Deep Learning Assisted Adaptive Index Modulation for mmWave Communications with Channel Estimation

Haochen Liu, Yaoyuan Zhang, Xiaoyu Zhang, *Senior Member, IEEE*,
Mohammed El-Hajjar, *Senior Member, IEEE*,
Lie-Liang Yang, *Fellow, IEEE*

***Abstract*—The efficiency of link adaptation in wireless communications relies greatly on the accuracy of channel knowledge and transmission mode selection. In this paper, a novel deep learning based link adaptation framework is proposed for the orthogonal frequency-division multiplexing (OFDM) systems with compressed-sensing-assisted index modulation, termed as OFDM-CSIM, communicating over millimeter-wave (mmWave) channels. To achieve link adaptation, a novel multi-layer sparse Bayesian learning (SBL) algorithm is proposed for accurately and instantaneously providing the required channel state information. Meanwhile, a deep neural networks (DNN)-assisted adaptive modulation algorithm is proposed to choose the best possible transmission mode to maximize the achievable throughput. Simulation results show that the proposed multi-layer SBL algorithm enables more accurate channel estimation than the conventional techniques. The DNN-based adaptive modulator is capable of achieving a higher throughput than the learning-assisted solution based on the k nearest neighbor (k-NN) algorithm, and also the classic average signal-to-noise ratio (SNR)-based solutions. Moreover, analysis shows that both the multi-layer SBL algorithm and the DNN-assisted adaptive modulator achieve better performance than their respective conventional counterparts while at a significantly lower computational complexity cost.**

***Index Terms*—OFDM, mmWave, sparse Bayesian learning, channel estimation, adaptive modulation, machine learning, neural networks.**

I. INTRODUCTION

MILLIMETER-wave (mmWave) communication has been actively studied and considered for the standardization in the next generation wireless systems, owing to its potential to meet the continued growing wireless capacity demands in the numerous wireless applications [1]. Due to its short wavelength, which ranges from 1 mm to 10 mm, mmWave signals suffer from substantial path-loss and shadowing losses [2]. However, the mmWave's short wavelength allows to implement a large number of antennas in a relatively small area on both transmitter and receiver sides. Consequently, advanced beamforming techniques can be utilized for overcoming the challenges generated by the propagation characteristics [3]. Furthermore, given the large available bandwidth at mmWave frequencies, the mmWave channels are usually frequency selective fading channels [1, 2]. Correspondingly, OFDM-based techniques have been considered as the typical signalling techniques for operation with mmWave communications, according to the 5G New Radio [4, 5].

On the other hand, index modulation (IM) has been introduced as a promising modulation scheme to multi-carrier systems including OFDM, as it is capable of providing additional degrees of freedom for data modulation, and the flexibility to strike a good trade-off between spectral efficiency (SE) and energy efficiency (EE) [6, 7]. In literature, OFDM with subcarrier IM (OFDM-IM) was proposed in [6], which was further developed in [8–12]. In OFDM-IM [9], information is conveyed by both the conventional amplitude phase modulation (APM) and the indices of subcarriers. In contrast to the conventional OFDM, OFDM-IM can achieve better error performance [9], and also attain a more flexible design trade-off among SE, EE and complexity [6, 13]. The authors of [14] proposed two enhanced OFDM-IM schemes. One scheme extended IM to a new dimension by employing both in-phase and quadrature components, which allows to further increase spectral efficiency. With the aid of linear constellation precoding, the other scheme spreads information symbols over two adjacent active subcarriers to convey information symbols over two subcarriers, which achieves additional diversity gain. Furthermore, by assuming the same number of IM bits, the scheme in [15] proposes to utilize different constellation sizes on different subcarriers, which further increases the flexibility of IM. Moreover, to improve the flexibility of OFDM-IM, the compressed sensing (CS) concept can be introduced to form a virtual index space for IM operation [13], forming the so-called OFDM with CS-assisted IM (OFDM-CSIM). This arrangement enables OFDM-IM to benefit from CS for further improving the trade-off between SE and EE [13, 16]. Additionally, by introducing appropriate interleaving, the OFDM-CSIM is also capable of attaining an improved performance, when compared with the OFDM-IM [13, 17].

The financial support of the Engineering and Physical Sciences Research Council (EPSRC) and the Royal Academy of Engineering is gratefully acknowledged.

Haochen Liu, Yaoyuan Zhang, Mohammed El-Hajjar and Lie-Liang Yang are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom (email:{hl14n17,yz2m19}@soton.ac.uk and {meh,lly}@ecs.soton.ac.uk).

Xiaoyu Zhang is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with the Optoelectronics Research Centre, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: xy.zhang@soton.ac.uk).

Simulation data is available here: https://doi.org/10.5258/SOTON/D2250.

It is widely acknowledged that the time-varying nature of wireless channels is among the most challenging aspects of wireless system design [2]. To mitigate its effect, link adaptation was proposed as a powerful technique to optimize the SE [17]. Conventionally, link adaptation is achieved via adaptive modulation and coding (AMC) [18], which selects the appropriate modulation order and coding rate to maximize throughput, while maintaining a required reliability [18–21]. In the conventional link adaptation, transmission modes are determined according to the thresholds specified based on the channel statistics [22–24]. However, the thresholds in the conventional link adaptation may not be optimal, due to the various impairments imposed by wireless systems, including the time-varying channels, the non-linearity of amplifiers, the transmission frequency instability, *etc.* [25]. By contrast, machine learning (ML)-based approaches treat the physical layer as the transfer between system state and data observations [26], which have the potential to overcome the challenges faced by the conventional link adaptation techniques. Owing to this, the application of ML for link adaptation in wireless communications has drawn good attention [20, 23, 27]. For example, a supervised learning algorithm was proposed in [17, 20], where the observation data correlated with the conditions are directly used to select the transmission modes.

Link adaptation can be based on the signal-to-noise ratio (SNR) information in both the conventional and learning-aided methods. Hence, accurate estimation of SNR is important for the implementation of link adaptation in wireless communications. To achieve this, a high-performance channel estimator is critical for achieving good performance [20]. Specially in mmWave communications, the spatial sparsity can be effectively exploited to estimate the channel coefficients of the significant spatial paths. Based on this observation, the sparse mmWave channels can be estimated with a considerably decreased pilot overhead using the CS-based methods, such as the sparse Bayesian learning (SBL) [28–30], orthogonal matching pursuit (OMP) [31, 32], approximate message passing (AMP) [33], etc. For the SBL and OMP methods, the angles of arrival (AoA) and angles of departure (AoD) spaces are divided into a finite number of grid-points to obtain the beamspace representation of mmWave channels. In [34], the mmwave channel estimation problem was formulated as a structured sparse signal recovery problem with a *priori* channel knowledge determined by the channel characteristics. To further improve the performance of the SBL-based channel estimation, the authors of [29] developed an expectation maximization (EM)-based SBL channel estimator. However, it has high computational complexity, which is challenging, especially when a high angular resolution is needed. By contrast, the AMP approach applies the quadratic and Taylor series approximations to loopy the belief propagation and achieves low complexity channel estimation [33]. Furthermore, based on the original AMP, a generalized AMP (GAMP) algorithm was proposed in [35] to reconstruct signals from nonlinear measurements. The studies show that although the Gaussian prior and likelihood case are not sufficient for dealing directly with the sparse signal recovery problem, it is effective to use them to replace the inversion step of the standard EM-based implementation of the SBL [36]. Thus, a GAMP-based low complexity SBL algorithm was proposed in [36], which can address the convergence limitations of the AMP, while reducing the complexity of the SBL. Nevertheless, for the link adaptation over time-varying wireless channels, not only accurate but also low-complexity channel estimation techniques are demanded, *so that the estimated SNR can be fed back to transmitter in a timely fashion to achieve the best possible performance*.

Given the above background, in this contribution, the benefits of adaptive transmission and ML are amalgamated for the proposed OFDM-CSIM system. In order to realize the full potential of adaptive systems, accurate and efficient estimation of channel state information (CSI) is critical. Hence, a low-complexity SBL-based channel estimation method is proposed for the adaptive OFDM-CSIM system communicating over mmWave channels. Explicitly, we first introduce the OFDM-CSIM system, which takes the advantages of both IM and CS, in order to attain a good trade-off between SE and EE. Then, a multi-layer SBL algorithm is developed for the OFDM-CSIM system operated in mmWave channels. Finally, the novel adaptive OFDM-CSIM system employing the proposed multi-layer SBL channel channel estimation is investigated, in conjunction with the conventional and ML-assisted adaptation, respectively. In our previous research [17], perfect channel estimation was assumed and the supervised-learning classifier was employed to obtain a statistically-consistent solution. To further improve the system performance, in this paper, a novel DNN-based multi-label classifier is proposed for the OFDM-CSIM system. Moreover, we consider the practical scenario, where the CSI is estimated by an enhanced SBL algorithm with low complexity, so that the performance of link adaptation can be guaranteed. To summarize, our novel contributions are as follows.

- We intrinsically amalgamate the concept of adaptive modulation with the OFDM-CSIM system communicating over mmWave channels. In our adaptive OFDM-CSIM system, the modulation order and the number of active subcarriers are adaptively adjusted according to the communication conditions, with the objective of maximizing the system's throughput at a constrained target bit error ratio (BER). The main advantage of our proposed work, as compared to the state-of-the-art ones, such as that in [18, 22], is the attainable high flexibility in system design, which allows to strike a flexible trade-off among SE, EE and complexity.
- To realize the potential of the proposed adaptive system, we propose a multi-layer SBL channel estimation method for our OFDM-CSIM mmWave system, which has lower computational complexity than the conventional SBL-based algorithm. In addition, the proposed multi-layer SBL channel estimation method enables high degrees of freedom in system design, allowing for attaining flexible trade-off between estimation performance and computational complexity.
- We conceive a novel DNN-based multi-label classifier for the proposed adaptive OFDM-CSIM system, which

is shown to be capable of choosing better transmission modes than the conventional and also the k-NN-based adaptive schemes. After proper training, the DNN-based classifier enables a significant improvement of throughput, while maintaining the required target BER performance.

The rest of the paper is organized as follows. Section II introduces the concept of the OFDM-CSIM system employed for mmWave communications along with the system model. In Section III, the proposed multi-layer SBL approach is introduced, along with the Bayesian Cramr-Rao lower bound (BCRB) analysis to characterize the efficiency of the proposed channel estimation schemes. In Section IV, both the conventional adaptive modulation and the ML-assisted adaptive modulation are addressed. In Section V, the simulation results are presented and analyzed. Finally, the conclusions and some suggestions for future research are provided in Section VI.

Notations: Lower-case boldface letter $\boldsymbol{a}$ and upper-case boldface letter $\boldsymbol{A}$ denote vectors and matrices, respectively; $[\cdot]_{\rm i}$ denotes the i-th element of a vector; $(\cdot)^{\rm T}$ and $(\cdot)^{\rm H}$ express transpose and conjugate transpose operations, respectively; Λ denotes complex signal symbol; $\mathbb{C}^{A\times B}$ is the set of $(A\times B)$-element in the complex field; $\{\cdot\}_a^b$ is a sequence with the indices from a to b; $\mathbb{E}[\cdot]$ is the expectation operator; $\mathrm{diag}(\boldsymbol{a})$ expresses a diagonal matrix formed from vector $\boldsymbol{a}$; $\mathrm{Tr}(\cdot)$ and $|\cdot|$ represent trace and absolute value, respectively; $\binom{n}{k}$ is the combination of the selection of k items from a collection of n items; $\|\cdot\|_{\rm F}$ is the Frobenius norm; $\langle\cdot\rangle$ expresses the inner product operation; $\lfloor\cdot\rfloor$ returns the largest integer less than the value.

II. System Model

In this section, we introduce the adaptive OFDM-CSIM system operating in mmWave environment.

Fig. 1 shows the block diagram of the transmitter with hybrid analog and digital precoder. The information bit sequence input to the transmitter is firstly split into G groups as shown in Fig. 1. The bits for each group is further split into multiple parts, with one part used to implement the IM $\boldsymbol{i}_g$, whilst the other part is used to form the amplitude-phase modulation (APM) symbols, such as PSK or QAM, to form $\boldsymbol{s}_g$. Afterwards, as shown in Fig. 1, a virtual[1] OFDM block $\boldsymbol{X}$ is generated by combining the outputs of the IMs and QAM symbols. Then this virtual OFDM block is compressed from the virtual domain (high dimension) to the frequency domain (low dimension) with the aid of a CS matrix $\boldsymbol{A}_{\rm CS}$. Then, a digital precoder is applied to process the transmit OFDM block. Then the signals are transformed to the time domain by imposing the inverse fast Fourier transforms (IFFT) followed by inserting the cyclic prefixes (CP). Finally, as shown in Fig. 1, analog beamforming is implemented before the signals are transmitted from the transmit antennas. As shown in Fig. 1, an adaptive control part is needed to achieve adaptive modulation, for which the receiver uses the signal-to-noise ratio (SNR) obtained from multi-layer SBL algorithm to select a most suitable transmission mode so as to maximize the achievable SE of the system. The details of the transmitter will be detailed in Section IV.

The block diagram of the receiver is shown in Fig. 2, where the received signal is firstly processed by an analog combiner. Then, after CP removal and the fast Fourier transform (FFT) processing, the signal is input to the digital combiner in the frequency domain. The compressed OFDM blocks $\boldsymbol{Y}_m$, given by the outputs of the digital combiner, are divided into several sub-blocks $\boldsymbol{Y}_g$ for signal detection. Let us describe in detail the operations in the transmitter and receiver in Fig. 1 and Fig. 2, respectively.

