Received: 20 August 2021 | Revised: 9 December 2021 | Accepted: 5 January 2022

IET Communications

DOI: 10.1049/cmu2.12329

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Exploiting angular spread in channel estimation of millimeter wave MIMO system

You You[1,2] | Li Zhang[1]

[1] School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK

[2] Purple Mountain Laboratories, Nanjing, China

Correspondence
Li Zhang, the School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK.
Email: l.x.zhang@leeds.ac.uk

Abstract

Millimeter wave (mmWave) frequency spectrum can mitigate severe spectrum shortage caused by the explosive growth of mobile data demand. To overcome the high propagation loss of mmWave signals, massive multi-input multi-output (MIMO) and hybrid architecture are employed. As in microwave communication systems, channel state information (CSI) is essential to fully achieve the advantages of mmWave communication. However, due to the massive number of antennas and hybrid architecture, the CSI acquisition is challenging. The sparsity of mmWave channel can be utilized to reduce the training overhead. In addition to sparsity, real-world measurements in dense urban propagation environments reveal that the mmWave channel may spread in form of cluster of paths over the angular domains, namely the angular spread. In this paper, it is utilized to formulate the channel estimation as a block-sparse signal recovery problem. The block orthogonal matching pursuit (BOMP) is used to validate the model. Then, block fast Bayesian matching pursuit (BFBMP) algorithm is proposed to solve the above problem. Compared with other existing channel estimation methods, simulation results show that the angular spread feature and the proposed BFBMP can considerably improve the CSI estimation with less complexity.

1 | INTRODUCTION

Millimeter wave (mmWave) is a promising approach for the fifth Generation (5G) and beyond wireless networks by virtue of huge available bandwidth [1]. However, mmWave communication has to overcome the high propagation loss which is typical in this frequency range. Thanks to the small wavelength, massive MIMO can be deployed at both transmitter and receiver to provide sufficient beamforming gains. Hybrid architecture has been proved to achieve desirable beamforming performance with much less power consumption compared with fully digital architecture. As in microwave communications, reliable channel state information (CSI) is also required for mmWave communication to fully unleash the advantages. However, applying hybrid beamforming and massive MIMO makes CSI acquisition challenging because of the large number of antennas and hybrid architecture [2].

Recently, the mmWave channel has been proved to have sparsity in angular domain [3]. The virtual channel representation [4] can be exploited to avoid estimating all the entries in the channel matrix. Instead only the angle-of-departure (AoD), the angle-of-arrival (AoA), and the corresponding path gains of dominant paths need to be estimated. Then the channel estimation problem can be formulated as a sparse signal recovery problem and compressive sensing (CS) methods are widely applied to estimate the CSI efficiently. A closed-loop beam scanning technique was proposed in [5]. This method can avoid an exhaustive beam search and significantly reduce the search complexity by the multistage process but the performance of beam scanning is greatly affected by beamforming dictionary (codebook). For example, an improved codebook was proposed in [6]. Through using continuous basis pursuit (CBP), it significantly improves the estimation accuracy. However, close-loop beam scanning techniques are difficult for the outdoor environment where the communication needs larger beamforming gain, because it is difficult to achieve sufficient beamforming gain for wide searching beams at initial stages with limited transmitted power. An alternative approach is to use an open-loop channel estimator which applies fixed width training beam and does not need feedback from the receiver

[7]. In [7], the orthogonal matching pursuit (OMP) algorithm was used to solve the sparse signal recovery problem.

In order to further improve the channel estimation performance, Bayesian based compressive sensing methods such as sparse Bayesian learning (SBL) [8] and fast Bayesian matching pursuit (FBMP) [9] have been applied in mmWave channel estimation. SBL is a learning method. It assumes that each element follows Gaussian distribution with unknown variances which are assigned Gamma conjugate prior. SBL utilizes Expectation maximization (EM) method to learn the variances and compute a maximum a posteriori (MAP) estimate. FBMP is another Bayesian based method. It makes appropriate assumptions of noise variance and non-zero element variance according to the characteristics of mmWave channel and searches a set of candidate sparsity patterns (SP) with high posterior probabilities to estimate the CSI. Both SBL and FBMP show superior performance than OMP but with much higher complexity. SBL has the largest complexity among them because of the learning process.

The characteristics of mmWave channel are further studied. For instance, different delay taps of the wideband channel may share the same AoD/AoAs. This is utilized in [10] to reduce the training overhead. The clustering block sparse Bayesian learning (CBSBL) algorithm for mmWave channel estimation is proposed in [11]. It exploits the correlation between the mmWave channel to avoid the suboptimal solutions.

In addition to the above characteristic, mmWave channels are expected to exhibit limited scattering with only a few propagation paths even in urban environments. Several real-world measurements in dense-urban propagation environment reveal the clustered nature of the mmWave channels in the angular domains, with a few non-negligible angular spreads in both AoA and AoD [1, 12]. In this paper, we focus on limited scattering environment. the real-world channel measurements at 28 and 73 GHz were shown to have an angular spread of 15.5° and 15.4°, respectively, in terms of root mean-squared (rms) beam spread at AoA [1, 12]. While the measured AoD spreads (in terms of rms) are 10.2° and 10.5°, respectively. For example, in Figure 1, a signal sent from the transmitter reaches the receiver via a few clusters of paths. Since the wavelength of electromagnetic waves in the mmWave system is likely comparable with the roughness of the object surfaces that bounce off the waves, the departures at the transmitter and arrivals at the receiver are likely to be spread in angular domain as **a**, **b** and **c**, **d** at transmitter and receiver respectively. Path cluster power profiles were generated in [13] using the proposed statistical channel model with fitted large-scale parameters in [12]. As demonstrated in [13], the angular spreads give rise to a structured sparsity pattern that can be exploited to improve the mmWave channel estimation performance. Thus, the structurally limited scattering channel model has been adopted in the literature [14] [13] and [15]. A two-stage compressed sensing scheme was proposed and it was shown that the proposed scheme achieves a lower sample complexity than a conventional compressed sensing method that exploits only the sparse structure of mmWave channels[14]. The proposed methods in [13] essentially couple the channel path power at one angular direction with its two-dimensional

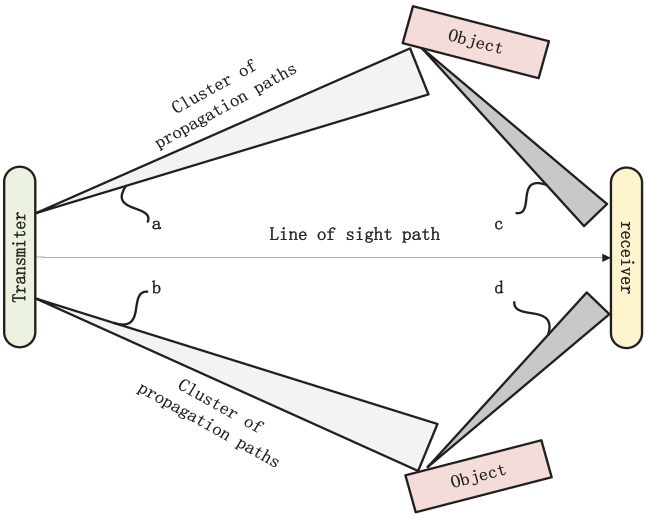

FIGURE 1 Angular spreads in mmWave communication [16]

AoD-AoA neighboring directions and adopt coupled sparse Bayesian learning to estimate the CSI. The proposed methods in[15] address the channel estimation problem within a Bayesian framework. Specifically, they adopt a matrix factorization formulation and translate the problem of channel estimation into searching for two-factor matrices. Then a variational Bayesian inference method is proposed for the mmWave channel estimation. However, both [13] and [15] are based on the Bayesian learning method which adopts expectation-maximization (EM) algorithm to estimate the hyperparameters with huge computational complexity.

In this paper, we exploit the sparsity in the angular domain and make use of the angular spread of path clusters in the AoA domain. Two-dimensional joint AoD-AoA spread will be our future work. Different from [13], [14], and [15] which study angular spread based on the low-rank structure or statistical probability, we derive the AoA angular spreads as blocks in channel matrix directly with some assumptions and utilize the block sparsity by formulating the channel estimation to a block signal recovery problem. Block orthogonal matching pursuit (BOMP) algorithm is applied to validate our channel estimation formulation. We then utilize this block property in the Bayesian matching based mmWave channel estimation and proposed the block fast Bayesian matching pursuit (BFBMP) method. Simulation results show that BFBMP produces superior performance compared with the existing methods including BOMP at all SNRs with fixed block patterns (without overlaps). BFBMP also performs better than most algorithms at low SNRs without fixed block patterns (with overlaps). The contribution of this paper can be summarized as follows.

