performance of the adaptive codebook channel estimation [12] and that of the AC-SBL channel estimation with a resolution of $N = 16$. As shown in Fig. 10, the MSE of the AC-SBL channel estimation is 5 times lower than that of the original adaptive codebook channel estimation at the SNR of 20dB. When the SNR is lower than 5dB, the performance of both is similar, as the

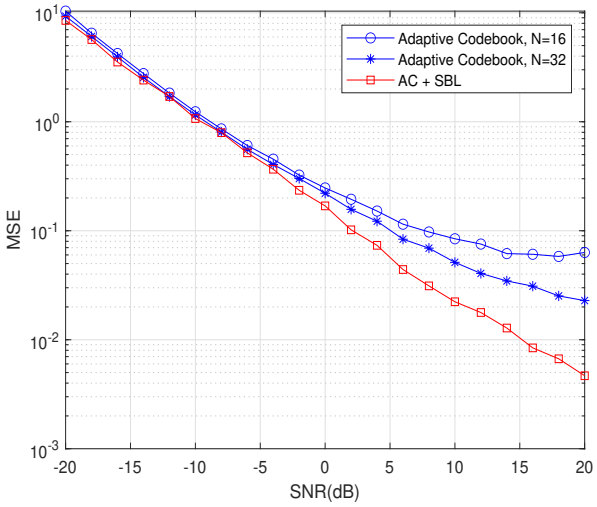

Fig. 10. MSE performance of the adaptive codebook channel estimation and the AC-SBL channel estimation with a resolution of $N = 16$.

result that the angle estimated by the adaptive codebook in high noise is not accurate. in the adaptive codebook algorithm. Therefore, the AC-SBL algorithm can be introduced to reduce the complexity of the multi-layer SBL algorithm in the relatively high SNR scenarios.

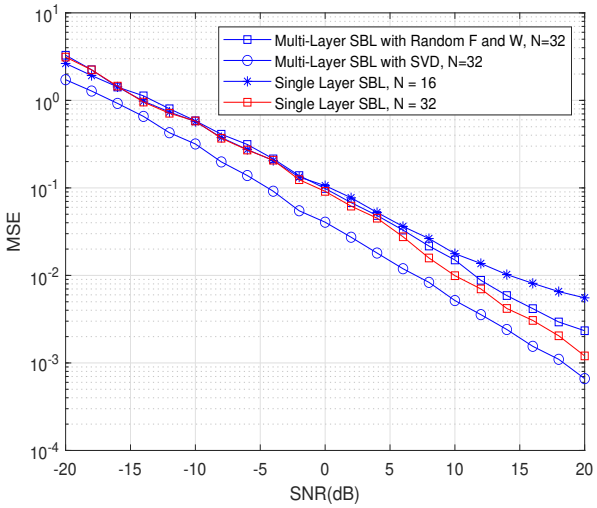

Fig. 11. MSE performance of channel estimation with different precoding and combining methods.

Finally, in Fig. 11, we demonstrate the MSE performance of the single-layer SBL channel estimation and the multi-layer SBL channel estimation with different precoding and combining methods. As shown in Fig. 11, the MSE performance of the multi-layer SBL algorithm with random precoder and combiner is capable of approaching the performance of the single-layer SBL algorithm with the same resolution, while requiring significantly lower computational complexity, as shown previously. By contract, the MSE performance of the multi-layer SBL with the precoder and combiner designed based on SVD of the estimated channel in the first layer is better than the MSE performance attained by the other considered schemes. Hence, the precoder and combiner as designed are more efficient, as they make the sensing matrix to be used in the next layer more efficient in signal recovery.

V. Conclusion

By taking the advantages of both the OMP-based algorithms and the SBL-based algorithms while circumventing their shortcomings, this paper proposed and investigated a multi-layer SBL algorithm in order to achieve the channel estimation of high-accuracy and low-complexity. Owing to the multi-layer structure, each layer of the multi-layer SBL algorithm only needs to deal with a small number of codewords, which reduces the overall complexity of the algorithm. Simultaneously, high-accuracy can be achieved by the algorithm via increasing the resolution layer-by-layer. Furthermore, the channel estimation algorithms for different layers can be flexibly implemented by, such as OMP- or SBL-based algorithms, so as to strike a best trade-off between complexity and performance. In this paper, the proposed algorithm was compared with a range of benchmark algorithms. It is shown that the complexity of the proposed algorithm can be 3 times lower than the traditional single-layer SBL algorithm, when both of them are operated at the same resolution and also attain a similar accuracy of channel estimation. The MSE of channel estimation achieved by the proposed algorithm is about 10 times less than that obtained by the OMP-based algorithm. Furthermore, in terms of the design of precoder/combiner, in comparison with the random precoder/combiner, the precoder/combiner designed based on the SVD of estimated channel allows our proposed multi-layer SBL algorithm to achieve much better performance, which is the closest one to the BCRB.

References

[1] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts*, vol. 20, no. 2, pp. 870–913, 2017.

[2] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts*, vol. 20, no. 2, pp. 836–869, 2017.

[3] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, "Software defined mobile networks: concept, survey, and research directions," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 126–133, 2015.