A. Virtual Domain Processing

Consider an OFDM-based mmWave system employing $N_{\rm t}$ transmit antennas with $N_{\rm t}^{\rm RF}$ RF chains and $N_{\rm r}$ receive antennas with $N_{\rm r}^{\rm RF}$ RF chains, which supports $N_{\rm s}$ data streams on each subcarrier. For each data stream, the input bit sequence of length $(p_1+p_2)G$ bits is firstly split into G groups with each group having (p_1+p_2) bits in the virtual domain. Within each group, p_1 bits are used to select the active virtual indices for transmitting the QAM symbols, which convey in total p_2 bits per symbol period. We assume that there are a total $N_{\rm v}$ number of indices in the virtual domain and each group occupies $N_{\rm g}$ indices, hence $N_{\rm v}=GN_{\rm g}$. Then, based on IM, in each virtual group, $N_{\rm a}$ indices are selected from the $N_{\rm g}$ to modulate the $N_{\rm a}$ QAM symbols. Thus $p_1=\left\lfloor \log_2\left[\binom{N_{\rm g}}{N_{\rm a}}\right]\right\rfloor$. After the IM, the active virtual indices of the G groups are represented by a set as

$$\mathcal{S}_{\rm index}=\{\boldsymbol{i}_1,\boldsymbol{i}_2,\ldots,\boldsymbol{i}_{\rm g},\ldots,\boldsymbol{i}_G\}, \tag{1}$$

for the $N_{\rm s}$ data streams, where $\boldsymbol{i}_{\rm g}\in\mathbb{Z}^{N_{\rm g}\times 1}$ is an integer matrix containing only '0' and '1' elements, with '1' denoting an active index and '0' denoting an inactive index. Hence, there are only $N_{\rm a}$ '1's in $\boldsymbol{i}_{\rm g}$. Then, in each virtual group, $p_2=N_{\rm a}(\log_2(Q))$ bits are used to generate the APM symbols to be associated with the active indices, where Q is the order of the APM. The APM symbols of G groups are expressed by a set as

$$\mathcal{S}_{\rm APM}=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_{\rm g},\ldots,\boldsymbol{s}_G\}, \tag{2}$$

where $\boldsymbol{s}_{\rm g}=[\boldsymbol{s}_{{\rm g},1},\ldots,\boldsymbol{s}_{{\rm g},N_{\rm a}}]^{\rm T}\in\mathbb{C}^{N_{\rm a}\times 1}$ contains the APM symbols of the g-th group, which are conveyed by the virtual active indices of the g-th group.

Afterwards, the virtual OFDM block is generated by mapping the G groups of the APM symbols in (2) to the active virtual indices in (1). Specifically, the APM symbols in $\boldsymbol{s}_{\rm g}$ are mapped to the virtual active indices represented by the '1' elements in $\boldsymbol{i}_{\rm g}$. The resultant index modulated symbol of group g can be expressed as

$$\boldsymbol{x}_{g,n_{\rm s}}=[0,\ldots,s_{g,1},\ldots,0,\ldots,s_{g,N_{\rm a}},\ldots,0]^{\rm T}, \tag{3}$$

where $\boldsymbol{x}_{g,n_{\rm s}}\in\mathbb{C}^{N_{\rm g}\times 1}$, $n_{\rm s}=1,\cdots,N_{\rm s}$, $N_{\rm s}$ is the number of data streams transmitted in spatial domain. Owing to the employment of transmit antenna array, we assume that each active subcarrier conveys $N_{\rm s}$ APM symbols and each data

[1]Virtual domain and frequency domain are used to differentiate the signal before and after the CS [17].

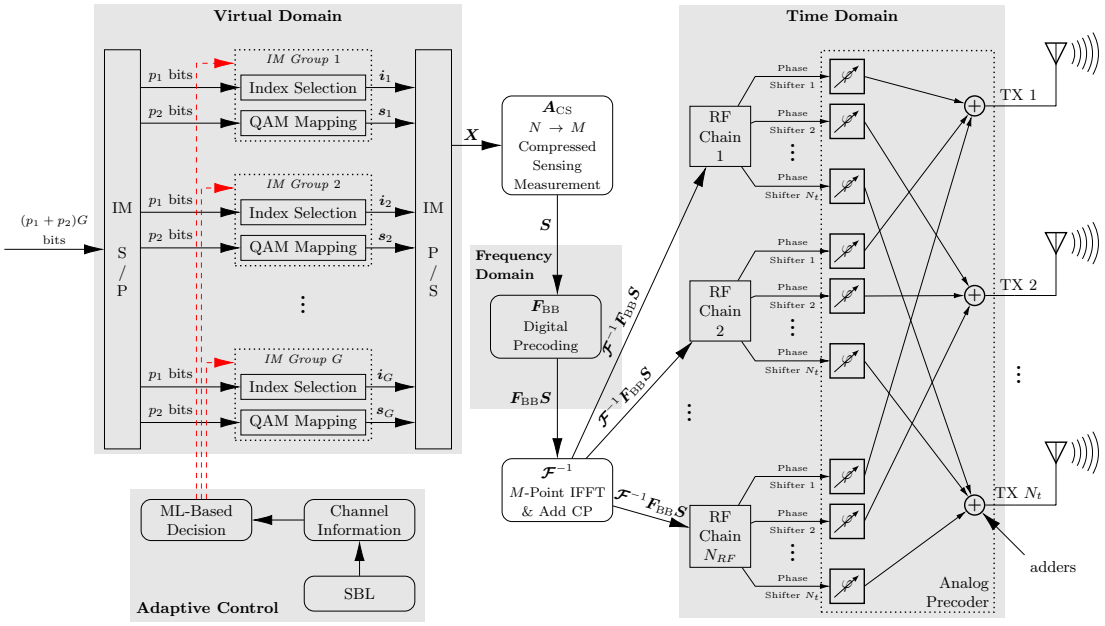


Fig. 1. Block diagram of the adaptive modulation-aided OFDM-CSIM transmitter with hybrid precoder.

stream is able to activate different virtual subcarriers of a group. Finally, the virtual OFDM block can be expressed as

$$\boldsymbol{X}_{\mathrm{VD}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_g, \ldots, \hat{\boldsymbol{x}}_G]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{v}} \times N_{\mathrm{s}}}, \tag{4}$$

where $\hat{\boldsymbol{x}}_g = [\boldsymbol{x}_{g,1}, \ldots, \boldsymbol{x}_{g,N_{\mathrm{s}}}]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{s}} \times N_g}$. Based on the above discussion, there are in total $N_{\mathrm{s}} N_{\mathrm{a}} G$ QAM symbols transmitted in one OFDM symbol period. Note that, the numbers of active virtual indices in different groups can be different. However, in this paper, we assume a fixed number of active virtual indices in each group for simplicity.

B. Compressed Sensing

As shown in Fig. 1, the CS measurement matrix $\boldsymbol{A}_{\mathrm{CS}} \in \mathbb{C}^{M_{\mathrm{f}} \times N_{\mathrm{v}}}$ is applied to compress $\boldsymbol{X}_{\mathrm{VD}}$ from N_{v}-dimension in the virtual domain to M_{f}-dimension in the frequency domain, where $M_{\mathrm{f}} < N_{\mathrm{v}}$. Hence, the matrix $\boldsymbol{X}_{\mathrm{VD}}$ should be a sparse matrix in order to benefit from the CS. The compressed matrix in the frequency domain can be expressed as

$$\boldsymbol{S}_{\mathrm{F}} = \boldsymbol{A}_{\mathrm{CS}} \boldsymbol{X}_{\mathrm{VD}}, \tag{5}$$

where $\boldsymbol{S}_{\mathrm{F}} \in \mathbb{C}^{M_{\mathrm{f}} \times N_{\mathrm{s}}}$, $\boldsymbol{S}_{\mathrm{F}} = [\boldsymbol{s}[1], \ldots, \boldsymbol{s}[c], \ldots, \boldsymbol{s}[M_{\mathrm{f}}]]^{\mathrm{T}}$, $\boldsymbol{s}[c] \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ contains the N_{s} data symbols transmitted by the c-th subcarrier.

Note that in the considered OFDM-CSIM system, the N_{v} virtual indices are divided into G groups, with each group containing N_{g} indices and $N_{\mathrm{v}} = N_{\mathrm{g}} G$. When compared with the conventional IM system, which uses no virtual domain operations, *i.e.* $M_{\mathrm{f}} = N_{\mathrm{v}}$, a higher throughput can be achieved by the OFDM-CSIM system [13, 17]. Furthermore, in comparison with the conventional OFDM-IM system, the improved throughput of OFDM-CSIM system is obtained without consuming additional power or bandwidth.

However, the sparse matrix $\boldsymbol{X}_{\mathrm{VD}}$ should be carefully designed to satisfy the strict sparsity level of N_{a}, *i.e.*, $N_{\mathrm{a}} \ll N_{\mathrm{v}}$, in order to achieve a good recovery performance at the receiver side. Moreover, the CS measurement matrix $\boldsymbol{A}_{\mathrm{CS}}$ plays an important role when recovery performance is considered, which should be designed to satisfy the mutual incoherence property (MIP) [13]. In general, the columns of $\boldsymbol{A}_{\mathrm{CS}}$ should be as uncorrelated as possible. According to [13], $\boldsymbol{A}_{\mathrm{CS}}$ can be generated by constructing a sub-matrix from an orthonormal dictionary, which can guarantee the efficiency of information recovery [13]. The quality of $\boldsymbol{A}_{\mathrm{CS}}$ can be evaluated by its mutual coherence $\mu(\boldsymbol{A}_{\mathrm{CS}})$, given as:

$$\mu(\boldsymbol{A}_{\mathrm{CS}}) = \max_{i \neq j} \frac{|\langle \boldsymbol{A}_{\mathrm{CS},i}, \boldsymbol{A}_{\mathrm{CS},j} \rangle|}{\|\boldsymbol{A}_{\mathrm{CS},i}\|_F \|\boldsymbol{A}_{\mathrm{CS},j}\|_F}, \tag{6}$$

where i and j indicate the different columns in $\boldsymbol{A}_{\mathrm{CS}}$. To provide a good recovery performance, $\mu(\boldsymbol{A}_{\mathrm{CS}})$ should satisfy the following specific conditions [13]:

$$\sqrt{\frac{M_{\mathrm{f}} - N_{\mathrm{v}}}{N_{\mathrm{v}}(M_{\mathrm{f}} - 1)}} \leq \mu(\boldsymbol{A}_{\mathrm{CS}}) \leq \frac{1}{2N_{\mathrm{a}} - 1}. \tag{7}$$

C. Frequency and Time Domain Processing

As shown in Fig. 1, the N_s transmitted symbols $\boldsymbol{s}[c] \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ on the c-th subcarrier, $c = 1, \ldots, M_{\mathrm{f}}$, are precoded by a digital precoder $\boldsymbol{F}_{\mathrm{D}}[c] \in \mathbb{C}^{N_{\mathrm{t}}^{\mathrm{RF}} \times N_s}$. Then, the M_{f} symbols corresponding to one data stream are transformed to the time-domain using the M_{f}-point IFFTs. After adding the CP and processing by analog precoder $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}^{RF}}$ common to all subcarriers, the final signal transmitted on the c-th subcarrier can be represented in a baseband equivalent form as

$$\boldsymbol{u}[c] = \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_D[c] \boldsymbol{s}[c] = \boldsymbol{F}[c] \boldsymbol{s}[c], \tag{8}$$

where $\mathbb{E}\left[||\boldsymbol{s}[c]||^2\right] = N_s$, $\boldsymbol{F}[c] = \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_D[c]$. Since the analog part of the hybrid beamforming is implemented by analog phase shifters, the elements in $\boldsymbol{F}_{\mathrm{RF}}$ are constrained with constant magnitude. The transmitted power is constrained by $\boldsymbol{F}_{\mathrm{D}}[c]$ so that $\|\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{D}}[c]\|_F^2 = N_s$. Assuming a block-fading channel model, the received signal from the c-th subcarrier is:

$$\boldsymbol{y}[c] = \boldsymbol{H}[c] \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{D}}[c] \boldsymbol{s}[c] + \boldsymbol{n}[c], c = 1, 2, \cdots, M_{\mathrm{f}}, \tag{9}$$

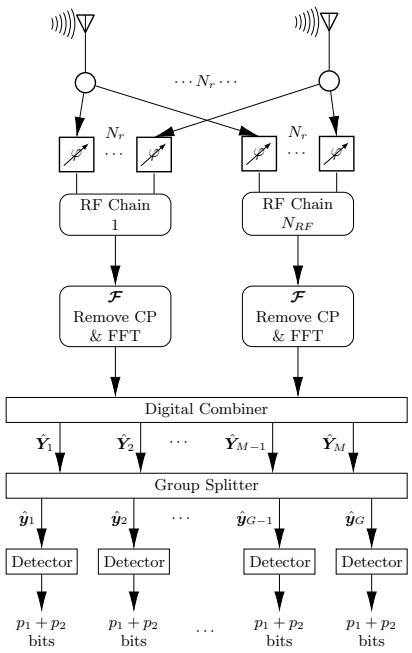


Fig. 2. Block diagram of the receiver.

where $\boldsymbol{H}[c] \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ is the mmWave channel matrix between transmitter and receiver, which will be presented later in this section, and $\boldsymbol{n}[c] \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \mathbf{I}_{N_s})$ is the vector of additive white Gaussian noise (AWGN).