1) We formulate channel estimation as a block sparse signal recovery problem exploiting the angular spread at AoAs. BOMP algorithm is applied to validate the channel estimation formulation.
2) In the proposed BFBMP algorithm, angular spread feature is utilized in mmWave channel estimation to further improve

the estimation performance. The extension from FBMP to BFBMP is not straightforward: The channel modelling and problem formulation are derived to show block structure and the parameters in BFBMP algorithm are calculated block by block instead of element by element. In addition, the update of metrics must also be redesigned.

3) Simulation results show that BFBMP is able to achieve accurate channel estimation at all SNR without overlaps. And it has less computational complexity compared with conventional Bayesian learning based and Bayesian matching pursuit based methods. Moreover, with overlaps of the AoA angular spread, BFBMP still achieves superior performance at low SNR.

The remainder of this paper is organized as follows. In Section 2, we introduce the mmWave communication system model and formulate the CE as a sparse signal recovery problem. In Section 3, we formulate channel estimation as a block sparse recovery problem exploiting AoA angular spreads. BOMP is applied to validate the formulation. Section 4, we propose the BFBMP algorithm to further improve the channel estimation performance and reduce the high complexity induced by the searching process in FBMP. In Section V, simulation results are presented to demonstrate the superiority of the BFBMP. In Section VI, we conclude the paper.

2 | SYSTEM MODEL AND PRIOR WORK

We consider a single-user hybrid MIMO system, where the base station (BS) is equipped with N_T antennas and N_{RF} RF chains communicating with a mobile station (MS) with N_R antennas and N_{RF} RF chains ($N_{RF} \ll \min(N_T, N_R)$).

In the channel estimation stage, BS and MS apply N_T^{Beam} ($N_T^{Beam} \leq N_T$) different pilot beam training patterns denoted as $\{\mathbf{f}_m \in \mathbb{C}^{N_T \times 1} : m = 1, \dots, N_T^{Beam}\}$ and N_R^{Beam} ($N_R^{Beam} \leq N_R$) different receive beam patterns denoted as $\{\mathbf{w}_n \in \mathbb{C}^{N_R \times 1} : n = 1, \dots, N_R^{Beam}\}$, respectively. We consider pilot beams are sent to MS successively and each $\mathbf{f}_m$ is received through all receive beam patterns. Because the MS is equipped with N_{RF} RF chains, MS is only able to generate N_{RF} different receive beams simultaneously so that receive beam patterns need $N_R^{Block} = \frac{N_R^{Beam}}{N_{RF}}$ time slots to scan over. The received signal for the m^{th} pilot beam in the q^{th} time slot is denoted as $\mathbf{y}_{q,m} \in \mathbb{C}^{N_{RF} \times 1}$ for $q \in \{1, \dots, N_R^{Block}\}$. We assume N_R^{Beam} and N_T^{Beam} are multiples of N_{RF}. The received vector for the q^{th} time slot and the m^{th} transmit beam is given by

$$\mathbf{y}_{q,m} = \mathbf{W}_q^H \mathbf{H} \mathbf{f}_m x_p + \mathbf{W}_q^H \mathbf{n}_{q,m}, \tag{1}$$

where $\mathbf{W}_q = [\mathbf{w}_{(q-1)N_{RF}+1}, \dots, \mathbf{w}_{qN_{RF}}] \in \mathbb{C}^{N_R \times N_{RF}}$ is the receive beam pattern matrix for $\mathbf{f}_m$ in the q^{th} time slot. x_p is the transmitted pilot symbol. $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ represents the channel matrix, and $\mathbf{n}_{q,m} \in \mathbb{C}^{N_R \times 1}$ is the noise vector in the q^{th} time slot. Collecting $\mathbf{y}_{q,m}$ for $q \in \{1, \dots, N_R^{Block}\}$, we get complete signal $\mathbf{y}_m \in \mathbb{C}^{N_R^{Beam} \times 1}$ for pilot beam pattern $\mathbf{f}_m$ as

$$\begin{aligned} \mathbf{y}_m = \mathbf{W}^H \mathbf{H} \mathbf{f}_m x_p + \operatorname{diag}\left(\mathbf{W}_1^H, \dots, \mathbf{W}_{N_R^{Block}}^H\right) \\ \times \left[\mathbf{n}_{1,m}^T, \dots, \mathbf{n}_{N_R^{Block},m}^T\right]^T, \end{aligned} \tag{2}$$

where $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_{N_R^{Block}}] \in \mathbb{C}^{N_R \times N_R^{Beam}}$. To represent the signals for all N_T^{Beam} transmit beams, we collect $\mathbf{y}_m$ for $m \in \{1, \dots, N_T^{Beam}\}$ to get

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{X} + \mathbf{N} \\ &= \sqrt{P} \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{N}, \end{aligned} \tag{3}$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_{N_T^{Beam}}] \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$, $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_{N_T^{Beam}}] \in \mathbb{C}^{N_T \times N_T^{Beam}}$ and $\mathbf{N} \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$ is the noise matrix given by

$$\begin{aligned} \mathbf{N} = \operatorname{diag}\left(\mathbf{W}_1^H, \dots, \mathbf{W}_{N_R^{Block}}^H\right) \left[\left[\mathbf{n}_{1,1}^T, \dots, \mathbf{n}_{N_R^{Block},1}^T\right]^T, \right. \\ \left. \dots, \left[\mathbf{n}_{1,N_T^{Beam}}^T, \dots, \mathbf{n}_{N_R^{Block},N_T^{Beam}}^T\right]^T\right]. \end{aligned} \tag{4}$$

The matrix $\mathbf{X} \in \mathbb{C}^{N_T^{Beam} \times N_T^{Beam}}$ is a diagonal matrix with x_p on its diagonal. Throughout the paper, we assume identical pilot symbols so that $\mathbf{X} = \sqrt{P}\mathbf{I}_{N_T^{Beam}}$ where P is the pilot power.

In mmWave communication, hybrid MIMO architecture is employed. The hybrid beamforming matrix as shown in (3) can be decomposed as $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ and $\mathbf{W} = \mathbf{W}_{RF}\mathbf{W}_{BB}$, where $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_T^{Beam}}$ and $\mathbf{W}_{RF} \in \mathbb{C}^{N_R \times N_R^{Beam}}$ represent the RF beamforming matrices, $\mathbf{F}_{BB} \in \mathbb{C}^{N_T^{Beam} \times N_T^{Beam}}$ and $\mathbf{W}_{BB} \in \mathbb{C}^{N_R^{Beam} \times N_R^{Beam}}$ represent the baseband processing matrices. In this case, (3) can be formulated as

$$\mathbf{Y} = \sqrt{P}(\mathbf{W}_{RF}\mathbf{W}_{BB})^H \mathbf{H}(\mathbf{F}_{RF}\mathbf{F}_{BB}) + \mathbf{N}. \tag{5}$$

RF beamforming design and baseband processing design will be discussed in Section 5.