[4] M. Series, "Imt vision–framework and overall objectives of the future development of imt for 2020 and beyond," *Recommendation ITU*, vol. 2083, p. 0, 2015.

[5] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5g: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts*, vol. 20, no. 4, pp. 3060–3097, 2018.

[6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[7] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[8] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

[9] R. W. Heath, N. Gonzalez-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, 2016.

[10] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, 2011.

[11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *2013 Information Theory and Applications Workshop (ITA)*, pp. 1–5, IEEE, 2013.

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

[13] L.-L. Yang, *Multicarrier communications*. John Wiley & Sons, 2009.

[14] S. Katla, L. Xiang, Y. Zhang, M. El-Hajjar, A. A. Mourad, and L. Hanzo, "Deep learning assisted detection for index modulation aided mmwave systems," *IEEE Access*, vol. 8, pp. 202738–202754, 2020.

[15] H. Liu, S. Lu, M. El-Hajjar, and L.-L. Yang, "Machine learning assisted adaptive index modulation for mmwave communications," *IEEE open j. Commun. Soc.*, vol. 1, pp. 1425–1441, 2020.

[16] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. Veh. Technol*, vol. 68, no. 2, pp. 1394–1404, 2019.

[17] T. Wang, F. Long, T. Ma, L. Gao, Y. Jiang, and L. Chang, "Low-complexity matrix-based conjugate gradient channel estimation for cooperative wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4078–4083, 2019.

[18] F. Talaei and X. Dong, "Hybrid mmwave mimo-ofdm channel estimation based on the multi-band sparse structure of channel," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1018–1030, 2018.

[19] X. Gao, L. Dai, S. Han, I. Chih-Lin, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive mimo systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, 2017.

[20] J. Tan and L. Dai, "Wideband channel estimation for thz massive mimo," *China Communications*, vol. 18, no. 5, pp. 66–80, 2021.

[21] 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, 2007.

[22] L. Wei, C. Huang, Q. Guo, Z. Yang, Z. Zhang, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Joint channel estimation and signal recovery for ris-empowered multiuser communications," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4640–4655, 2022.

[23] X. Gan, C. Zhong, C. Huang, and Z. Zhang, "RIS-Assisted Multi-User MISO Communications Exploiting Statistical CSI," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6781–6792, 2021.

[24] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel Estimation for RIS-Empowered Multi-User MISO Wireless Communications," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4144–4157, 2021.

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

[26] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave mimo systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, 2018.

[27] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[28] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[29] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[30] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[31] K. Liu, X. Li, J. Fang, and H. Li, "Bayesian mmwave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Access*, vol. 7, pp. 48961–48970, 2019.

[32] X. Shao, L. Cheng, X. Chen, C. Huang, and D. W. K. Ng, "Reconfigurable intelligent surface-aided 6g massive access: Coupled tensor modeling and sparse bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10145–10161, 2022.

[33] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "Dnn-aided block sparse bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12000–12012, 2019.

[34] H. Tang, J. Wang, and L. He, "Off-grid sparse bayesian learning-based channel estimation for mmwave massive mimo uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, 2018.

[35] H. Liu, Y. Zhang, X. Zhang, M. El-Hajjar, and L.-L. Yang, "Deep learning assisted adaptive index modulation for mmwave communications with channel estimation," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2022.

[36] Z. Han, H. Li, and W. Yin, *Compressive sensing for wireless networks*. Cambridge University Press, 2013.

[37] Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, 2018.

[38] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive mimo with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, 2016.

[39] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[40] X. Wei, C. Hu, and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive mimo systems," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 182–193, 2021.

[41] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and G. Sharma, "Sparse bayesian learning (sbl)-based frequency-selective channel estimation for millimeter wave hybrid mimo systems," in *2019 National Conf. Commun. (NCC)*, pp. 1–6, 2019.

[42] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in mimo-ofdm systems: A sparse bayesian learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5369–5382, 2015.

[43] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the em algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 300–308, 1997.

[44] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.

[45] A. Mishra, N. Yashaswini, and A. K. Jagannatham, "Sbl-based joint sparse channel estimation and maximum likelihood symbol detection in ostbc mimo-ofdm systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4220–4232, 2018.

[46] A. Mishra, A. K. Jagannatham, and L. Hanzo, "Sparse bayesian learning-aided joint sparse channel estimation and ml sequence detection in space-time trellis coded mimo-ofdm systems," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1132–1145, 2020.

[47] B. Vucetic and J. Yuan, *Space-time coding*. John Wiley & Sons, 2003.

[48] H. Hijazi and L. Ros, "Bayesian cramer-rao bound for ofdm rapidly time-varying channel complex gains estimation," in *2009 IEEE Global Telecom. Conf. (GLOBECOM 2009)*, pp. 1–6, 2009.

[49] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE inf. theory workshop (ITW 2010)*, pp. 1–5, 2010.

[50] M. Hemnani, S. Palekar, P. Dixit, and P. Joshi, "Hardware optimization of complex multiplication scheme for dsp application," in *2015 International Conference on Computer, Communication and Control (IC4)*, pp. 1–4, 2015.

[51] M. Borgerding, P. Schniter, and S. Rangan, "Amp-inspired deep networks for sparse linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4293–4308, 2017.