At the receiver side shown in Fig. 2, the received signal of (9) is first processed by an analog combiner $\boldsymbol{W}_\mathrm{RF} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{RF}^\mathrm{r}}$, which is identical for all subcarriers with the elements of $\boldsymbol{W}_\mathrm{RF}$ constrained as in the analog precoder. Afterwards, the cyclic prefix is removed and M_f-point fast Fourier transform (FFT) is applied to recover the frequency domain signals. Then, the receiver applies the digital combiners $\boldsymbol{W}_\mathrm{D}[c] \in \mathbb{C}^{N_\mathrm{RF}^\mathrm{r} \times N_s}$ to generate the decision variables represented as:

$$\hat{\boldsymbol{y}}[c] = \boldsymbol{W}^\mathrm{H}[c]\boldsymbol{H}[c]\boldsymbol{F}[c]\boldsymbol{s}[c] + \boldsymbol{W}^\mathrm{H}[c]\boldsymbol{n}[c], \tag{10}$$

where $c = 1, 2, \cdots, M_\mathrm{f}$, and $\boldsymbol{W}[c] = \boldsymbol{W}_\mathrm{RF}\boldsymbol{W}_\mathrm{D}[c]$.

Based on (10), the N_s decision variables are grouped following the grouping at transmit, forming $\hat{\boldsymbol{y}}_\mathrm{g} \in \mathbb{C}^{N_\mathrm{g} \times N_\mathrm{s}}$ for the g-th group. Then, the joint maximum likelihood (JML) detectors are employed to recover the transmitted information, which can be represented as:

$$\hat{\boldsymbol{X}}_{\mathrm{VD},g} = \arg\min \|\hat{\boldsymbol{y}}_g - \hat{\boldsymbol{H}}_g \boldsymbol{A}_\mathrm{cs} \widetilde{\boldsymbol{X}}_{\mathrm{VD},g}\|_\mathrm{F}^2, \tag{11}$$

where $g = 1, 2, \cdots, G$, and $\tilde{\boldsymbol{X}}_{\mathrm{VD},g}$ is a test matrix for the $N_\mathrm{s}(p_1 + p_2)$ bits transmitted by the g-th group of subcarriers, $\hat{\boldsymbol{H}}_g$ is the resulted channels of the g-th group of subcarriers, with its elements from $\boldsymbol{W}^\mathrm{H}[c]\boldsymbol{H}[c]\boldsymbol{F}[c], c = 1, 2, \cdots, M_\mathrm{f}$.

D. mmWave Channel

Using the clustered channel model, the channel matrix $\boldsymbol{H}$ in (9), with N_cl clusters and N_ray rays in each cluster, can be expressed as:

$$\boldsymbol{H}_t = \tau \sum_{l=1}^{N_\mathrm{cl}} \sum_{i=1}^{N_\mathrm{ray}} \alpha_{l,i} \boldsymbol{v}_\mathrm{r}(\theta_{l,i}^\mathrm{r}, \phi_{l,i}^\mathrm{r}) \boldsymbol{v}_\mathrm{t}^\mathrm{H}(\theta_{l,i}^\mathrm{t}, \phi_{l,i}^\mathrm{t}) \tag{12}$$

where $\tau = \sqrt{\frac{N_\mathrm{t} N_\mathrm{r}}{N_\mathrm{cl} N_\mathrm{ray}}}$ is a normalization factor, ψ_l is proportional to the phase shift of each cluster, $\alpha_{l,i}$ is the complex channel gain of i-th ray of the l-th cluster, $\theta_{l,i}^\mathrm{r}$ and $\phi_{l,i}^\mathrm{r}$ are the elevation angle and azimuth angle of arrival, respectively, while $\theta_{l,i}^\mathrm{t}$ and $\phi_{l,i}^\mathrm{t}$ are the elevation angle and azimuth angle of departure. $\boldsymbol{v}_\mathrm{r}(\cdot)$ and $\boldsymbol{v}_\mathrm{t}(\cdot)$ express respectively the array manifold vectors of the transmitter antenna and receiver antenna arrays, which include all the spatial characteristics of the arrays [17]. Equivalently, the channel for each subcarrier can be expressed as [4, 37, 38]:

$$\boldsymbol{H}[c] = \tau \sum_{l=1}^{N_\mathrm{cl}} \sum_{i=1}^{N_\mathrm{ray}} \alpha_{l,i} \boldsymbol{v}_\mathrm{r}(\theta_{l,i}^\mathrm{r}, \phi_{l,i}^\mathrm{r}) \boldsymbol{v}_\mathrm{t}^\mathrm{H}(\theta_{l,i}^\mathrm{t}, \phi_{l,i}^\mathrm{t}) e^{-j2\pi\psi_l \frac{c}{M_\mathrm{f}}} \tag{13}$$

The channel matrix of (13) per subcarrier can also be written as:

$$\boldsymbol{H}[c] = \boldsymbol{V}_\mathrm{r} \mathrm{diag}(\boldsymbol{\alpha}[c]) \boldsymbol{V}_\mathrm{t}^\mathrm{H}, \tag{14}$$

where $\boldsymbol{V}_\mathrm{r} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{cl} N_\mathrm{ray}}$ and $\boldsymbol{V}_\mathrm{t} \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{cl} N_\mathrm{ray}}$ include the array manifold vectors of the receiver and transmitter arrays, respectively, $\boldsymbol{\alpha}[c] \in \mathbb{C}^{(N_\mathrm{cl} N_\mathrm{ray}) \times (N_\mathrm{cl} N_\mathrm{ray})}$ is a diagonal matrix with the diagonal elements given by $\{\alpha_{l,i}\}$.

For example, when the uniform linear arrays (ULA) with N elements is considered [17], the array manifold vector is

$$\begin{aligned} \boldsymbol{v}_\mathrm{ULA}(\theta) =& [\boldsymbol{v}_{\mathrm{ULA},0}(\theta), \cdots, \boldsymbol{v}_{\mathrm{ULA},N-1}(\theta)]^\mathrm{T}, \\ [\boldsymbol{v}_{\mathrm{ULA},n}(\theta)] =& e^{j(n - \frac{N-1}{2}) d_\mathrm{z} \frac{2\pi}{\lambda} \cos(\theta)}, n = 0, \cdots, N-1, \end{aligned} \tag{15}$$

where λ is the signal's wavelength, and $d_\mathrm{z} = \lambda/2$ denotes the distance between adjacent elements. Let us now develop the sparse channel model for the considered mmWave MIMO system in the next section, so that channel estimation can be built on it.

III. Sparse Bayesian Learning-based Channel Estimation

Due to the highly directional nature of mmWave propagation, mmWave channel can be represented using beamspace, where only some specific directions of the beamspace channel have non-negligible power [3]. In this section, the sparse beamspace mmWave channel model for channel estimation is introduced, and both the conventional SBL and our proposed multi-layer SBL are explained.

A. Sparse mmWave Channel Model

The beamspace representation of the mmWave channel matrix for the c-th subcarrier can be expressed as

$$\boldsymbol{H}[c] = \boldsymbol{D}_\mathrm{r}(\Theta_\mathrm{r}) \boldsymbol{H}_\mathrm{b}[c] \boldsymbol{D}_\mathrm{t}^\mathrm{H}(\Theta_\mathrm{t}), \tag{16}$$

where $\boldsymbol{H}_\mathrm{b}[c] \in \mathbb{C}^{G_\mathrm{r} \times G_\mathrm{t}}$ denotes the equivalent channel matrix in the beam space. As described in [39, 40], there are usually only a small fraction of elements in $\boldsymbol{H}_\mathrm{b}[c]$ being non-zero, owing to the signal propagation nature of mmWave signals. $\boldsymbol{D}_\mathrm{t}(\Theta_\mathrm{t}) = [\boldsymbol{v}_\mathrm{ULA}(\theta_1), \ldots, \boldsymbol{v}_\mathrm{ULA}(\theta_{G_\mathrm{t}})]^\mathrm{T}$ and $\boldsymbol{D}_\mathrm{r}(\Theta_\mathrm{r}) = [\boldsymbol{v}_\mathrm{ULA}(\theta_1), \ldots, \boldsymbol{v}_\mathrm{ULA}(\theta_{G_\mathrm{r}})]^\mathrm{T}$ denote respectively the transmit and receive array response dictionary matrices, where $\Theta_\mathrm{t} =$

$\{\theta_1, \ldots, \theta_{g_\mathrm{t}}, \ldots, \theta_{G_\mathrm{t}}\}$ and $\Theta_\mathrm{r} = \{\theta_1, \ldots, \theta_{g_\mathrm{r}}, \ldots, \theta_{G_\mathrm{r}}\}$, corresponding to the transmit and receive antenna arrays, respectively, are chosen to satisfy the conditions of $\cos\theta_{g_\mathrm{t}} = \frac{2}{G_\mathrm{t}}(g_\mathrm{t} - 1), 1 \le g_\mathrm{t} \le G_\mathrm{t}$ and $\cos\theta_{g_\mathrm{r}} = \frac{2}{G_\mathrm{r}}(g_\mathrm{r} - 1), 1 \le g_\mathrm{r} \le G_\mathrm{r}$ [29].

For the purpose of channel estimation, the channel matrix $\boldsymbol{H}[c]$ can be represented in a vector form of $\boldsymbol{h}[c] \in \mathbb{C}^{N_\mathrm{r}N_\mathrm{t}\times 1}$ as [34]

$$\boldsymbol{h}[c] = \mathrm{vec}(\boldsymbol{H}[c]) = \underbrace{(\boldsymbol{D}_\mathrm{t}^\mathrm{H}(\Theta_\mathrm{t}) \otimes \boldsymbol{D}_\mathrm{r}(\Theta_\mathrm{r}))}_{\boldsymbol{\Psi}}\boldsymbol{h}_\mathrm{b}[c], \tag{17}$$

where $\boldsymbol{h}_\mathrm{b}[c] = \mathrm{vec}(\boldsymbol{H}_\mathrm{b}[c]) \in \mathbb{C}^{G_\mathrm{t}G_\mathrm{r}\times 1}$ represents the equivalent beamspace channel vector obtained by the column wise stacking of the beamspace channel matrix $\boldsymbol{H}_\mathrm{b}[c]$, and $\otimes$ denotes the matrix Kronecker product. $\boldsymbol{\Psi} \in \mathbb{C}^{G_\mathrm{t}G_\mathrm{r}\times N_\mathrm{t}N_\mathrm{r}}$ denotes the angle codebook matrix. Note that, $\boldsymbol{h}_\mathrm{b}[c]$ is expected to be a sparse vector, most elements of which are '0'. By contrast, the non-zero elements in $\boldsymbol{h}_\mathrm{b}[c]$ with non-negligible power correspond to the significant paths in the specific directions. The specific directions with the majority of power share the key elements to recover the original channels in the beamspace, which is further elaborated in the following sub-sections.

B. Problem Formulation of Sparse mmWave Channel Estimation

To estimate the channels, pilot data are transmitted. Let the pilot symbols transmitted on the c-th subcarrier be denoted by $\boldsymbol{p}[c] \in \mathbb{C}^{N_\mathrm{RF}\times 1}$. Then, the received signal from the c-th subcarrier after the RF combiner can be denoted as:

$$\begin{aligned}\boldsymbol{y}_\mathrm{p}[c] &= \boldsymbol{W}_\mathrm{RF}^\mathrm{H}\boldsymbol{H}[c]\boldsymbol{F}_\mathrm{RF}\boldsymbol{p}[c] + \boldsymbol{W}_\mathrm{RF}^\mathrm{H}\boldsymbol{n}[c] \\ &= \underbrace{\sqrt{\rho}(\boldsymbol{p}[c]^\mathrm{T}\boldsymbol{F}_{RF}^\mathrm{T}) \otimes \boldsymbol{W}_\mathrm{RF}^\mathrm{H}}_{\boldsymbol{\Phi}[c]}\boldsymbol{h}[c] + \boldsymbol{e}[c],\end{aligned} \tag{18}$$

where $\boldsymbol{e}[c] = \boldsymbol{W}_\mathrm{RF}^\mathrm{H}\boldsymbol{n}[c]$ and its covariance matrix is $\boldsymbol{R}[c] = \mathbb{E}[\boldsymbol{e}[c]\boldsymbol{e}^\mathrm{H}[c]] = \sigma^2(\boldsymbol{W}_{RF}^\mathrm{H}\boldsymbol{W}_{RF}) \in \mathbb{C}^{N_{RF}\times N_{RF}}$, $\boldsymbol{\Phi}[c] \in \mathbb{C}^{N_\mathrm{RF}\times(N_\mathrm{t}N_\mathrm{r})}$.

Note that, at the channel estimation stage, only the RF precoder $\boldsymbol{F}_\mathrm{RF}$ and the RF combiner $\boldsymbol{W}_\mathrm{RF}$ are employed. Both of them contain the magnitude normalized elements, whose phases are randomly drawn with an equal probability from the following set [34]:

$$\boldsymbol{\mathcal{A}} = \left\{0, \tfrac{2\pi}{2^{N_\mathrm{q}}}, \cdots, \tfrac{(2^{N_\mathrm{q}}-1)2\pi}{2^{N_\mathrm{q}}}\right\}, \tag{19}$$

where N_q denotes the number of the quantization bits of the phase shifter. Thus, we can express $\boldsymbol{F}_\mathrm{RF} = \frac{1}{\sqrt{N_\mathrm{t}}}e^{j\boldsymbol{\mathcal{A}}_\mathrm{f}}$ and $\boldsymbol{W}_\mathrm{RF} = \frac{1}{\sqrt{N_\mathrm{r}}}e^{j\boldsymbol{\mathcal{A}}_\mathrm{w}}$, where $\boldsymbol{\mathcal{A}}_\mathrm{f}$ and $\boldsymbol{\mathcal{A}}_\mathrm{w}$ are the randomly selected elements from $\boldsymbol{\mathcal{A}}$ for $\boldsymbol{F}_\mathrm{RF}$ and $\boldsymbol{W}_\mathrm{RF}$, respectively.