The mmWave channel can be approximated by a narrowband physical channel model with L scatters due to its limited scattering feature [5]. Each scatterer contributes only one path of propagation between BS and MS. The channel matrix can be written as

$$\mathbf{H} = \sqrt{\frac{N_T N_R}{L}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}_R(\theta_\ell^r) \mathbf{a}_T^H(\theta_\ell^t), \tag{6}$$

where α_ℓ is the complex gain of the l^{th} path, θ_l^r and θ_l^t are the AoA and AoD of the l^{th} path, respectively. $\mathbf{a}_T(\theta_l^t)$ and $\mathbf{a}_R(\theta_l^r)$

are array response vector for BS and MS. Assuming that we use N_T and N_R uniform linear array (ULA), $\mathbf{a}_T(\theta_l^t)$ and $\mathbf{a}_R(\theta_l^r)$ can be given by

$$\begin{aligned}\mathbf{a}_T(\theta_l^t) &= \left[1, e^{-j2\pi\frac{d}{\lambda}\cos\theta_l^t}, \dots, e^{-j2\pi\frac{d}{\lambda}\cos\theta_l^t(N_T-1)}\right]^T,\\ \mathbf{a}_R(\theta_l^r) &= \left[1, e^{-j2\pi\frac{d}{\lambda}\cos\theta_l^r}, \dots, e^{-j2\pi\frac{d}{\lambda}\cos\theta_l^r(N_R-1)}\right]^T,\end{aligned} \tag{7}$$

where d denotes the antenna spacing, λ denotes the wavelength of operation. In this paper, we consider $d = \frac{\lambda}{2}$. The channel gains $\{\alpha_\ell\}_{\ell=1}^L$ are modelled by i.i.d. random variables with distribution $\mathcal{CN}(0, \sigma^2)$. The AoAs and AoDs are modelled by a Laplacian distribution whose mean is uniformly distributed over $[0, \pi)$, and the angular standard deviation is σ_{AS}. (6) can be rewritten in matrix form as

$$\mathbf{H} = \mathbf{A}_R\mathbf{H}_a\mathbf{A}_T^H, \tag{8}$$

where $\mathbf{H}_a = \sqrt{\frac{N_TN_R}{L}}\text{diag}(\alpha_1, \dots, \alpha_\ell, \dots, \alpha_L)$, $\mathbf{A}_R = [\mathbf{a}_r(\theta_1^r), \dots, \mathbf{a}_r(\theta_\ell^r), \dots, \mathbf{a}_r(\theta_L^r)] \in \mathbb{C}^{N_R\times L}$, and $\mathbf{A}_T = [\mathbf{a}_t(\theta_1^t), \dots, \mathbf{a}_t(\theta_\ell^t), \dots, \mathbf{a}_t(\theta_L^t)] \in \mathbb{C}^{N_T\times L}$.

To apply CS techniques to channel estimation, virtual channel representation is used. Specifically, we assume that all the angles fall onto a set of discrete angles called grid. In this paper, we choose uniform grid as $[\frac{\pi}{2G}, \frac{\pi}{2G} + \frac{\pi}{G}, \dots, \frac{\pi}{2G} + \frac{\pi}{G}(G-1)]$, and $G \gg L$ to achieve the desired resolution. Using discrete angle grid, the channel matrix $\mathbf{H}$ in (3) can be approximated as

$$\mathbf{H} = \mathbf{A}_R\mathbf{H}_\mathcal{A}\bar{\mathbf{A}}_T^H, \tag{9}$$

where $\bar{\mathbf{A}}_R = [\mathbf{a}_R(\frac{\pi}{2G}), \mathbf{a}_R(\frac{\pi}{2G} + \frac{\pi}{G}), \dots, \mathbf{a}_R(\frac{\pi}{2G} + \frac{\pi}{G}(G-1))] \in \mathbb{C}^{N_R\times G}$, $\bar{\mathbf{A}}_T = [\mathbf{a}_T(\frac{\pi}{2G}), \mathbf{a}_T(\frac{\pi}{2G} + \frac{\pi}{G}), \dots, \mathbf{a}_T(\frac{\pi}{2G} + \frac{\pi}{G}(G-1))] \in \mathbb{C}^{N_T\times G}$ and $\mathbf{H}_\mathcal{A} \in \mathbb{C}^{G\times G}$ is an L-sparse channel gain matrix. The virtual channel representation is not exactly equal to the real channel matrix $\mathbf{H}$ because of the quantized grid error as demonstrated in the simulation in [17].

Considering the system model in (5) and channel model in (6), the mmWave channel estimation problem can be formulated as a sparse signal recovery problem by vectorizing $\mathbf{Y}$ in (2). Using the property of Khatri-Rao product $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{B})$ for $\mathbf{Y}$ and $\mathbf{H}$, we can get

$$\begin{aligned}\mathbf{y}_v &= \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H) \cdot \text{vec}(\mathbf{H}) + \text{vec}(\mathbf{N})\\ &= \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H)\text{vec}(\bar{\mathbf{A}}_R\mathbf{H}_\mathcal{A}\bar{\mathbf{A}}_T^H) + \mathbf{n}_\mathcal{Q}\\ &= \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H)\mathbf{A}_D\mathbf{h}_\mathcal{A} + \mathbf{n}_\mathcal{Q}\\ &= \mathbf{Q}\mathbf{h}_\mathcal{A} + \mathbf{n}_\mathcal{Q},\end{aligned} \tag{10}$$

where $\mathbf{y}_v \in \mathbb{C}^{N_T^{Beam}N_R^{Beam}\times 1}$ is the vectorized received signal. $\mathbf{Q} = \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H)\mathbf{A}_D \in \mathbb{C}^{N_T^{Beam}N_R^{Beam}\times G^2}$ is the sensing matrix. $\mathbf{A}_D = \bar{\mathbf{A}}_T^* \otimes \bar{\mathbf{A}}_R$ is a dictionary matrix that consists of the G^2 column vectors of the form $\mathbf{a}_T^H(\theta_u) \otimes \mathbf{a}_R(\theta_v)$, with θ_u and θ_v, the u^{th} and v^{th} points, respectively, of the angle uniform grid. $\mathbf{h}_\mathcal{A} = \text{vec}(\mathbf{H}_\mathcal{A})$ represents the path gains of the corresponding quantized directions. Because $\mathbf{h}_\mathcal{A} \in \mathbb{C}^{G^2\times 1}$ is an $L(L \ll G^2)$ sparse vector, it can be recovered by CS algorithms.

3 | PROPOSED MODEL FOR EXPLOITING AOA ANGULAR SPREAD IN MMWAVE CHANNEL ESTIMATION

3.1 | System model and formulation of mmWave channel estimation problem

We consider that each scatterer contributes only one path of propagation and the AoAs have angular spreads. The continuous angular spreads are modelled as M grid points length blocks with discrete angle grid G. M can be approximated by rounding up $(\frac{\theta_s}{180°}G - 0.5)$ where θ_s is the AoA angular spread in degree based on real-world measurements [12]. Then the channel model (6) can be formulated as

$$\mathbf{H} = \sqrt{\frac{N_TN_R}{LM}}\sum_{\ell=1}^{L}\sum_{m=1}^{M}\alpha_{\ell,m}\mathbf{a}_R(\theta_{\ell,m}^r)\mathbf{a}_T^H(\theta_\ell^t), \tag{11}$$

where θ_l^t is the AoD of the ℓ^{th} path, $\{\theta_{\ell,m}^r\}_{m=1}^M$ are the discrete AoA points for angular spread of the ℓ^{th}, $\alpha_{\ell,m}$ is the complex path gain for the path between θ_ℓ^t and $\theta_{\ell,m}^r$. Considering the sparsity of the mmWave channel, the limited number of the clusters are unlikely to overlap with each other. So it is reasonable to make the non-overlapping assumption to simplify the channel estimation problem. The overlapping scenario will also be discussed briefly in Section 5. In summary, we assume that the AoA angular spread of different paths do not overhead each other and each AoA angular spread is approximated to a predetermined group pattern as shown in Figure 2 where the number of columns and rows represent the grid points for AoDs and AoAs, respectively. Then, (11) can be rewritten as a matrix form

$$\mathbf{H} = \mathbf{B}_R\mathbf{H}_b\mathbf{B}_T^H, \tag{12}$$

where

$$\begin{aligned}\mathbf{B}_R &= [\mathbf{B}_r(\theta_1^r), \dots, \mathbf{B}_r(\theta_\ell^r), \dots, \mathbf{B}_r(\theta_L^r)] \in \mathbb{C}^{N_R\times ML},\\ \mathbf{B}_r(\theta_\ell^r) &= [\mathbf{a}_R(\theta_{\ell,1}^r), \dots, \mathbf{a}_R(\theta_{\ell,m}^r), \dots, \mathbf{a}_R(\theta_{\ell,M}^r)] \in \mathbb{C}^{N_R\times M},\\ \mathbf{B}_T &= [\mathbf{B}_t(\theta_1^t), \dots, \mathbf{B}_t(\theta_\ell^t), \dots, \mathbf{B}_t(\theta_L^t)] \in \mathbb{C}^{N_T\times ML},\\ \mathbf{B}_t(\theta_\ell^t) &= [\mathbf{a}_T(\theta_\ell^t), \dots, \mathbf{a}_T(\theta_\ell^t), \dots, \mathbf{a}_T(\theta_\ell^t)] \in \mathbb{C}^{N_T\times M},\end{aligned}$$

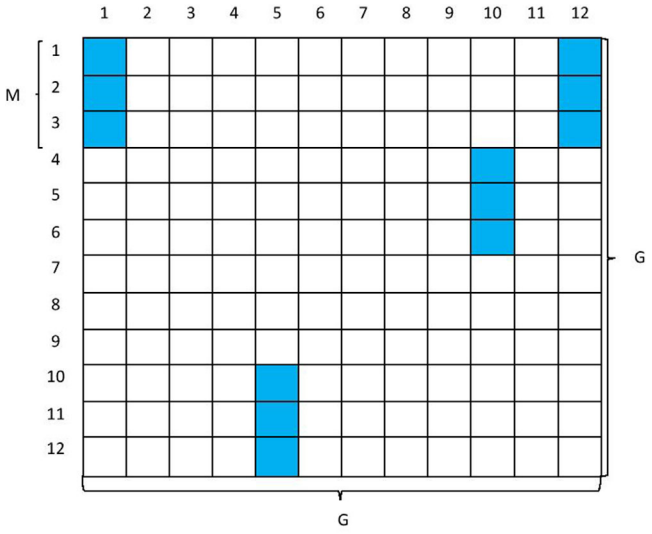

FIGURE 2 Block sparse structure of $\mathbf{H}_B$

$$\mathbf{H}_b = \text{diag}(\boldsymbol{\alpha}_b(1), \dots, \boldsymbol{\alpha}_b(\ell), \dots, \boldsymbol{\alpha}_b(L)) \in \mathbb{C}^{ML\times ML},$$
$$\boldsymbol{\alpha}_b(\ell) = \text{diag}(\alpha_{\ell,1}, \dots, \alpha_{\ell,m}, \dots, \alpha_{\ell,M}) \in \mathbb{C}^{M\times M}. \quad (13)$$

We choose G as times of ML. (12) can be represented as

$$\mathbf{H} = \bar{\mathbf{A}}_R \mathbf{H}_B \bar{\mathbf{A}}_T^H, \quad (14)$$

where $\mathbf{H}_B \in \mathbb{C}^{G\times G}$ is an ML-sparse channel gain matrix.

To apply compressive algorithms, we vectorize the received signal $\mathbf{Y}$, as we did in (10), and we get

$$\mathbf{y}_v = \mathbf{Q}\mathbf{h}_B + \mathbf{n}_Q, \quad (15)$$

where $\mathbf{Q}$ and $\mathbf{n}_Q$ are the same as in (10). $\mathbf{h}_B = \text{vec}(\mathbf{H}_B)$ is an $G^2 \times 1$ vector with block sparsity. (15) is a block sparse signal recovery problem. Structure CS methods can be leveraged to recover $\mathbf{h}_B$ from noisy received signal $\mathbf{y}_v$.

3.1.1 | Block orthogonal matching pursuit method for mmWave MIMO channels

The standard sparsity model in the conventional sense assumes that non-zero elements can appear anywhere in $\mathbf{h}_B$ [18]. As discussed in block-sparse model [19], the non-zero entries of $\mathbf{h}_B$ appear in blocks rather than arbitrarily spread over the vector. We assume that the vector $\mathbf{h}_B \in \mathbb{C}^{G^2\times 1}$ is a concatenation of $N = \frac{G^2}{M}$ blocks and each block has M elements. The vector $\mathbf{h}_B$ is described as:

$$\mathbf{h}_B = [\mathbf{h}_B^T[1], \mathbf{h}_B^T[2], \dots, \mathbf{h}_B^T[N]]^T, \quad (16)$$

where $\mathbf{h}_B[i] \in \mathbb{C}^{M\times 1}$ for $i = 1, \dots, N$. The vector $\mathbf{h}_B$ has only L non-zero blocks. In mmWave communication, the centre of the blocks appears randomly and these blocks will be adjusted to the nearest block pattern to fit (16). Accordingly, the sensing matrix $\mathbf{Q}$ is divided as a concatenation of N matrix as

$$\mathbf{Q} = [\mathbf{Q}[1], \mathbf{Q}[2], \dots, \mathbf{Q}[N]], \quad (17)$$

where size $\mathbf{Q}[i] \in \mathbb{C}^{N_T^{Beam} N_R^{Beam}\times M}$ for $i = 1, \dots, N$ are termed as blocks.

The block OMP has been proposed for this block sparse recovery problem [19]. In Section 5, simulation results are presented to demonstrate that BOMP achieves better accuracy of estimation than OMP with less complexity. However, the performance of OMP and BOMP will deteriorate at low SNR [5]. In this case, $\mathbf{h}_B$ is overwhelmed by noise, and the support of $\mathbf{h}_B$ detected by classic non-Bayesian based compressive sensing method is inaccurate, leading to the deteriorated performance as demonstrated by the simulation results. In addition, sparsity information is usually unknown in mmWave channel estimation. To increase the accuracy of estimation without sparsity information, a Bayesian based block compressive sensing method exploiting angular spreads is proposed in the next section.

4 | PROPOSED BLOCK BAYESIAN MATCHING PURSUIT BASED MMWAVE CHANNEL ESTIMATION

4.1 | Assumptions for mmWave channel

FBMP method is first proposed for sparse signal recovery in [20]. By making appropriate statistical assumptions according to the characteristics of the mmWave channel, FBMP is employed for the mmWave channel estimation in [9]. In this section, we further consider the characteristic of angular spreads. The extension from FBMP to BFBMP is not straightforward. New assumptions and metric update methods are proposed.

In order to apply the Bayesian based compressive sensing method to estimate the mmWave channel, appropriate statistic assumptions need to be made according to the characteristics of mmWave channel. The noise $\mathbf{n}_Q$ in (15) is assumed to be white circular Gaussian with variance σ^2, that is, $\mathbf{n}_Q \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{M_s})$ where $M_s = N_T^{Beam} N_R^{Beam}$ is the number of measurements. $\{h_i\}_{i=1}^{N_s}$ are the elements in sparse vector $\mathbf{h}_B$ where $N_s = G^2$ is the number of elements in channel matrix. We assume that $\{h_i\}_{i=1}^{N_s}$ are drawn from two specific Gaussian distributions. Considering the block structure shown in (16), the block sparsity can be explicitly expressed as

$$\mathbf{h}_B[n] = \mathbf{h}_B^c[n] s_n, \quad (18)$$

where $\mathbf{h}_B^c[n] = \{h_i\}_{i=M(n-1)}^{Mn}$ are the channel coefficients of the n^{th} block. $s_n \in \{0, 1\}$ is a binary index used as a mixture parameter for the distribution of the n^{th} block

$$Pr(s_n = t) = \begin{cases} p_1, & \text{for } t = 1, \\ 1 - p_1, & \text{for } t = 0. \end{cases} \quad (19)$$

$\{s_n\}_{n=1}^N$ are treated as i.i.d random variables as $\Pr\{s_n = 1\} = p_1$ $(0 < p_1 \leq 1)$. $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ is the mixture vector of N blocks. p_1 is the probability that the channel coefficients in the n^{th} block follow Gaussian distribution indexed by $s_n = 1$. When $s_n = 0$, $(\mu_0, \sigma_0^2) = (0, 0)$ is set to make sure that $\mathbf{h}_B[n] = 0$. When $s_n = 1$, $(\mu_1, \sigma_1^2) = (0, 100)$ is set as the assumed distribution for the active non-zero coefficients. In fact, σ_1^2 can be any positive value. We set 100 because relative large variance can improve the accuracy. Simulation based analysis shows that variance larger than 100 would not improve performance further [9]. More details will be discussed in the following derivation. We make $p_1 \ll 1$ to ensure the sparsity. $\mathbf{z}_n$ is the support of the n^{th} block

$$\mathbf{z}_n = s_n \otimes \mathbf{I}_M, \tag{20}$$

where $\mathbf{I}_M \in \mathbb{C}^{M\times 1}$ is a vector with all entries equal to 1. Considering $\mathbf{h}_B = [\mathbf{h}_B^T[1], \mathbf{h}_B^T[2], \ldots, \mathbf{h}_B^T[N]]^T$ and $\mathbf{z} = [\mathbf{z}[1]^T, \mathbf{z}[2]^T, \ldots, \mathbf{z}[N]^T]^T$ is the support pattern (SP) of $\mathbf{h}_B$, the priors can be written as

$$\mathbf{h}_B \mid \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}(\mathbf{z})), \tag{21}$$

where covariance matrix $\mathbf{R}(\mathbf{s})$ has the structure-property, that is,

$$\mathbf{R}(\mathbf{z}) = \begin{bmatrix} \sigma_{s_1}^2\mathbf{I}_M & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \sigma_{s_2}^2\mathbf{I}_M & \ldots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \sigma_{s_n}^2\mathbf{I}_M \end{bmatrix}_{MN\times MN}. \tag{22}$$