Substituting (17) into (18), the system model for the estimation of the beamspace channel vector $\boldsymbol{h}_\mathrm{b}[c]$ can be expressed as

$$\boldsymbol{y}_\mathrm{p}[c] = \boldsymbol{\Phi}[c]\boldsymbol{\Psi}\boldsymbol{h}_\mathrm{b}[c] + \boldsymbol{e}[c]. \tag{20}$$

As reviewed in Section I, there are the CS-based channel estimation techniques for solving the mmWave channel estimation problem. For example, an adaptive algorithm was proposed in [40] to estimate the mmWave channel parameters, which employed the orthogonal matching pursuit (OMP) algorithm. In comparison with the existing CS-based channel estimation techniques, the SBL algorithm has been demonstrated to provide a superior signal recovery performance by overcoming the drawback of the other techniques, such as that in [39, 41]. Hence, it has been applied in many different application scenarios [42]. However, the SBL-based channel estimation suffers from the high computational complexity, especially in the high resolution scenarios. Since mobile communications channels may change rapidly, it is of vital importance to have a channel estimation algorithm that is not only accurate but also low complexity.

In the following parts of this section, we first present a brief overview of the conventional SBL algorithm. Then the multi-layer SBL algorithm with low computational complexity is introduced. We also derive the Bayesian Cramer-Rao lower bound (BCRB) so as to characterize the efficiency of the proposed multi-layer SBL algorithm.

C. Conventional SBL Algorithm

The SBL algorithm begins by assigning the following parameterized Gaussian prior to the unknown beamspace channel vector $\boldsymbol{h}_\mathrm{b}[c]$ [29, 30]:

$$p(\boldsymbol{h}_\mathrm{b}[c]; \boldsymbol{\Gamma}[c]) = \textstyle\prod_{i=1}^{G_\mathrm{t}G_\mathrm{r}}(\pi\gamma_i[c])^{-1}\exp\left(-\frac{|\boldsymbol{h}_{\mathrm{b},i}[c]|}{\gamma_i[c]}\right), \tag{21}$$

where the hyperparameter $\gamma_i[c]$, $1 \le i \le G_\mathrm{t}G_\mathrm{r}$, represents the prior variance of the i-th element of $\boldsymbol{h}_\mathrm{b}[c] \in \mathbb{C}^{G_\mathrm{t}G_\mathrm{r}\times 1}$. $\boldsymbol{\Gamma}[c] = \mathrm{diag}(\gamma_1[c], \cdots, \gamma_{G_\mathrm{t}G_\mathrm{r}}[c])$ is the diagonal matrix of the hyperparamaters $\gamma_i[c]$. The *a posteriori* probability density function of the beamspace channel vector $\boldsymbol{h}_\mathrm{b}[c]$ can be evaluated as $p(\boldsymbol{h}_\mathrm{b}[c] \mid \boldsymbol{y}_\mathrm{p}[c]; \boldsymbol{\Gamma}[c]) \sim \mathcal{CN}(\boldsymbol{\mu}[c], \boldsymbol{\Sigma}[c])$, where $\boldsymbol{\mu}[c] \in \mathbb{C}^{G_\mathrm{t}G_\mathrm{r}\times 1}$ and $\boldsymbol{\Sigma}[c] \in \mathbb{C}^{G_\mathrm{t}G_\mathrm{r}\times G_\mathrm{t}G_\mathrm{r}}$ can be expressed as[29]

$$\begin{aligned}\boldsymbol{\mu}[c] &= \boldsymbol{\Sigma}[c]\boldsymbol{\Omega}^\mathrm{H}[c]\boldsymbol{R}^{-1}\boldsymbol{y}_\mathrm{p}[c], \\ \boldsymbol{\Sigma}[c] &= (\boldsymbol{\Omega}^\mathrm{H}[c]\boldsymbol{R}^{-1}\boldsymbol{\Omega}[c] + \boldsymbol{\Gamma}[c]^{-1})^{-1},\end{aligned} \tag{22}$$

where $\boldsymbol{\Omega}[c] = \boldsymbol{\Phi}[c]\boldsymbol{\Psi}$ denotes the sensing matrix with $\boldsymbol{\Omega}[c] \in \mathbb{C}^{N_\mathrm{RF}\times G_\mathrm{t}G_\mathrm{r}}$. From (22), it can be shown that the estimation of the beamspace channel $\boldsymbol{h}_\mathrm{b}[c]$ can be transformed to the estimation of the associated hyperparameter $\boldsymbol{\gamma}[c] = [\gamma_1[c], \cdots, \gamma_{G_\mathrm{t}G_\mathrm{r}}[c]]^\mathrm{T}$.

The aim of the SBL algorithm is to maximize the Bayesian evidence $p(\boldsymbol{y}_\mathrm{p}[c]; \boldsymbol{\Gamma}[c])$ to achieve an improved estimation performance of the sparse beamspace channel $\boldsymbol{h}_\mathrm{b}[c]$. To achieve this objective, the iterative expectation-maximization (EM) algorithm can be employed. In the expectation step (E-step), a function for the expectation of the log-likelihood evaluated using the current estimation of the parameters is created. In the maximization step (M-step), the parameters are computed to maximize the expected log-likelihood found on the E-step. The computed parameters are then used to determine the expectation function of the log-likelihood in the next E-step [43].

To be more specific, the expectation (E-step) in the k-th iteration evaluates the average log-likelihood of the complete data as [29]

$$\mathcal{L}(\boldsymbol{\Gamma}[c] \mid \hat{\boldsymbol{\Gamma}}^{(k-1)}[c]) = \mathbb{E}_{\boldsymbol{h}_{\mathrm{b}}[c]|\boldsymbol{y}_{\mathrm{p}}[c];\hat{\boldsymbol{\Gamma}}^{(k-1)}[c]}\left\{ \log\left[p(\boldsymbol{y}_{\mathrm{p}}[c], \boldsymbol{h}_{\mathrm{b}}[c]; \boldsymbol{\Gamma}[c])\right] \right\}, \quad (23)$$

where $\hat{\boldsymbol{\Gamma}}^{(k-1)}[c]$ denotes the matrix of the hyperparameter matrix $\boldsymbol{\Gamma}[c]$ estimated in the $(k-1)$-th iteration. Then, the M-step maximizes (23) with respect to the hyperparameter vector $\boldsymbol{\gamma}[c]$ as [30]

$$\hat{\boldsymbol{\gamma}}^{(k)}[c] = \arg\max_{\boldsymbol{\gamma}} \sum_{i=1}^{G_{\mathrm{t}}G_{\mathrm{r}}} \left(-\log \gamma_i[c] - \frac{\mathbb{E}_{\boldsymbol{h}_{\mathrm{b}}[c]|\boldsymbol{y}_{\mathrm{p}}[c];\hat{\boldsymbol{\Gamma}}^{(k-1)}[c]}\left\{ |\boldsymbol{h}_{\mathrm{b},i}[c]|^2 \right\}}{\gamma_i[c]} \right), \quad (24)$$

where the maximization problem for estimating the hyperparameter vector $\boldsymbol{\gamma}[c]$ can be simplified by decoupling the hyperparameter vector $\boldsymbol{\gamma}[c]$ into the estimates $\hat{\gamma}_i[c]$, which can be denoted by $\hat{\gamma}_i^{(k)}[c]$ in the k-th iteration. Finally, the mathematical expression of $\hat{\gamma}_i^{(k)}[c]$ can be given as

$$\hat{\gamma}_i^{(k)}[c] = \boldsymbol{\Sigma}^{(k)}[c](i,i) + |\boldsymbol{\mu}^{(k)}[c](i)|^2, \quad (25)$$

where $\boldsymbol{\Sigma}^{(k)}[c]$ and $\boldsymbol{\mu}^{(k)}[c]$ can be obtained from (22), and $\boldsymbol{\Gamma}^{(k)}[c] = \mathrm{diag}(\boldsymbol{\gamma}^{(k)})$. Upon the convergence, the estimate of the beamspace channel by the SBL algorithm is determined as *the posteriori* mean, *i.e.*, $\hat{\boldsymbol{h}}_{\mathrm{b}}[c] = \boldsymbol{\mu}^{(k)}[c]$. After obtaining the estimated beamspace channel matrix, the mmWave channel matrix of the c-th subcarrier in the frequency domain can be recovered based on (16) as

$$\boldsymbol{H}[c] = \boldsymbol{D}_{\mathrm{r}}(\Theta_{\mathrm{r}})\mathrm{vec}^{-1}(\hat{\boldsymbol{h}}_{\mathrm{b}}[c])\boldsymbol{D}_{\mathrm{t}}^{\mathrm{H}}(\Theta_{\mathrm{t}}), \quad (26)$$

where $\mathrm{vec}^{-1}(\cdot)$ denotes the conversion from the vector form to the matrix form, which is the inversion operation of (17).

The implementation of the SBL algorithm uses matrix inversions at each iteration. Hence, its complexity may be too high in the high resolution case, which is the main challenge of the SBL algorithm.

D. Algorithm for Multi-Layer SBL Channel Estimation

According to (17), the resolution of the transmit and receive array response dictionary matrices $\boldsymbol{D}_{\mathrm{t}}(\Theta_{\mathrm{t}})$ and $\boldsymbol{D}_{\mathrm{r}}(\Theta_{\mathrm{r}})$ significantly affects the channel estimation performance, the channel estimation accuracy increases as the resolution of $\boldsymbol{D}_{\mathrm{t}}(\Theta_{\mathrm{t}})$ and $\boldsymbol{D}_{\mathrm{r}}(\Theta_{\mathrm{r}})$ increases. However, increasing the resolution results in an increased channel estimation complexity. To further improve the estimation performance while without experiencing the increase of complexity, the multi-layer SBL algorithm is proposed for channel estimation in this paper. Compared to the conventional SBL algorithm, the proposed multi-layer SBL algorithm is able to achieve a similar MSE performance with lower complexity, when the same resolution of the dictionary matrices is employed.

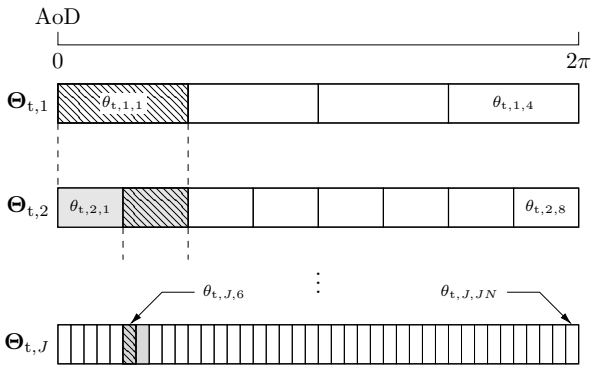


Fig. 3. The method of updating response dictionary for J layers.

Assume that there are J layers in the multi-layer SBL algorithm. Then the proposed algorithm can be detailed as follows. Let us use $\hat{\boldsymbol{h}}_{\mathrm{b},j}[c]$, $j = 1, \ldots, J$, to denote the estimated beamspace equivalent channel at the j-th layer. Firstly, the conventional SBL algorithm is employed at the first layer to estimate a beamspace vector equivalent channel $\hat{\boldsymbol{h}}_{\mathrm{b},1}[c]$. In the proposed multi-layer SBL algorithm, to reduce the estimation complexity, the lower resolution response dictionary matrices $\boldsymbol{D}_{\mathrm{t},j}$ and $\boldsymbol{D}_{\mathrm{r},j}$ are used at the first layer. As mentioned in Section III-B, the RF precoder and combiner employed at the first SBL layer are generated randomly, as CSI is unknown at this stage. Correspondingly, the beamspace channel $\hat{\boldsymbol{h}}_{\mathrm{b},1}[c]$ and $\hat{\boldsymbol{H}}_1[c]$ are the outputs of the first layer. Then, both the response dictionary matrices, precoder and combiner are updated based on the CSI obtained from the first layer of channel estimation. As an example, the transmitter array response dictionary matrix is updated as seen in Fig. 3. Specifically, for $j = 1$, the angle range $[0, 2\pi]$ is divided into N_1 sections, where N_1 is referred to as the resolution of the response dictionary at the first layer. In this example, a low resolution of $N_1 = 4$ is considered. Since $\boldsymbol{h}_{\mathrm{b}}[c]$ is a sparse vector, we can know that most of the elements of $\hat{\boldsymbol{h}}_{\mathrm{b},1}[c]$ should be close to zero. Hence, the proposed multi-layer SBL algorithm filters the elements [29], so as to further enhance its performance. Specifically, in this paper, two threshold-based approaches are introduced to determine the dominant elements:

- Approach-1: let ν be a value-wise threshold. Then, any beamspace channel elements with their values higher than ν are flagged as the channels with dominant power.
- Approach-2: let a percentage threshold be 95%, Then, the beamspace channel elements with their total power being 95% of the total power are flagged as the channels with dominant power.

The above two approaches are independently applied, only those elements flagged by both approaches are kept for further consideration, while the rest are set to zero values.

After identifying the close-to-zero elements, the positions of the non-zero elements in $\hat{\boldsymbol{h}}_{\mathrm{b},1}[c]$ correspond to the rows in $\boldsymbol{D}_{\mathrm{t},1}$ and the columns in $\boldsymbol{D}_{r,1}$, as shown in (26), which convey the main channel power in the AOA direction of $(\boldsymbol{\Theta})_{\mathrm{t},1}$ and AOD direction of $(\boldsymbol{\Phi})_{\mathrm{t},1}$. As shown in Fig. 3, $\theta_{\mathrm{t},1,1}$ is selected according to the positions of the non-zero elements in $\hat{\boldsymbol{h}}_{\mathrm{b},1}[c]$.