Considering (15), the channel vector $\mathbf{h}_B$ and the received signal $\mathbf{y}_v$ are joint Gaussian conditioned on the mixture parameters $\mathbf{z}$ as

$$\begin{bmatrix} \mathbf{y}_v \\ \mathbf{h}_B \end{bmatrix} \Big| \mathbf{z} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{\Phi}(\mathcal{Z}) & \mathbf{Q}R(\mathcal{Z}) \\ \mathbf{R}(\mathbf{z})\mathbf{Q}^H & \mathbf{R}(\mathbf{z}) \end{bmatrix} \right), \tag{23}$$

where

$$\mathbf{\Phi}(\mathcal{Z}) \triangleq \mathbf{Q}\mathbf{R}(\mathbf{z})\mathbf{Q}^H + \sigma^2\mathbf{I}_M. \tag{24}$$

4.1.1 | MMSE coefficient estimation

For channel estimation, MMSE estimate of $\mathbf{h}_B$ from $\mathbf{y}_v$ is

$$\hat{\mathbf{h}}_{mmse} \triangleq \mathrm{E}\{\mathbf{h}_B|\mathbf{y}_v\} = \sum_{\mathbf{z}\in\mathbf{Z}} p(\mathbf{z}|\mathbf{y}_v)\mathrm{E}\{\mathbf{h}_B|\mathbf{y}_v, \mathbf{z}\}. \tag{25}$$

From (23) it is straightforward [21] to obtain

$$\mathrm{E}\{\mathbf{h}_B|\mathbf{y}_v, \mathbf{z}\} = \mathbf{R}(\mathbf{z})\mathbf{Q}^H\mathbf{\Phi}(\mathcal{Z})^{-1}\mathbf{y}_v. \tag{26}$$

We collect the set of all possible SPs in the matrix $\mathbf{Z}$. $p(\mathbf{z}|\mathbf{y}_v)_{\mathbf{z}\in Z}$, (13) can be calculated when all the possible posterior probabilities are known. Although employing block structure is able to reduce the number of possible SPs from 2^{MN} to 2^N, it remains impractical to compute all possible 2^N posterior probability. Fortunately, the size of $\mathbf{Z}_\Omega$ which includes the SPs with non-negligible posterior probability $p(\mathbf{z}|\mathbf{y}_v)_{\mathbf{z}\in\mathbf{Z}_\Omega}$ can be small and practical to compute because of the sparsity. Making use of the dominant SPs in $\mathbf{Z}_\Omega$ yields the approximate MMSE estimate

$$\hat{\mathbf{h}}_{\text{ammse}} \triangleq \mathrm{E}\{\mathbf{h}_B|\mathbf{y}_v\} = \sum_{\mathbf{z}\in\mathbf{Z}_\Omega} p(\mathbf{z}|\mathbf{y}_v)\mathrm{E}\{\mathbf{h}_B|\mathbf{y}_v, \mathbf{z}\}. \tag{27}$$

We first leverage a fast method to search for $\mathbf{Z}_\Omega$.

Search for dominant SPs

We search for $\mathbf{Z}_\Omega$ by selecting $\mathbf{z} \in \mathbf{Z}$ with the significant posterior probability $p(\mathbf{z}|\mathbf{y}_v)$. According to Bayesian rule, the posterior probability can be written as

$$p(\mathbf{z}|\mathbf{y}_v) = \frac{p(\mathbf{y}_v|\mathbf{z})p(\mathbf{z})}{\sum_{\mathbf{z}'\in\mathbf{Z}} p(\mathbf{y}_v|\mathbf{z}')p(\mathbf{z}')}, \tag{28}$$

where $p(\mathbf{z}|\mathbf{y}_v)$ are equal to $p(\mathbf{y}_v|\mathbf{z})p(\mathbf{z})$ up to a scale. For convenience, we work in logarithm domain and define $\alpha(\mathbf{z}, \mathbf{y}_v)$ as SP selection metric:

$$\begin{aligned} \alpha(\mathbf{z}, \mathbf{y}_v) &\triangleq \ln p(\mathbf{y}_v|\mathbf{z})p(\mathbf{z}) \\ &= \ln\left(\frac{1}{(2\pi)^{\frac{M_s}{2}}|\mathbf{\Phi}(\mathcal{Z})|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}_v^H\mathbf{\Phi}(\mathcal{Z})^{-1}\mathbf{y}_v\right) \right. \\ &\quad \left. p_1^L(1-p_1)^{N-L} \right) \\ &= -\frac{M_s}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Phi}(\mathcal{Z})| - \frac{1}{2}\mathbf{y}_v^H\mathbf{\Phi}(\mathcal{Z})^{-1}\mathbf{y}_v \\ &\quad + \frac{\|\mathbf{z}\|_0}{M}\ln\frac{p_1}{(1-p_1)} + N\ln(1-p_1). \end{aligned} \tag{29}$$

The significant $p(\mathbf{z}|\mathbf{y}_v)$ corresponds to the significant value of $\alpha(\mathbf{s}, \mathbf{y}_v)$. As a result, $\mathbf{Z}_\Omega$ can be selected based on metric $\alpha(\mathbf{z}, \mathbf{y}_v)$ using the non-exhaustive tree search method.

The search starts with $\mathbf{z} = 0$. In the first stage, only one block elements $\mathbf{z}_n$ is changed to 1. It has N different 'one block elements active' SP. All these possible SPs are stored as $\mathbf{Z}^{(1)}$ and the metric of these SPs can be calculated. We choose D SPs with the largest metrics and store them as $\mathbf{Z}_\Omega^{(1)}$. In the second step, we activate one more block elements from the D chosen SPs in $\mathbf{Z}_\Omega^{(1)}$ so that we have $(N-1) + (N-2) + \cdots + (N-D)$ possible 'two block elements active' SPs in $\mathbf{Z}^{(2)}$. Then we choose D SPs with the largest metrics among these possible SPs and store them as $\mathbf{Z}_\Omega^{(2)}$. We do this procedure J times to get D 'J block elements active' SPs with the largest posterior possibility as candidate SPs.

The value of D is fixed and chosen as 5 because simulation shows the benefits of increasing D diminish quickly for $D > 5$. The value of J is determined by the sparsity of the channel. However, we do not know the real sparsity of mmWave channel. So we define a virtual sparsity L'. We choose an arbitrary small integer (from 5 to 10) as the virtual sparsity. Based on virtual sparsity, p_1 can be calculated as: L'/N, L' follows Binomial (N, p_1) distribution. Simulation results will show that our proposed algorithm is able to achieve superior performance without the need to know the real sparsity.

Fast metric update

In the above search, metric α needs to be calculated for each possible SP. We adopted a fast metric update method based on [20] to reduce the computational complexity.

For the case that $\mathbf{z}[n] = \mathbf{0}$ and $\mathbf{z}_{\text{new}}[n] = \mathbf{1}$, where $\mathbf{z}$ and $\mathbf{z}_{\text{new}}$ are identical except for the coefficients in the n^{th} block. For brevity, we use $\Delta_n(\mathbf{z}, \mathbf{y}_v) \triangleq \alpha(\mathbf{z}_{\text{new}}, \mathbf{y}_v) - \alpha(\mathbf{z}, \mathbf{y}_v)$ below. According to (29), the root node ($\mathbf{Z}_{\Omega}^{(0)} = \mathbf{0}$) has the following metric

$$\alpha(\mathbf{0}, \mathbf{y}_v) = -\frac{M_s}{2}\ln(2\pi) - M_s \ln \sigma_n - \frac{1}{2\sigma_n^2}\|\mathbf{y}_v\|_2^2 + L\ln(1-p_1). \tag{30}$$

When SP is updated, the primary challenge in the computation of metrics is to obtain $\mathbf{\Phi}(\mathbf{z}_{\text{new}})$ and $\mathbf{\Phi}(\mathbf{z}_{\text{new}})^{-1}$. First of all, we compute $\mathbf{\Phi}(\mathbf{z}_{\text{new}})$ due to the support update. For any n and $\mathbf{z}_{\text{new}}$, we have

$$\begin{aligned}\mathbf{\Phi}(\mathbf{z}_{\text{new}}) &= \mathbf{Q}\mathbf{R}(\mathbf{z}_{\text{new}})\mathbf{Q}^H + \sigma^2\mathbf{I}_{M_s} \\ &= \mathbf{Q}\mathbf{R}(\mathbf{z})\mathbf{Q}^H + \sigma^2\mathbf{I}_{M_s} + \sum_{i=(n-1)M+1}^{Mn}\sigma_1^2\mathbf{q}_i\mathbf{q}_i^H \\ &= \mathbf{\Phi}(\mathbf{z}) + \sigma_1^2\mathbf{Q}[n]\mathbf{Q}[n]^H,\end{aligned} \tag{31}$$

where $\mathbf{q}_n$ is the n^{th} column of $\mathbf{Q}$. $\mathbf{Q}[n]$ is defined in (17). The matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$ implies

$$\begin{aligned}\mathbf{\Phi}(\mathbf{z}_{\text{new}})^{-1} &= \mathbf{\Phi}(\mathbf{z})^{-1} - \mathbf{\Phi}(\mathbf{z})^{-1}\mathbf{Q}[n]\left(\frac{1}{\sigma_1^2}\mathbf{I}_M + \mathbf{Q}[n]^H\mathbf{\Phi}(\mathbf{z})^{-1}\mathbf{Q}[n]\right)^{-1}\mathbf{Q}[n]^H\mathbf{\Phi}(\mathbf{z})^{-1} \\ &= \mathbf{\Phi}(\mathbf{z})^{-1} - \mathbf{C}_n\mathbf{\Pi}_n\mathbf{C}_n^H,\end{aligned} \tag{32}$$

where

$$\mathbf{C}_n \triangleq \mathbf{\Phi}(\mathbf{z})^{-1}\mathbf{Q}[n], \tag{33}$$

$$\mathbf{\Pi}_n \triangleq \left(\frac{1}{\sigma_1^2}\mathbf{I}_M + \mathbf{Q}[n]^H\mathbf{C}_n\right)^{-1}. \tag{34}$$

According to (33), we can observe that the update of $\mathbf{C}_l$ includes matrix inversion which has high complexity. Fortunately, the previous information can be exploited. We assume that $\mathbf{z}$ is the SP which is obtained from changing $\mathbf{z}_{\text{pre}}$. $\mathbf{z}$ and $\mathbf{z}_{\text{pre}}$ are identical except for the coefficients in the $n_{\text{pre}}^{\text{th}}$ block that $\mathbf{z}_{\text{pre}}[n_{\text{pre}}] = \mathbf{0}$ and $\mathbf{z}[n_{\text{pre}}] = \mathbf{1}$. Based on (32), we have

$$\mathbf{\Phi}(\mathbf{z})^{-1} = \mathbf{\Phi}(\mathbf{z}_{\text{pre}})^{-1} - \mathbf{C}_{n_{\text{pre}}}^{\text{pre}}\mathbf{\Pi}_{n_{\text{pre}}}^{\text{pre}}\mathbf{C}_{n_{\text{pre}}}^{\text{pre}\,H}, \tag{35}$$

so that $\mathbf{C}_n$ can be calculated by applying previous information as

$$\begin{aligned}\mathbf{C}_n &= \left(\mathbf{\Phi}(\mathbf{z}_{\text{pre}})^{-1} - \mathbf{C}_{n_{\text{pre}}}^{\text{pre}}\mathbf{\Pi}_{n_{\text{pre}}}^{\text{pre}}\mathbf{C}_{n_{\text{pre}}}^{\text{pre}\,H}\right)\mathbf{Q}[n] \\ &= \mathbf{C}_n^{\text{pre}} - \mathbf{C}_{n_{\text{pre}}}^{\text{pre}}\mathbf{\Pi}_{n_{\text{pre}}}^{\text{pre}}\mathbf{C}_{n_{\text{pre}}}^{\text{pre}\,H}\mathbf{Q}[n],\end{aligned} \tag{36}$$

where

$$\mathbf{C}_n^{\text{pre}} = \mathbf{\Phi}(\mathbf{z}_{\text{pre}})^{-1}\mathbf{Q}[n], \tag{37}$$

$$\mathbf{C}_{n_{\text{pre}}}^{\text{pre}} = \mathbf{\Phi}(\mathbf{z}_{\text{pre}})^{-1}\mathbf{Q}[n_{\text{pre}}], \tag{38}$$

$$\mathbf{\Pi}_{n_{\text{pre}}}^{\text{pre}} = \left(\frac{1}{\sigma_1^2}\mathbf{I}_M + \mathbf{Q}[n_{\text{pre}}]^H\mathbf{\Phi}(\mathbf{z}_{\text{pre}})^{-1}\mathbf{Q}[n_{\text{pre}}]\right)^{-1}. \tag{39}$$

To this end, we are able to calculate metrics fast and we have

$$\begin{aligned}\alpha(\mathbf{z}_{\text{new}}) &= -\frac{M_s}{2}\ln(2\pi) - \frac{1}{2}\ln|(\mathbf{\Phi}(\mathbf{z}_{\text{new}})| \\ &\quad -\frac{1}{2}\mathbf{y}_v^H\mathbf{\Phi}(\mathbf{z}_{\text{new}})^{-1}\mathbf{y}_v + \frac{\|\mathbf{z}_{\text{new}}\|_0}{M}\ln\frac{p_1}{(1-p_1)} \\ &\quad +N\ln(1-p_1) \\ &= -\frac{M_s}{2}\ln(2\pi) - \frac{1}{2}\left(\ln|(\mathbf{\Phi}(\mathbf{z})| + M\ln\sigma_1^2 - \ln|\mathbf{\Pi}_n|\right) \\ &\quad -\frac{1}{2}\left(\mathbf{y}_v^H\mathbf{\Phi}(\mathbf{z})^{-1}\mathbf{y}_v - \mathbf{y}_v^H\mathbf{C}_n\mathbf{\Pi}_n\mathbf{C}_n^H\mathbf{y}_v\right) \\ &\quad +\left(\frac{\|\mathbf{z}\|_0}{M}\ln\frac{p_1}{(1-p_1)} + \ln\frac{p_1}{(1-p_1)}\right) \\ &\quad +N\ln(1-p_1) \\ &= \alpha(\mathbf{z}) + \Delta_n(\mathbf{z}),\end{aligned} \tag{40}$$

where

$$\Delta_n(\mathbf{z}) = -\frac{M}{2}\ln\sigma_1^2 + \frac{1}{2}\mathbf{y}_v^H\mathbf{C}_n\mathbf{\Pi}_n\mathbf{C}_n^H\mathbf{y}_v + \frac{1}{2}\ln|\mathbf{\Pi}_n| + \ln\frac{p_1}{(1-p_1)}. \tag{41}$$

$\Delta_n(\mathbf{z})$ quantifies the change to $\alpha(\mathbf{z})$ corresponding to the change of the coefficients in $\mathbf{z}[n]$ from $\mathbf{0}$ to $\mathbf{1}$. In this way, once the SP

is updated, the metric of new SP can be fast computed base on the metric of the previous SP.

In summary, the proposed block Bayesian Matching Pursuit based method is a non-exhaustive tree-search using the SP selection metric (29) with a fast metric update method. The algorithm is shown in Algorithm 2, where the approximate posterior probability of $\mathbf{z}$ is estimated as

$$p(\mathbf{z}|\mathbf{y}_v) = \frac{\exp\{\alpha(\mathbf{z}, \mathbf{y}_v)\}}{\sum_{\mathbf{z}' \in \mathbf{Z}} \exp\{\alpha(\mathbf{z}', \mathbf{y}_v)\}} \approx \frac{\exp\{\alpha(\mathbf{z}, \mathbf{y}_v)\}}{\sum_{\mathbf{z}' \in \mathbf{Z}_\Omega} \exp\{\alpha(\mathbf{z}', \mathbf{y}_v)\}}. \tag{42}$$

According to the characteristics of mmWave channel, $(\mu_0, \sigma_0^2) = (0, 0)$, $(\mu_1, \sigma_1^2) = (0, 100)$, $D = 5$, $L' = 6$, $p_1 = L'/N$ and $J = L'$ are applied.