Then, the updated response dictionary $\boldsymbol{D}_{\mathrm{t},2}$, with a higher resolution of $N_2 = 8$ as shown in the example of Fig. 3, is employed at the 2nd layer. Simultaneously, the RF precoder $\boldsymbol{F}_{\mathrm{RF}}$ and combiner $\boldsymbol{W}_{\mathrm{RF}}$ are given by $\boldsymbol{F}_{\mathrm{SVD}}$ and $\boldsymbol{W}_{\mathrm{SVD}}$, which are obtained from the singular value decomposition (SVD) of $\hat{\boldsymbol{H}}_1[c]$ obtained from the first layer [3]. Note that, during the 2nd layer search, only the selected angle ranges in $\boldsymbol{D}_{\mathrm{t},1}$ that correspond to the non-zero elements in $\boldsymbol{h}_{\mathrm{b},1}[c]$ are considered in $\boldsymbol{D}_{\mathrm{t},2}$. This allows to achieve more accurate estimation of $\hat{\boldsymbol{h}}_{\mathrm{b},2}[c]$, while simultaneously at a lower complexity, when compared to the approach searching all the possible angles. In layer J, the resolution is $N_J = 2^{(J-1)}N_1$, which tends to be a large number. However, with our multi-layer SBL algorithm, only the angles in $\boldsymbol{D}_{\mathrm{t},J-1}$ that are identified with desired signal power are considered. Hence, the search space at layer J is small, meaning a relatively low complexity.

A summary of the proposed multi-layer SBL algorithm for OFDM mmWave system is given as Algorithm 1. Notice that, the hybrid precoder ($\boldsymbol{F}_{\mathrm{RF}}$, $\boldsymbol{F}_{\mathrm{D}}[c]$) and the combiner ($\boldsymbol{W}_{\mathrm{RF}}$, $\boldsymbol{W}_{\mathrm{D}}[c]$) in Algorithm 1 can be designed based on $\boldsymbol{F}_{\mathrm{SVD},J}$ and $\boldsymbol{W}_{\mathrm{SVD},J}$, respectively [3]. Since the focus of this paper is not on the design of the hybrid beamformers, but on the multi-layer SBL assisted adaptive IM, the approach in [3] is directly employed for the hybrid beamformer/combiner design.

More specifically, the proposed channel estimation algorithm is an iterative algorithm, where the codebook matrix of $\boldsymbol{\Psi}$, the precoder of $\boldsymbol{F}$ and the combiner of $\boldsymbol{W}$ are updated after each layer of the SBL algorithm, which are then fed to the next layer for further enhancement. In comparison with the algorithms using the full high resolution dictionary matrix, the proposed algorithm only needs to consider a fraction of the high resolution dictionary matrix after the first layer of operation. Hence, it is capable of achieving a similar estimation performance as the approach using the full high resolution dictionary, but with a substantially reduced complexity. Furthermore, as the number of layers of J increases, the accuracy of channel estimation improves and hence more accurate precoders and combiners can be obtained, which ensures an improved channel estimation in the following layers of estimation and finally an improved performance of the system.

Algorithm 1: Multi-layer SBL-based OFDM mmWave channel estimation

Input: Observation $\boldsymbol{y}_{\mathrm{p}}[c]$, sensing matrix $\boldsymbol{\Omega} = \boldsymbol{\Phi}\boldsymbol{\Psi}$, where $\boldsymbol{\Phi} = \sqrt{\rho}(\boldsymbol{p}[c]^{\mathrm{T}}\boldsymbol{F}_{RF}^{\mathrm{T}}) \otimes \boldsymbol{W}_{\mathrm{RF}}^{\mathrm{H}}$ and $\boldsymbol{\Psi} = \boldsymbol{D}_{\mathrm{T},1}^{\mathrm{H}}(\Theta_{\mathrm{t}}) \otimes \boldsymbol{D}_{R,1}(\Theta_{\mathrm{r}})$, noise covariance $\boldsymbol{R}$, maximum number of iterations $k_{\max}$, stopping parameter ϵ

Output: $\boldsymbol{H}_J[c] = \hat{\boldsymbol{D}}_{r,J}\boldsymbol{\mu}_J^{(k)}[c]\hat{\boldsymbol{D}}_{\mathrm{t},J}^{\mathrm{H}}$

Initialization: $\hat{\boldsymbol{\Gamma}}^0[n] = \boldsymbol{I}$, $\hat{\boldsymbol{\Gamma}}^{-1}[n] = 0$, $\boldsymbol{F}_{\mathrm{RF}} = \frac{1}{\sqrt{N_{\mathrm{t}}}}e^{j\boldsymbol{\mathcal{A}}_{\mathrm{f}}}$, $\boldsymbol{W}_{\mathrm{RF}} = \frac{1}{\sqrt{N_{\mathrm{r}}}}e^{j\boldsymbol{\mathcal{A}}_{\mathrm{w}}}$, $k = 1$
for $j = 1, 2, \ldots, J$, **do**
  $k \leftarrow k + 1$
  E-step: Evaluate the *posteriori* probability density
  $\boldsymbol{\mu}_j^{(k)}[c] = \boldsymbol{\Sigma}_j\boldsymbol{\Omega}_j^{\mathrm{H}}\boldsymbol{R}_j^{-1}\boldsymbol{y}_{p,v,j}[c]$
  $\boldsymbol{\Sigma}_j^{(k)}[c] = (\boldsymbol{\Omega}_j^{\mathrm{H}}\boldsymbol{R}_j^{-1}\boldsymbol{\Omega}_j + \boldsymbol{\Gamma}_j^{(k)}[c]^{-1})^{-1}$.
  M-step: Estimate the hyperparameters:
  for $i = 1, 2, \ldots, G_{\mathrm{t}}G_{\mathrm{r}}$, **do**
    $i = i + 1$
    $\hat{\gamma}_i^{(k)}[c] = \boldsymbol{\Sigma}^{(k)}[c](i,i) + |\boldsymbol{\mu}^{(k)}[c](i)|^2$
  end for
  if $\|\hat{\boldsymbol{\Gamma}}_j^k[n] - \hat{\boldsymbol{\Gamma}}_j^{(k-1)}[n]\|_2 \geq \epsilon$ and $k < k_{\max}$ **then**
    go 4
  end if
  $\boldsymbol{H}_j[c] = \boldsymbol{D}_{\mathrm{r}}\boldsymbol{\mu}_j^{(k)}[c]\boldsymbol{D}_{\mathrm{t}}^{\mathrm{H}}$
  $\mathrm{SVD}(\boldsymbol{H}_j[c]) = \boldsymbol{U}_j\boldsymbol{\Pi}_j\boldsymbol{V}_j^{-1}$
  $\boldsymbol{F}_{\mathrm{SVD},j} = \boldsymbol{U}_j^{1:N_s,1:N_{\mathrm{t}}}$, $\boldsymbol{W}_{\mathrm{SVD},j} = \boldsymbol{V}_j^{1:N_s,1:N_{\mathrm{r}}}$
  update: $\boldsymbol{D}_{\mathrm{t}} \leftarrow \hat{\boldsymbol{D}}_{\mathrm{T},j}$, $\boldsymbol{D}_{\mathrm{r}} \leftarrow \hat{\boldsymbol{D}}_{R,j}$
  $\boldsymbol{F}_{\mathrm{RF}} \leftarrow \boldsymbol{F}_{\mathrm{SVD},j}$, $\boldsymbol{W}_{\mathrm{RF}} \leftarrow \boldsymbol{W}_{\mathrm{SVD},j}$
end for
return $\boldsymbol{H}_J[c]$;

E. Bayesian Cramer-Rao Bound of Multi-Layer SBL Channel Estimator

This section derives the BCRB as a lower bound on the MSE of the proposed multi-layer SBL assisted channel estimator. The BCRB for the beamspace channel $\boldsymbol{h}_b[c]$ of the c-th subcarrier is given by the inverse of the Bayesian Fisher information matrix (FIM) $\boldsymbol{J}_{\mathrm{F}}$ expressed as:

$$\boldsymbol{J}_{\mathrm{F}} = \underbrace{-\mathbb{E}_{(\boldsymbol{y}_{\mathrm{p}}[c],\boldsymbol{h}_{\mathrm{b}}[c])}\left\{\frac{\partial^2\mathcal{L}(\boldsymbol{y}_{\mathrm{p}}[c] \mid \boldsymbol{h}_{\mathrm{b}}[c];\boldsymbol{\Gamma}[c])}{\partial\boldsymbol{h}_{\mathrm{b}}[c]\partial\boldsymbol{h}_{\mathrm{b}}^{\mathrm{H}}[c]}\right\}}_{\boldsymbol{J}_{\mathrm{D}}} - \underbrace{\mathbb{E}_{\boldsymbol{h}_{\mathrm{b}}[c]}\left\{\frac{\partial^2\mathcal{L}(\boldsymbol{y}_{\mathrm{p}}[c];\boldsymbol{\Gamma}[c])}{\partial\boldsymbol{h}_{\mathrm{b}}[c]\partial\boldsymbol{h}_{\mathrm{b}}^{\mathrm{H}}[c]}\right\}}_{\boldsymbol{J}_{\mathrm{P}}}, \quad (27)$$

where the matrices $\boldsymbol{J}_{\mathrm{D}}$ and $\boldsymbol{J}_{\mathrm{P}}$ denote the FIMs of the observation vector $\boldsymbol{y}_{\mathrm{p}}[c]$ and the parameter vector $\boldsymbol{h}_{\mathrm{b}}[c]$, respectively. Upon applying the results from [29, 30, 44] along with some further simplification, $\boldsymbol{J}_{\mathrm{D}}$ and $\boldsymbol{J}_{\mathrm{P}}$ can be formulated as $\boldsymbol{J}_{\mathrm{D}} = \boldsymbol{\Omega}^{\mathrm{H}}[c]\boldsymbol{R}[c]\boldsymbol{\Omega}[c]$ and $\boldsymbol{J}_{\mathrm{P}} = \boldsymbol{\Gamma}^{-1}[c]$. Finally, the BCRB for the MSE of the estimated beam space channel $\hat{\boldsymbol{h}}_b[c]$ of the c-th subcarrier satisfies:

$$\mathbb{E}\left\{\left\|\hat{\boldsymbol{h}}_b[c] - \boldsymbol{h}_b[c]\right\|^2\right\} \geq \mathrm{Tr}\left\{\left(\boldsymbol{\Omega}^{\mathrm{H}}[c]\boldsymbol{R}[c]\boldsymbol{\Omega}[c] + \boldsymbol{\Gamma}[c]^{(-1)}\right)^{-1}\right\}. \quad (28)$$

Correspondingly, the MSE of the estimated mmWave channel in the frequency domain can be obtained by substituting (17) into (28), satisfying:

$$\mathbb{E}\left\{\left\|\hat{\boldsymbol{H}}[c] - \boldsymbol{H}[c]\right\|^2\right\} \geq \mathrm{Tr}\left\{\boldsymbol{\Psi}\left(\boldsymbol{\Omega}^{\mathrm{H}}[c]\boldsymbol{R}[c]\boldsymbol{\Omega}[c] + \boldsymbol{\Gamma}[c]^{(-1)}\right)^{-1}\boldsymbol{\Psi}^{\mathrm{H}}\right\}. \quad (29)$$

Above we have addressed the channel estimation in mmWave OFDM systems. In practice, mmWave communication environment is time varying. In this case, it is expected that the mmWave communications systems' throughput can be improved with the adoption of adaptive modulation [17]. To implement adaptive modulation, the required CSI can be obtained from the proposed multi-layer SBL-based channel estimation. However, in real-time adaptive modulation, it is often difficult to switch the transmission mode at the right point, as the switching point depends on the SNR, which

in turn on the time-varying channel. Therefore, it is highly challenging for the transmitter to switch at the near optimum point. In order to circumvent this problem, below we introduce a learning-assisted adaptive modulation scheme, which does not depend on thresholds for switching modulation mode, but learns based on the observation of the communication environment.

IV. Adaptive Modulation

In our OFDM-CSIM system, the transmissions parameters that may be adapted include the modulation order Q, the number of active indices $N_{\rm a}$ per virtual sub-block, the total number $N_{\rm v}$ of virtual indices, and the number of subcarriers $N_{\rm g}$ of a sub-block, *etc.*. As introduced in Section II, the transmitted information is not only carried by the APM symbols, but also by the virtual indices, which are mapped to the symbols transmitted by a sub-block of subcarriers. Therefore, the system throughput can be adjusted via adaptively setting transmissions parameters, which also affect the achievable BER performance.

In the following two subsections, we present the conventional threshold-based adaptive modulation, followed by the learning-aided adaptive modulation. In our adaptive modulation schemes, the post-processing SNR is used as the metric for selecting the parameters Q, $N_{\rm v}$, $N_{\rm a}$, and $N_{\rm g}$ in the OFDM-CSIM system. The post-processing SNR can be expressed as [23]:

$$\xi = \mathbb{E}\left\{ \sum_{c=1}^{M_{\rm f}} \frac{{\rm Tr}\left\{ (\boldsymbol{W}^{\rm H}[c]\boldsymbol{H}[c]\boldsymbol{F}[c])^{\rm H}\boldsymbol{W}^{\rm H}[c]\boldsymbol{H}[c]\boldsymbol{F}[c] \right\}}{{\rm Tr}\left\{ \boldsymbol{W}[c]\boldsymbol{W}^{\rm H}[c]\sigma^2[c] \right\}} \right\}, \quad (30)$$

where $\boldsymbol{W}[c] = \boldsymbol{W}_{\rm RF}\boldsymbol{W}_{\rm D}[c]$ and $\boldsymbol{F}[c] = \boldsymbol{F}_{\rm RF}\boldsymbol{F}_{\rm D}[c]$.

A. Conventional Adaptive Modulation

In the conventional adaptive modulation, the parameters Q, $N_{\rm v}$, $N_{\rm g}$, and $N_{\rm a}$ can be selected according to the post-processing SNR by referring to the pre-defined thresholds set to maximize throughput, while maintaining a target BER. Specifically, the number of combinations of Q, $N_{\rm v}$, $N_{\rm g}$, and $N_{\rm a}$ can be defined as MODEs, expressed as MODE$_1$, MODE$_2$, $\cdots$, respectively, and each mode corresponds to a given data rate. Note that the number of modes M can be any value not exceeding the total number of combinations, depending on the trade-off between the implementation complexity and throughput.

As an example, Fig. 4 shows the BER performance of three different transmission modes, where the transmission parameters and corresponding data rates provided by these three modes are shown in Table I. Assume that the target BER is 10^{-3}. Then, as shown in Fig. 4, the specific SNR values T_1 and T_2 can be selected as the thresholds for determining the system to be operated in MODE$_1$, MODE$_2$ or MODE$_3$. In detail, MODE$_1$ is operated in the SNR range lower than T_1. MODE$_2$ with a higher data rate than MODE$_1$ is employed when the SNR is between T_1 and T_2, while MODE$_3$ with the highest rate is employed when the SNR is higher than T_2. Therefore, in order to implement the adaptive modulation, the receiver needs to decide the transmission mode by comparing

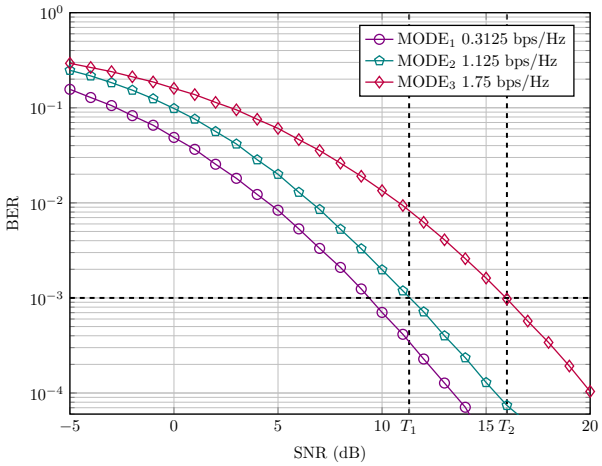


Fig. 4. BER versus SNR performance of the OFDM-CSIM system under different transmission modes, when the multi-layer SBL channel estimation is adopted.

TABLE I
System Parameters used in Simulations

	Q	$N_{\rm a}$	$N_{\rm v}$	$N_{\rm g}$	R (bps/Hz)
MODE$_1$	2	1	31	16	0.3125
MODE$_2$	4	3	31	16	1.125
MODE$_3$	16	2	16	8	1.75

the instantaneous post-processed SNR against the thresholds and feed back the selected mode to transmitter. In general, a higher data rate can be achieved if more optional modes are used, and the post-processing SNR is measured with higher accuracy. However, these are hard to achieve in the OFDM-CSIM systems, due to the many factors involved. Hence, in the next subsection, we propose the learning-based adaptive modulation for the OFDM-CSIM system to maximize its throughput.

B. Machine-Learning Aided Adaptive Modulation

The operation of adaptive modulation can be understood as a classification problem, which can hence be solved using ML techniques. Specifically in our system, adaptive modulation is the process of selecting a transmission mode from a range of candidate modes for a given post-processing SNR, which reflects the channel state. Let us use m as the index of a class. Then, each MODE$_m$ corresponds to a given data rate. Therefore, for a given channel state, adaptive modulation selects a class m that maps to a specific transmission mode MODE$_m$, so as to achieve the highest data rate under the constraint of the target BER. In this paper, both the k-NN and DNN techniques are investigated in the context of adaptive modulation.

There are two phases in the ML-aided adaptive modulation, namely the training phase and the learning phase. Firstly, the training data is generated separately from the testing data used in the learning phase, so as to avoid any correlation between them. In the proposed learning-based adaptive modulation system, for a MODE$_m$ of the modulation modes, a large number

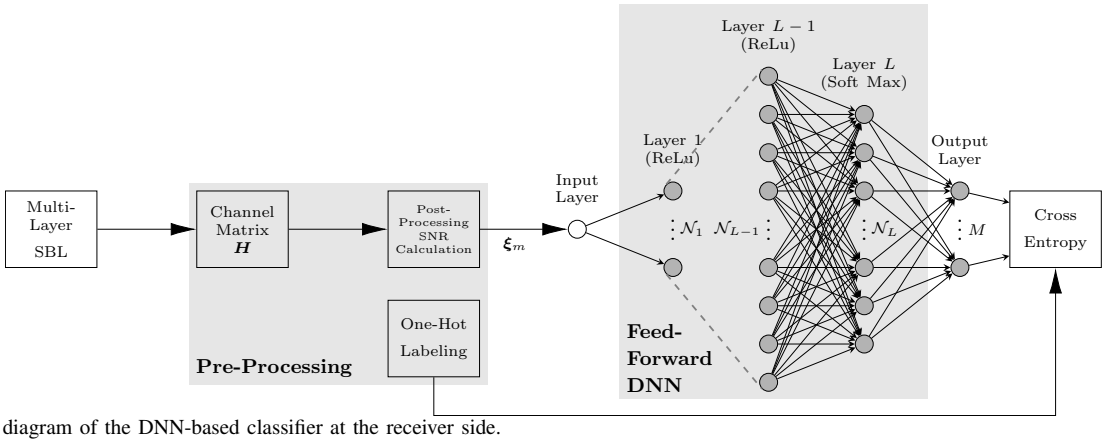


Fig. 5. Block diagram of the DNN-based classifier at the receiver side.

of post-processing SNR values are randomly generated, and the corresponding BER is evaluated. Then, a total N_m of the post-processings SNR values having the BER below the target of 10^{-3} are recorded and stored in a training set $\boldsymbol{T}^{(m)}$, expressed as

$$\boldsymbol{T}^{(m)} = [\xi_1^m, \ldots, \xi_n^m, \ldots, \xi_{N_m}^m]^{\mathrm{T}}, \tag{31}$$

where $m = 1, 2, \ldots, M$, $n = 1, 2, \ldots, N_m$, and ξ_n^m is the n-th generated SNR and a feature in the training set $\boldsymbol{T}^{(m)}$, with m being the corresponding label. Also note that $\sum_{m=1} N_m = N$. Then, the complete training set employed in the learning phase can be expressed as:

$$\boldsymbol{T} = [\boldsymbol{T}^{(1)}, \ldots, \boldsymbol{T}^{(m)}, \ldots, \boldsymbol{T}^{(M)}]^{\mathrm{T}}, m = 1, 2, \ldots, M. \tag{32}$$

After the training data is generated, the normalization of the feature set in (32) is implemented, in order to avoid any bias in the learning phase. The procedure of the normalization is expressed as:

$$\widetilde{\xi}_n^m = \frac{\left(\xi_n^m - \mathbb{E}\{\boldsymbol{T}\}\right)}{\left(\max(\boldsymbol{T}) - \min(\boldsymbol{T})\right)}. \tag{33}$$

In the learning phase, the learning system is established based on the normalized feature set and the corresponding labels. When the learning process is completed, a classifier capable of predicting the class labels for newly observed data is obtained. Below, we first employ a supervised-learning classifier (k-NN) to solve the adaptive modulation problem. Then, a DNN-based multi-label classifier is proposed to further improve the performance of the adaptive modulation system.

1) k-NN-Assisted Adaptive Modulation: k-NN and its variants are popular classification techniques due to the easy implementation and excellent classification performance [45]. In the context of the ML-based adaptive modulation supported by the k-NN based classification, the principle can be explained as follows:

- When a new observation data, which is given by the post-processing SNR ξ_{new}, is available, its normalization feature set $\widetilde{\xi}_{\mathrm{new}}$ is first obtained using (33). Then, the classifier searches over $\boldsymbol{T}$ to find the k nearest neighbors and their corresponding labels in the training sets, based on the distance metric $d(\cdot)$. When the Euclidian distance is employed, we have

$$d(\widetilde{\xi}_n^m, \widetilde{\xi}_{\mathrm{new}}) = \|\widetilde{\xi}_n^m - \widetilde{\xi}_{\mathrm{new}}\|_{\mathrm{F}}^2. \tag{34}$$

- Based on (34) and for a given value of k, the transmission mode is selected by the majority vote rule on the labels of the k nearest neighbors, as the one providing the highest number of neighbors to the observation data. Note that, it is possible that two or more MODEs provide the same maximum number of neighbors. In that case, the mode yielding the highest throughput is selected.

k-NN has the advantages of achieving the classification without requiring the information about the functional mapping between classifier and feature sets, and also has good performance for noisy training data [25, 46–48], but it is sensitive to the parameter setting and the value of k is the most important parameter to the classification performance of k-NN. The optimal value k should be searched for all test data before the training phase. In this adaptive system, the optimal value k is determined by considering the trade-off between the BER and throughput performance, as the maximum throughput is pursued under the requirement of target BER. Other challenges of k-NN include the high computational complexity for searching the nearest neighbors and the large memory space required for storing the training data, implying that the k-NN may be hard to implement in practice. More details can be found in [17].

2) DNN-Aided Adaptive Modulation: In the spirit of further reducing the processing complexity as well as the requirement for memory resource, in this section, we present a DNN-based classifier as a superior alternative to the k-NN solution discussed above. Fig. 5 shows the block diagram of the feed-forward DNN implemented at receiver side. In the pre-processing part, the system takes the estimates from the SBL algorithm of Section III to form a complex-valued channel matrix $\boldsymbol{H}$. A real-valued counterpart is subsequently obtained by calculating the corresponding post-processing SNR based on $\boldsymbol{H}$, which is then forwarded to the input layer of the DNN

The DNN is initially trained and then tested. As shown on the right side of Fig. 5, the feed-forward DNN consists of L layers of each having $\mathcal{N}_l$ neurons, where each layer has the input $\boldsymbol{v}_m^{l-1} \in \mathbb{R}^{\mathcal{N}_{l-1} \times 1}$ obtained from its previous layer. During the training phase, the training data $\widetilde{\boldsymbol{\xi}}_m$ is sent to the first layer of DNN, we have $\boldsymbol{v}_m^{(0)} = \widetilde{\boldsymbol{\xi}}_m$. The operation on the l-th layer ($1 \leq l \leq L-1$) can be expressed as

$$\boldsymbol{v}_m^{(l)} = \max(\boldsymbol{W}^{(l)} \boldsymbol{v}_m^{(l-1)} + \boldsymbol{b}^{(l)}, 0), \tag{35}$$

where $\boldsymbol{W}^{(l)} \in \mathbb{R}^{\mathcal{N}_l \times \mathcal{N}_{l-1}}$ is the weight matrix, and $\boldsymbol{b}^{(l)} \in \mathbb{R}^{\mathcal{N}_l \times 1}$ is the bias of the l-th layer, while $\max(\cdot, 0)$ is the Rectifier linear unit (ReLU) activation operation. Finally, at the last hidden layer ($l = L$), a softmax function is employed to replace the ReLU function, which is expressed as

$$\begin{aligned} \boldsymbol{v}_m^{\text{out}} = \boldsymbol{v}_m^{(L)} &= \text{softmax}(\boldsymbol{W}^{(L)}\boldsymbol{v}_m^{(l-1)} + \boldsymbol{b}^{(L)}) \\ &= \frac{\exp(\boldsymbol{W}^{(L)}\boldsymbol{v}_m^{(l-1)} + \boldsymbol{b}^{(L)})}{\sum_{q=1}^{\mathcal{N}_L} \exp([\boldsymbol{W}^{(L)}\boldsymbol{v}_m^{(l-1)} + \boldsymbol{b}^{(L)}]_q)}. \end{aligned} \tag{36}$$

Note that the last layer has M outputs, corresponding to the probabilities of the input $\widetilde{\boldsymbol{\xi}}_m$ belonging to the M classes, respectively. Note additionally that the actually desired label m of the input $\widetilde{\boldsymbol{\xi}}_m$ can be generated and formatted in a one-hop style as [49]

$$\boldsymbol{v}_m^{\text{desire}} = f_{\text{oh}}(m) = [\mathbb{I}(m = 1), \mathbb{I}(m = 2), \ldots, \mathbb{I}(m = M)], \tag{37}$$

where $f_{\text{oh}}(m)$ is the one-hop mapping function, while $\mathbb{I}(m = m')$ is the indicator function, whose value is 1 only when the condition $(m = m')$ is true and 0 otherwise. For example, the one-hot mapping for $M = 3$ possible classes is defined as

$$\begin{aligned} m = 1 &\iff \boldsymbol{v}_1^{\text{desire}} = f_{\text{oh}}(1) = [1, 0, 0], \\ m = 2 &\iff \boldsymbol{v}_2^{\text{desire}} = f_{\text{oh}}(2) = [0, 1, 0], \\ m = 3 &\iff \boldsymbol{v}_3^{\text{desire}} = f_{\text{oh}}(3) = [0, 0, 1]. \end{aligned} \tag{38}$$

Furthermore, during the training phase of the DNN, the weights $\boldsymbol{W}^{(l)}$ and the bias $\boldsymbol{b}^{(l)}$ of all layers are tuned using the Adam's algorithm [50] and a cross-entropy-based cost function $\mathcal{L}$ is employed, which is formulated as

$$\mathcal{L}_m(\boldsymbol{v}_m^{\text{desire}}, \boldsymbol{v}_m^{\text{out}}) = -\sum_q \boldsymbol{v}_m^{\text{desire}}[q] \log(\boldsymbol{v}_m^{\text{out}}[q]), \tag{39}$$

where $\boldsymbol{v}_m^{\text{out}}[q]$ denotes the q-th element of $\boldsymbol{v}_m^{\text{out}}$ and represents the predicted probability of the feature data ξ belonging to class m.