In Algorithm 2, lines 1 and 2 are the initialization. Lines 3–6 compute the metric when only one block is active. Steps 7–24 update the metrics with the fast method and apply a tree search for significant SPs. After obtaining D candidate J elements SPs, we can compute the posterior probability based on (42). At the end, according to (27), the algorithm would return the channel approximate MMSE estimate $\hat{\mathbf{h}}_{\text{ammse}}$.

5 | SIMULATION RESULTS

In this section, computer simulations are presented to evaluate performance of the proposed methods. All the simulations are averaged over 500 channel realizations and the system parameters are listed in Table 1. Specifically, N_T, N_R, N_T^{Beam} and N_R^{Beam} are the number of transmitter antennas, receiver antennas, transmit beam patterns and receive beam patterns. L is the number of scatterers. $\alpha_{\ell,m}$ is the Gaussian distributed ($\mathcal{CN}(0, 1)$) complex path gain for the paths between the AoD (θ_ℓ^t) and AoA ($\theta_{\ell,m}^r$) as defined in (11). The AoA and AoD are randomly distributed over $[0, \pi]$. The carrier frequency is 60 GHz. Size of the set of discrete angles is 64 as defined in (9). And we assume that the angular spreads are randomly distributed from 9.95° to 12.78°.

We use phase shifts to generate DFT beams for analog beamforming. $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$ can be designed as DFT matrices. We use the approach in [7] to design the precoding matrix for baseband through minimizing the coherence of sensing matrix $\mathbf{Q}$. Specifically, $\mathbf{F}_{BB}$ and $\mathbf{W}_{BB}$ are block diagonal matrices given by $\mathbf{F}_{BB} = \text{diag}(\mathbf{F}_{BB,1}, \dots, \mathbf{F}_{BB,i}, \dots, \mathbf{F}_{BB,N_T^{block}})$ and $\mathbf{W}_{BB} = \text{diag}(\mathbf{W}_{BB,1}, \dots, \mathbf{W}_{BB,i}, \dots, \mathbf{W}_{BB,N_R^{block}})$ whose diagonal entries, $\mathbf{F}_{BB,i}$ and $\mathbf{W}_{BB,i}$, consist of $N_{RF} \times N_{RF}$ complex valued matrices. $N_R^{Block} = \frac{N_R^{Beam}}{N_{RF}}$ and $N_T^{Block} = \frac{N_T^{Beam}}{N_{RF}}$ are the number of receive blocks and transmit block, respectively. It is shown in [7] that the optimal solution of $\mathbf{W}_{BB}$ and $\mathbf{F}_{BB}$ to minimize coherence of sensing matrix are given by (43) and (44).

$$\mathbf{W}_{BB,i} = \mathbf{U}_1 (\mathbf{\Lambda}_1^{-1/2})^H, 1 \le i \le N_R^{Block}, \tag{43}$$

ALGORITHM 1 block Bayesian Matching Pursuit mmWave Channel Estimation

Input:

Received signal $\mathbf{y}_v$, sensing matrix $\mathbf{Q}$, block length M, number of transmit and receive antenna N_T, N_R, number of transmit and receive beam patterns N_T^{Beam}, N_R^{Beam} and hypotheses of channel statistics σ_1^2, σ_n^2, L';

Output:

Channel approximate MMSE estimate $\hat{\mathbf{h}}_{\text{ammse}}$ in (27);

1: $M_s = N_T^{Beam} N_R^{Beam}$, $N = \frac{N_T N_R}{M}$, $p_1 = \frac{L'}{N}$, $\mathbf{Z}_\Omega = \emptyset$

2: $\alpha^0 = -\frac{M_s}{2}\ln(2\pi) - M_s \ln \sigma_n - \frac{1}{2\sigma_n^2}\|\mathbf{y}_v\|_2^2 + L\ln(1-p_1)$

3: **for** $n = 1 : N$ **do**

4: $\mathbf{C}_n^0 = \mathbf{\Phi}(\mathcal{Z})^{-1}\mathbf{Q}[n]$, $\mathbf{\Pi}_n^0 = (\frac{1}{\sigma_1^2}\mathbf{I}_M + \mathbf{Q}[n]^H \mathbf{C}_n)^{-1}$

5: $\alpha_n^0 = \alpha^0 + -\frac{M}{2}\ln\sigma_1^2 + \frac{1}{2}\mathbf{y}_v^H \mathbf{C}_n \mathbf{\Pi}_n \mathbf{C}_n^H \mathbf{y}_v + \frac{1}{2}\ln|\mathbf{\Pi}_n| + \ln\frac{p_1}{(1-p_1)}$

6: **end for**

7: **for** $d = 1 : D$ **do**

8: $\mathbf{n} = []$, $\hat{\mathbf{z}}^{(d,0)} = \mathbf{0}$,

9: **for** $n = 1 : N$ **do**

10: $\mathbf{C}_n = \mathbf{C}_n^0$, $\mathbf{\Pi}_n = \mathbf{\Pi}_n^0$

11: $\alpha_n = \alpha_n^0$

12: **end for**

13: **for** $j = 1 : J$ **do**

14: $n_* = n$ indexing the largest element in $\{\alpha_n\}_{n=1:N}$ which leads to an as-of-yet unexplored node.

15: $\alpha^{(d,j)} = \alpha_{n_*}$, update $\hat{\mathbf{z}}^{(d,j)} \xleftarrow{\mathbf{z}[n_*]=\mathbf{1}^{M\times 1}} \hat{\mathbf{z}}^{(d,j-1)}$

16: $\mathbf{n} = [\mathbf{n}, n_*]$

17: **for** $n = 1 : N$ **do**

18: Update $\mathbf{C}_n$ via (36)

19: Update $\mathbf{\Pi}_n$ via (34) and $\Delta_n(\mathbf{z})$ via (41)

20: Obtain $\alpha_n = \alpha^{(d,j)} + \Delta_n(\mathbf{z})$

21: **end for**

22: **end for**

23: $\mathbf{Z}_\Omega = \mathbf{Z}_\Omega \cup \hat{\mathbf{z}}^{(d,j)}$

24: **end for**

25: Compute $p(\mathbf{z}|\mathbf{y}_v)$ via (42)

26: Compute estimation $\hat{\mathbf{h}}_{\text{ammse}}$ via (27)

TABLE 1 System parameters in the simulations

Parameters	Values
(N_T, N_N), (N_T^{Beam}, N_R^{Beam})	(32, 32), (32, 32)
Sparsity (L)	$\max\{P_{10}, 1\}$
Channel gain ($\alpha_{\ell,m}$)	$\mathcal{CN}(0,1)$
AoA, AoD	$[0, \pi]$
Carrier frequency	60 GHz
Grid size (G)	64
Size of AoA angular spread	[9.95°, 12.78°]

∗ P_{10} is the outcome of the Poisson random variable with mean 2.The channel gains in each cluster are assumed to have internal coherence as 0.95.

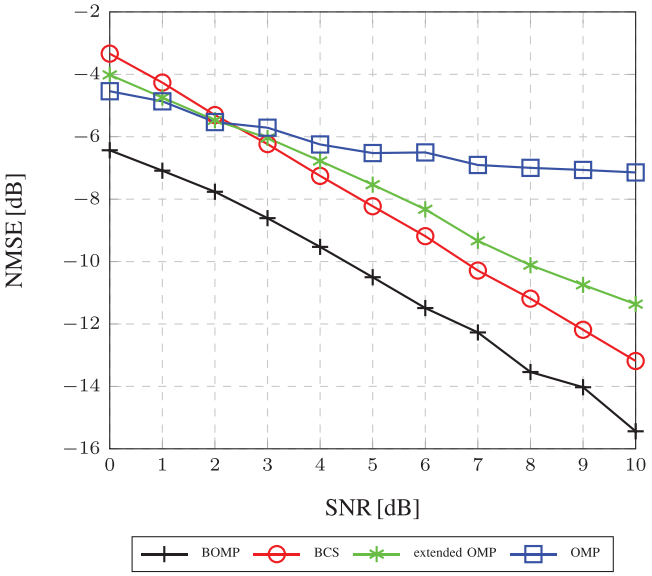

FIGURE 3 NMSE performances (without overlaps) at different SNR [dB]