After the DNN is trained, the system can provide an output $\boldsymbol{v}_{\text{new}}$ for any new SNR observation $\widetilde{\boldsymbol{\xi}}_{\text{new}}$ by going through the network based on (35)-(39), and the index of the element in $\boldsymbol{v}_{\text{new}}$ having the highest value (probability) is the predicted class, i.e, $m_{\text{new}} = \arg\max_q \boldsymbol{v}_{\text{new}}$, yielding the selected modulation mode.

C. Complexity Analysis

In this subsection, the computational complexity of the proposed ML-based adaptive modulation is evaluated. Since the training phase can be processed off-line, we concentrate on the analysis of the testing phase. The complexity of the k-NN algorithm can be analysed from two perspectives, namely the memory required for storing the training data and the computational requirement for searching the nearest neighbors. The requirement of a large memory space is a major disadvantage of the k-NN algorithm, especially when single-chip devices are considered. In the meantime, a brute-force search for the nearest neighbors demands a high search complexity of $O(k\hat{d}M)$, where $\hat{d}$ represents the dimensions of the feature set.

For the proposed DNN-based algorithm, the computational complexity of the testing phase depends on the structure of the network, which includes the number of layers and the number of neurons in each layer. Also, the different activation functions lead to different computational complexity. More specifically, the complexity of the fully-connected layer with ReLU activation function is dominated by computing $\boldsymbol{v}_m^l = \max(\boldsymbol{W}^l\boldsymbol{v}_m^{l-1} + \boldsymbol{b}^l, 0)$, which is $O(2\mathcal{N}_l\mathcal{N}_{l-1} - \mathcal{N}_{l-1})$. For the last fully-connected layer with softmax activation function, the complexity depends on the computation of $\boldsymbol{v}_m^L = \text{softmax}(\boldsymbol{W}^L\boldsymbol{v}_m^{L-1} + \boldsymbol{b}^L, 0)$, which has the complexity of $O(2\mathcal{N}_L\mathcal{N}_{L-1} - \mathcal{N}_{L-1} + 2\mathcal{N}_{L-1})$. Thus, the computational complexity of the DNN-based algorithm in total is:

$$\mathcal{C} = O\left(\sum_{l-1}^{L}(2\mathcal{N}_l\mathcal{N}_{l-1} - \mathcal{N}_{l-1}) + 2\mathcal{N}_{L-1}\right). \tag{40}$$

Based on the above analysis, it is obvious that DNN is able to avoid the large storage required for saving the training data. The testing efficiency of DNN is also much higher than that of the brute-force search of k-NN. Moreover, the DNN-based adaptive modulation system is capable of achieving higher throughput than the conventional and k-NN-based systems, as detailed in the following Section V.

TABLE II
PARAMETERS USED FOR SIMULATIONS IN SECTION V

Parameter	Value
mmWave Channel	
Carrier frequency (f_{c})	28 GHz
Number of channel clusters (N_{cl})	4
Number of rays per cluster (N_{ray})	1
Number of transmit antenna (N_{t})	8
Number of receive antenna (N_{r})	8
Number of RF chains in transmitter ($\text{N}_{\text{RF}}^{\text{t}}$)	4
Number of RF chains in receiver ($\text{N}_{\text{RF}}^{\text{r}}$)	4
Distance between adjacent antennas ($d_{\text{t}} = d_{\text{r}}$)	$\lambda/2$
Number of subcarriers (N_{sub})	1024
SBL Channel Estimation	
Maximum number of iterations ($k_{\max}$)	300
Stopping parameter (ϵ)	10^{-6}
Number of SBL layers (J)	2
Adaptive Modulation	
Number of channel realizations for training	160000
Number of channel realizations for testing	5000
Number of neighbors in k-NN searching k	20
Number of fully-Connected (FC) layers in DNN (L)	3
Number of neurons in each FC layer	$[64, 128, 256]$
Activation functions after each FC layer	ReLU
Number of neurons on the output layer (M) Number of transmission modes / Classes (M)	3
Activation function after the output layer	soft max

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the OFDM-CSIM systems with our proposed multi-layer SBL channel estimation. We start by comparing the channel estimation performance of our multi-layer SBL algorithm assisted channel estimator and of the classic methods, which are also compared with the benchmarks of BCRBs. Then the performance of the OFDM-CSIM systems is compared when receiver for

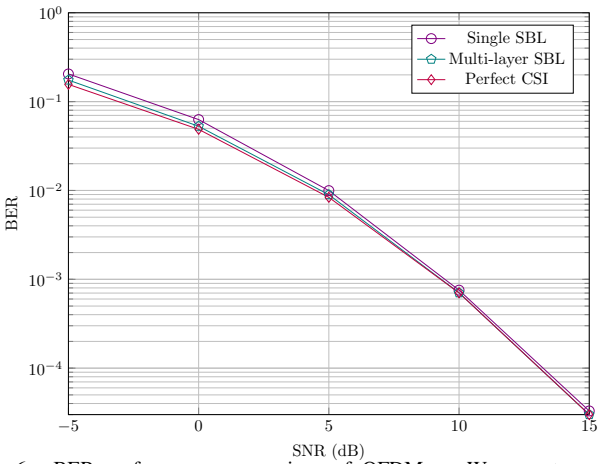


Fig. 6. BER performance comparison of OFDM mmWave systems with perfect CSI, the conventional SBL and multi-layer SBL algorithms assisted channel estimation.

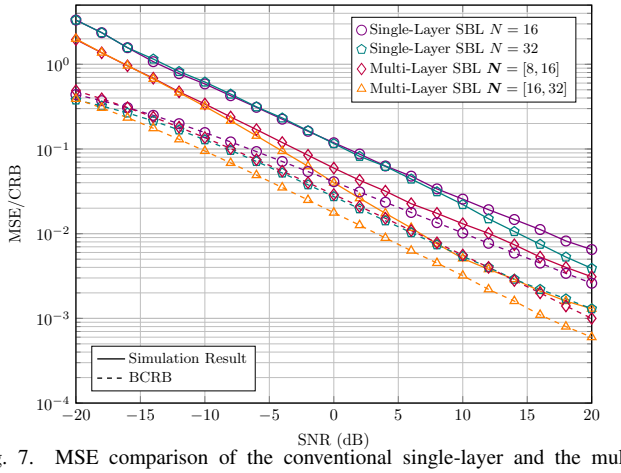


Fig. 7. MSE comparison of the conventional single-layer and the multi-layer SBL algorithms with different resolutions N, benchmarked against their corresponding BCRBs.

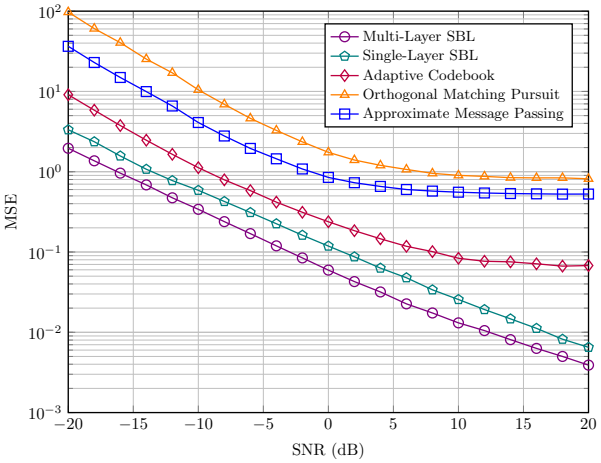


Fig. 8. MSE performance comparison of OFDM mmWave systems with different channel estimation techniques. The multi-layer SBL has the resolution of $\boldsymbol{N} = [8, 16]$, while for the rest cases of $N = 16$.

comparing the k-NN- and DNN-based adaptive modulations are involved. Finally, the computational complexity of the proposed scheme is investigated. Unless otherwise specified, the parameters of Table II are adopted for all the simulations in this section. Moreover, the elevation- and azimuth-angles for both the arrival and the departure rays are assumed to obey the Laplace distribution [3].

A. Multi-Layer SBL Channel Estimation

We first study the multi-layer SBL algorithm by comparing the system's BER and MSE performance, as well as the computational complexity, when the single- and multi-layer SBL algorithms are employed. It is worth mentioning that we adopted $J = 2$ layers in our multi-layer SBL algorithm for the sake of simplicity, while more layers can be added if required.

1) BER Performance: Fig. 6 shows the BER performance comparison of an OFDM system over mmWave channels using respectively the single- and multi-layer SBL algorithms, while the case with perfect CSI is also provided. It is clear from Fig. 6 that both the single- and multi-layer SBL algorithms are capable of providing near accurate channel estimation to achieve the similar BER performance as that with perfect CSI. Moreover, the multi-layer SBL shows slightly better BER performance than the single-layer SBL algorithm, especially in the low SNR region.

2) MSE Performance: In Fig. 7, we depict the MSE performance comparison of the single- and multi-layer SBL algorithms with different resolutions. The performance is also benchmarked against their corresponding BCRBs developed in Section III-E. It is clear that the MSE performance can be improved by using higher resolution provided by larger N, which leads to lower MSE in both the single- and the multi-layer SBL cases. Meanwhile, observe from Fig. 7 that the conventional single-layer SBL with $N = 16$ is outperformed by the proposed 2-layer SBL with $\boldsymbol{N} = [8, 16]$, as shown by both the simulated MSE and the respective BCRBs. The same observation applies to the single-layer SBL with $N = 32$ and the multi-layer SBL with $\boldsymbol{N} = [16, 32]$, where the latter one achieves lower MSE than the former, implying that more accurate angle estimation can be obtained by the multi-layer SBL algorithm. Furthermore, as the SVD-based hybrid beamformer is employed to replace the random-generated RF beamformer after the first layer of the multi-layer SBL algorithm, the sensing matrix becomes more efficient and accurate. This can be an additional contributor for achieving an improved MSE performance, in comparison with the conventional one.

Moreover, Fig. 8 compares the multi-layer SBL estimator with the other channel estimation techniques, including the single-layer SBL, adaptive codebook [40], OMP [39] and AMP [51] based estimators. For comparison, we adopt the resolution of $\boldsymbol{N} = [8, 16]$ for the 2-layer multi-layer SBL, while $N = 16$ is assumed with the other four techniques considered in Fig. 8. From the results we can clearly see that the proposed multi-layer SBL outperforms all the other channel estimators, when given the same value of resolution. The advantage of the multi-layer SBL estimator becomes more explicit in the high SNR region, as the performance of the OMP-based techniques are bottlenecked by the sensitivity of the dictionary matrix selection and the stopping criterion, which the SBL-based estimators can avoid.

3) Computational Complexity: According to Section III-C, the overall computation task in the SBL-based algorithms is predominately contributed by calculating (22) and (25). Therefore, their computational complexity can be characterized by

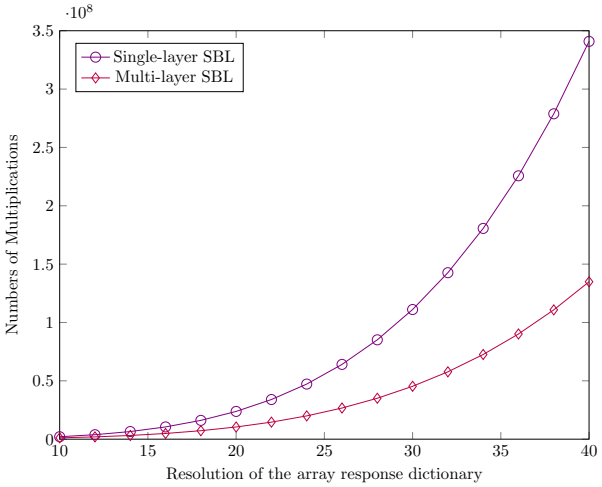


Fig. 9. The numbers of multiplications of the conventional SBL and multi-layer SBL algorithms with different resolution.

the number of multiplications in these two equations. Fig. 9 shows the relationship between the number of multiplications and the value N [2] of the resolution. It can be observed that the complexity of both algorithms increases as the resolution N increases. Moreover, the deviation between the two approaches increases with the increase of N. This in turn means that, when compared with the conventional SBL, a significant complexity reduction of the multi-layer SBL can be available, in particular, when high resolution dictionaries are employed. It is worth noting that the number of multiplications from the SVD employed in the multi-layer SBL is negligible, when compared to (22) and (25), and thus it can be ignored [52].

B. Adaptive Modulations

In this section, we apply the 2-layer SBL assisted detector having a resolution of $\boldsymbol{N} = [16, 32]$ to the k-NN- and DNN-based adaptive modulation systems proposed in Section IV. For the sake of a fair comparison, we use the data sets of the same size for both training and testing the k-NN- and DNN-based systems, as detailed in Table II.

Fig. 10 shows the probability distribution of choosing different transmission modes versus channel SNR for the conventional, k-NN-based and DNN-based adaptive modulations. It is worth noting that Fig. 10 is plotted against the *actual* channel SNR experienced by the signal, while the mode decisions are made based on the multi-layer SBL *estimated* SNR, thus gradual transitions from one mode to another are observed for all cases. According to the analysis in Section IV and Fig. 4, the switching thresholds for conventional adaptive modulation are $T_1 = 11.3$ dB and $T_2 = 16$ dB. Correspondingly, the conventional adaptive modulation of Fig. 10(a) uses MODE_1 when $\text{SNR} < T_1$, while the probability of choosing MODE_2 gradually arises after T_1 and reaches its peak at T_2, when the probability of choosing MODE_3 starts to increase. On the other hand, the two machine-learning-aided adaptive modulators, namely the k-NN-based one of Fig. 10(b) and the DNN-based one of Fig. 10(c), generally switch to higher SE modes at lower SNRs than the conventional one. Explicitly, at $T_1 = 11.3$ dB, when the conventional adaptive modulator just starts switching to use MODE_2 with a probability of around 2%, the probabilities of the same mode have already reached 3% and 5% for the k-NN and DNN, respectively. As another example, both the k-NN- and the DNN-based adaptive modulators manage to fully switch to MODE_3 after $\text{SNR} > 20$ dB, just about 4 dB higher than the conventional one, which results in a reduced achievable rate.

The reason behind the observations is that the learning-assisted adaptive modulations are capable of utilizing the *instantaneous* post-processing SNR to select a best possible mode, while the conventional adaptive modulation has to use the *pre-defined average* SNR-based thresholds for mode selection. Consequently, the learning-assisted designs are capable of making more accurate decisions than the conventional design, resulting in a significantly improved throughput for a given SNR, as seen in Fig. 11. Furthermore, the DNN-based adaptive modulation is able to achieve higher throughput than the k-NN-based one, where more accurate decisions can be made by the DNN classifier than the k-NN classifier.

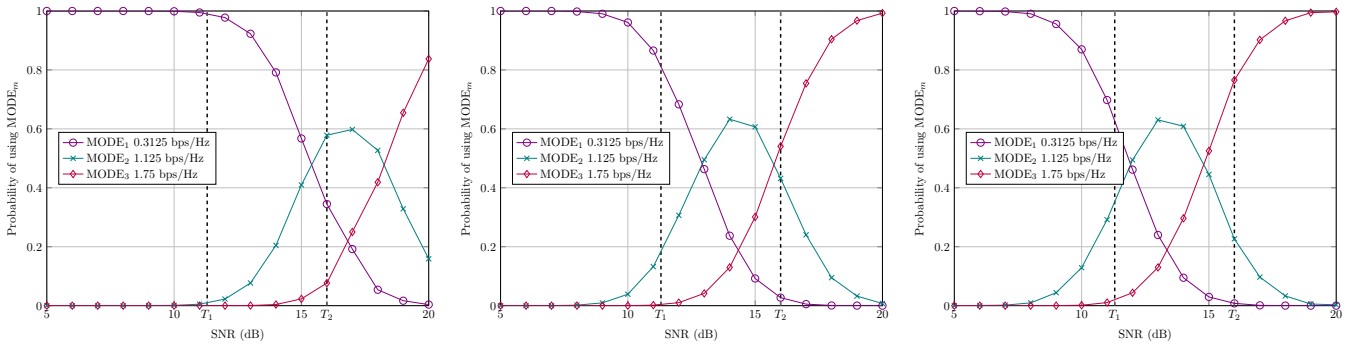
Fig. 12 shows the BER versus SNR performance of the conventional and the two learning-assisted adaptive modulation systems, where the BER curves of the three individual modulation modes are also included. Explicitly, all three adaptive modulations have the same BER performance as the MODE_1 curve until SNR reaches about 9 dB, i.e. when the 10^{-3} BER target is met. Afterwards, the adaptive modulation systems gradually switching to the next mode, in order to achieve higher throughput, while trying to retain the same BER level. It is however clear from Fig. 12 that the two learning-assisted adaptive modulations are capable of retaining the BER closer to the 10^{-3} target than the conventional adaptive modulation and hence attaining a higher throughput. Similarly, the DNN-based adaptive modulation is even closer to the BER target than k-NN-based design, as seen in Fig. 11.

Finally, we compare the computational complexity of the k-NN- and DNN-based adaptive modulations in Table III. We can see that the complexity of the DNN-based adaptive modulation is about 40 times lower than the k-NN one, due to the fact that k-NN requires to execute the exhaustive search over all possible CSI realizations, while DNN can be trained off-line for choosing the best transmission mode. Compared with k-NN, the complexity of the DNN-based adaptive modulation is dominated by the number of fully-connected layers and the size of each layer, rather than the number of components in the training set. Consequently, there is no requirement for memory to store the training data in the DNN-based adaptive modulation. It is noteworthy that in the conventional adaptive modulation, the transmission modes are selected by referring to the pre-defined thresholds, which do not consider the reserved memory or search complexity compared with the ML-based approaches. However, its achievable data rate may be significantly lower than that of the ML-based approaches.

VI. Conclusion

A deep-learning-assisted adaptive modulation relying on the multi-layer SBL channel estimation was proposed for

[2]For multi-layer SBL, N stands for the resolution for the second layer, while the resolution of the first layer is $2N$.



(a) Conventional adaptive modulation (b) k-NN-based adaptive modulation (c) DNN based adaptive modulation

Fig. 10. Probabilities for the adaptive system employing MODE$_1$, MODE$_2$ and MODE$_3$.

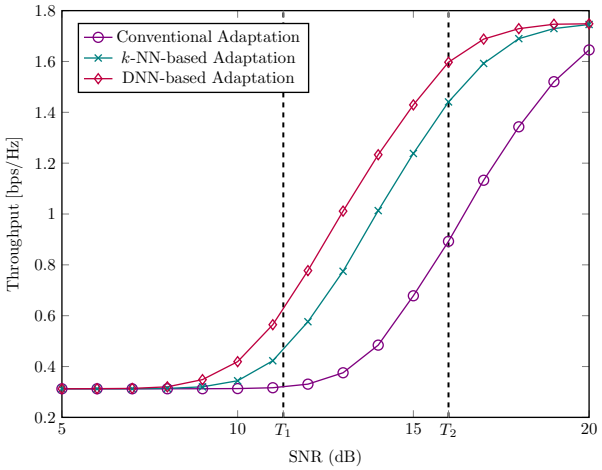


Fig. 11. Throughput comparison of the conventional, k-NN-based and DNN-based adaptive modulation.

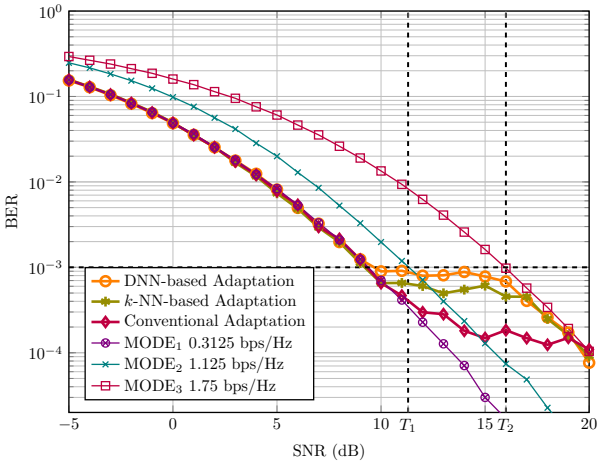


Fig. 12. BER versus E_s/N_0 performance of the conventional, k-NN-based and DNN-based adaptive modulations, as well as of the three individual MODEs.

TABLE III
COMPLEXITY ORDERS OF THE ML-ASSISTED ADAPTIVE SCHEMES

System	Complexity Equation	Complexity Order
k-NN	$O(k\hat{d}M)$	3.2×10^6
DNN	(40)	8.8×10^4

supporting the OFDM-CSIM systems communicating over mmWave channels. It is shown that the multi-layer SBL algorithm is capable of providing more accurate estimation at a lower cost of computational complexity, while achieving better BER and MSE performance, when compared with the conventional single-layer SBL algorithm and the other approaches relying on, such as, adaptive codebook and OMP. Then, a DNN-based adaptive modulator was conceived for utilizing the channel estimation to maximize the system throughput under a given BER target. Simulation results show that this DNN-based approach outperforms the conventional average-SNR-based adaptive modulations. Furthermore, it also has a lower computational complexity but a higher throughput than the previously proposed k-NN-based approach. Our studies imply that the ML-based adaptive modulation is able to "learn through experience", leveraging the existing training data to self-optimize on fly. Consequently, it is able to make more accurate decision and hence, attain higher throughput.

One of the most important applications of ML is the data-driven decision making, which enables adaptive transmission without CSI or SNR. A critical extension of our work is to directly use the received signals as the feature set, so as to reduce the complexity of channel estimation. In this case, however, the other techniques should be correspondingly designed for detecting the OFDM-CSIM signals.

REFERENCES

[1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Apr. 2015.

[2] 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. Surv. Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart. 2018.

[3] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[4] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[5] S. Lien, S. Shieh, Y. Huang, B. Su, Y. Hsu, and H. Wei, "5G new radio: Waveform, frame structure, multiple access, and initial access," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, Jun. 2017.

[6] R. Abu-alhiga and H. Haas, "Subcarrier-index modulation OFDM," in *IEEE 20th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Tokyo, Japan, Sep. 2009, pp. 177–181.

[7] M. Wen, B. Zheng, K. J. Kim, M. Di Renzo, T. A. Tsiftsis, K.-C. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging

wireless systems: Research progresses and applications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 1949–1972, 2019.

[8] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (SIM) OFDM," in *Proc. 2011 IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, TX, Dec. 2011, pp. 728–732.

[9] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536–5549, Nov. 2013.

[10] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.

[11] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.

[12] I. Al-Nahhal, O. A. Dobre, and S. Ikki, "Low complexity decoders for spatial and quadrature spatial modulations - Invited paper," in *IEEE 87th Vehicular Technology Conference (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–5.

[13] H. Zhang, L. Yang, and L. Hanzo, "Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, 2016.

[14] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4786–4801, 2017.

[15] M. Wen, Q. Li, E. Basar, and W. Zhang, "Generalized multiple-mode ofdm with index modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6531–6543, 2018.

[16] S. Lu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Compressed-sensing-aided space-time frequency index modulation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6259–6271, Jul. 2018.

[17] 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.

[18] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 108–115, Jun. 2002.

[19] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, no. 5, pp. 609–640, May 2000.

[20] R. C. Daniels, C. M. Caramanis, and R. W. Heath, "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 114–126, Jan. 2010.

[21] T. Keller and L. Hanzo, "Adaptive modulation techniques for duplex ofdm transmission," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1893–1906, 2000.

[22] S. S. Das, E. D. Carvalho, and R. Prasad, "Performance analysis of OFDM systems with adaptive sub carrier bandwidth," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1117–1122, Apr. 2008.

[23] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Multi-user hybrid beamforming relying on learning-aided link-adaptation for mmWave systems," *IEEE Access*, vol. 7, pp. 23 197–23 209, 2019.

[24] L. Hanzo, C. H. Wong, and M.-S. Yee, *Adaptive wireless transceivers: turbo-coded, turbo-equalised and space-time coded TDMA, CDMA and OFDM systems*. John Wiley & Sons, 2002.

[25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[26] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[27] R. Daniels and R. W. Heath, "Online adaptive modulation and coding with support vector machines," in *2010 Eur. Wireless Conf. (EW)*, Lucca, Italy, Apr. 2010, pp. 718–724.

[28] S. Srivastava, P. Sharma, S. Dwivedi, A. K. Jagannatham, and L. Hanzo, "Fast block LMS based estimation of angularly sparse channels for single-carrier wideband millimeter wave hybrid MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 666–681, 2021.

[29] A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Sparse bayesian learning-based channel estimation in millimeter wave hybrid mimo systems," in *IEEE 18th Intl. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[30] 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 *Nat. Conf. Commun. (NCC)*, Bangalore, India, Feb. 2019, pp. 1–6.

[31] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theor.*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[32] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theor.*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[33] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, 2010, pp. 1–5.

[34] R. T. Suryaprakash, M. Pajovic, K. J. Kim, and P. Orlik, "Millimeter wave communications channel estimation via Bayesian group sparse recovery," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3406–3410.

[35] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 2168–2172.

[36] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A gamp-based low complexity sparse bayesian learning algorithm," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 294–308, 2018.

[37] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[38] X. Yu, J. Zhang, and K. B. Letaief, "Alternating minimization for hybrid precoding in multiuser ofdm mmwave systems," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 281–285.

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

[40] 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, Oct. 2014.

[41] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, 2017.

[42] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in ofdm systems using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, 2014.

[43] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[44] H. Hijazi and L. Ros, "Bayesian Cramer-Rao bounds for complex gain parameters estimation of slowly varying rayleigh channel in OFDM systems," *Signal Process.*, vol. 89, no. 1, pp. 111–115, 2009.

[45] Y. Bengio, I. Goodfellow, and A. Courville, *Deep learning*. MIT press Massachusetts, USA:, 2017.

[46] Y. Zhang and X. Wu, "Integrating induction and deduction for noisy data mining," *Information Sciences*, vol. 180, no. 14, pp. 2663–2673, 2010.

[47] H. Liu and S. Zhang, "Noisy data elimination using mutual k-nearest neighbor for classification mining," *Journal of Systems and Software*, vol. 85, no. 5, pp. 1067–1074, 2012.

[48] X. Wang, "A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality," in *Int. Joint Conf. Neural Netw.* IEEE, 2011, pp. 1293–1299.

[49] P. Yang, Y. Xiao, M. Xiao, Y. L. Guan, S. Li, and W. Xiang, "Adaptive spatial modulation MIMO based on machine learning," *IEEE J. Sel. Areas Commun*, vol. 37, no. 9, pp. 2117–2131, 2019.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

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

[52] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Siam, 1997, vol. 50.