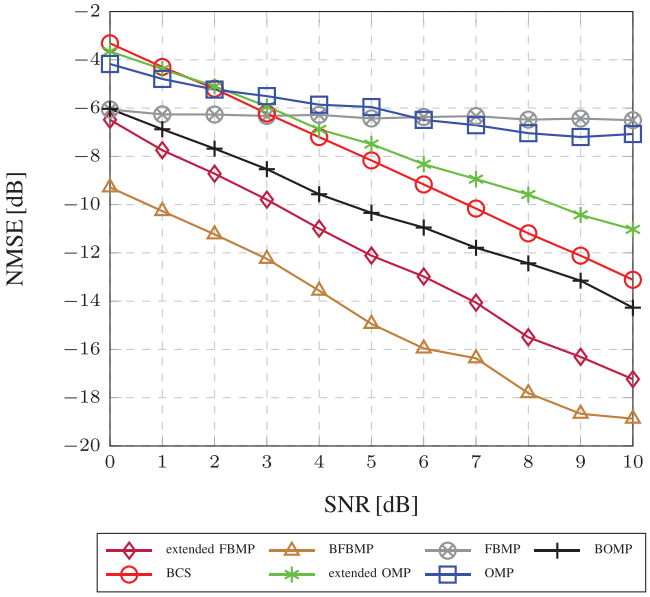

FIGURE 4 NMSE performances (without overlaps) at different SNR [dB]

where $\mathbf{U}_1$ and $\mathbf{\Lambda}_1$ are the matrices of the eigenvectors and eigenvalues, respectively, satisfying $\mathbf{W}_{RF,i}^H \bar{\mathbf{A}}_R \bar{\mathbf{A}}_R^H \mathbf{W}_{RF,i} = \mathbf{U}_1 \mathbf{\Lambda}_1 {\mathbf{U}_1}^H$.

$$\mathbf{F}_{BB,i} = {\mathbf{U}_2}^* ({\mathbf{\Lambda}_2}^{-1/2})^T, 1 \le i \le N_T^{Block}, \tag{44}$$

where $\mathbf{U}_2$ and $\mathbf{\Lambda}_2$ are the matrices of the eigenvectors and eigenvalues, respectively, satisfying $\mathbf{F}_{RF,i}^T \bar{\mathbf{A}}_T^* (\mathbf{F}_{RF,i}^T \bar{\mathbf{A}}_T^*)^H = \mathbf{U}_2 \mathbf{\Lambda}_2 {\mathbf{U}_2}^H$.

In Figure 3, the performance of channel estimation precision is measured by the normalized mean square error (NMSE) defined as $10\log_{10}(\mathbb{E}(\|\mathbf{H} - \mathbf{H}^{estimate}\|_F^2 / \|\mathbf{H}\|_F^2))$. We compare the NMSE performance for OMP, extended OMP, BOMP and BCS. According to real-world measurement [12], we assume that the AoA angular spread is between 9.95° and 12.78° and this results in a block length of $M = 4$ when $G = 64$. OMP takes L (number of non-zero paths) and extended OMP takes ML (number of non-zero elements) as the sparsity. BOMP adapts L as the block sparsity. BCS is included for comparison because it is a Bayesian based learning method that is able to achieve stable performance at all SNR without sparsity information. As shown, the worst performance is achieved by OMP because L is the number of paths which is far less than the number of non-zero elements. Extended OMP takes ML as the sparsity and it achieves much better performance at high SNR with sufficient iterations. However, extended OMP is worse than BOMP at all SNR because it does not take advantages of the block structure. BCS is better than extended OMP at high SNR. In summary, Figure 3 shows that the angular spreads can be utilized to improve the accuracy of the mmWave channel estimation.

Figure 4 adds FBMP, extended FBMP, and the proposed BFBMP into comparison. FBMP and extended FBMP take L' and ML' as virtual sparsity, respectively. BFBMP adopts L' as virtual block sparsity. Considering the sparsity of the mmWave channel, we choose $L' = 6$ in the simulations. In Figure 4, extended FBMP performs better than BOMP at cost of a large computational complexity caused by Bayesian based searching. FBMP has an almost flatten performance because the adopted virtual sparsity is too small to provide an accurate estimation. Among all algorithms, BFBMP has the best performance. It achieves nearly 4 dB better performance than BOMP and 2 dB better performance than extended FBMP at all SNR.

BCS is a learning based method that is difficult to calculate the exactly computational complexity. In order to compare the magnitude of complexity among OMP based methods, FBMP based methods, and Bayesian learning based methods, in Figure 5, non-optimized MATLAB codes are used to show the average runtime of OMP, BCS and extended FBMP as baselines. The computational complexity of OMP based methods and FBMP based methods are straightforward to be verified. From Algorithm 1, the number of multiplications required by FBMP based algorithms is $\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 PD)$. FBMP and BFBMP choose $P = L'$, and extended FBMP adapts $P = ML'$. As a result, extended FBMP has a much larger computational complexity compared with FBMP and BFBMP. BOMP and OMP are on the same order of computational complexity as $\mathcal{O}(LN_N^{Beam} N_R^{Beam} G^2)$. Extended OMP has M times higher complexity as $\mathcal{O}(MLN_N^{Beam} N_R^{Beam} G^2)$ because of M times iterations. Figure 5 shows that BCS is the slowest even with the given noise power. Extended FBMP is faster than BCS but also requires nearly $ML'D/L$ longer time than OMP. Note that, OMP based methods require real sparsity. FBMP based methods only require a range of the possible sparsity and utilize virtual sparsity to achieve accurate channel estimation. BCS does not need any information on sparsity. It can be found that less computational complexity is required with more accurate

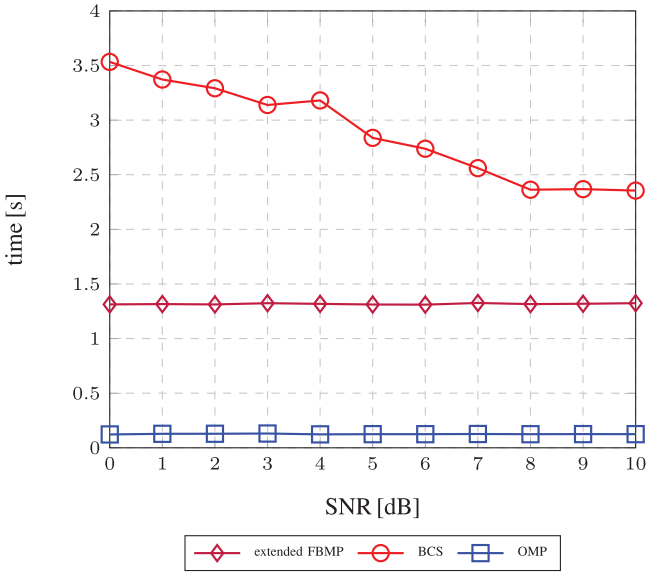

FIGURE 5 Runtime at different SNR [dB]

TABLE 2 Computational complexity analysis of proposed methods

Algorithms	Computational complexity
FBMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 L' D)$
BFBMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 L' D)$
Extended FBMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 ML' D)$
OMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 L)$
BOMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 L)$
Extended OMP	$\mathcal{O}(N_N^{Beam} N_R^{Beam} G^2 ML)$
BCS	Runtime shown as Figure 5

∗ In the mmWave system, we have $D = 5$.

sparsity information. The complexity analysis of the proposed methods is summarized in Table 2.

Figure 6 compares the performance of the above methods with a varying number of scatters L at 4 dB SNR. We assume that $L' = 6$. Although OMP, extended OMP, BOMP have the real sparsity, the performances are slightly worse when sparsity increases because of the growing number of non-zero elements in the channel matrix. BCS is a learning based method which is almost the same at all SNR with varying L. FBMP decreases significantly when the real sparsity L grows. It is because that the virtual sparsity L' is much smaller than the growing real non-zero elements in the channel matrix. Compared with FBMP, the larger virtual sparsity ML' of extended FBMP ensures the algorithm to achieve stable performance. BFBMP achieves the best accuracy of estimation when L grows to 6 which is exactly the same as the assumed virtual block sparsity L'. This best accuracy decreases when the true block sparsity further grows larger than assumed L'. But BFBMP still performs better than most algorithms even at $L = 2$ or $L = 10$.

In the above simulations, we assume that the angular spreads fall in the pre-determined block patterns perfectly without overlaps. It is worth noting that, considering the size of the surface

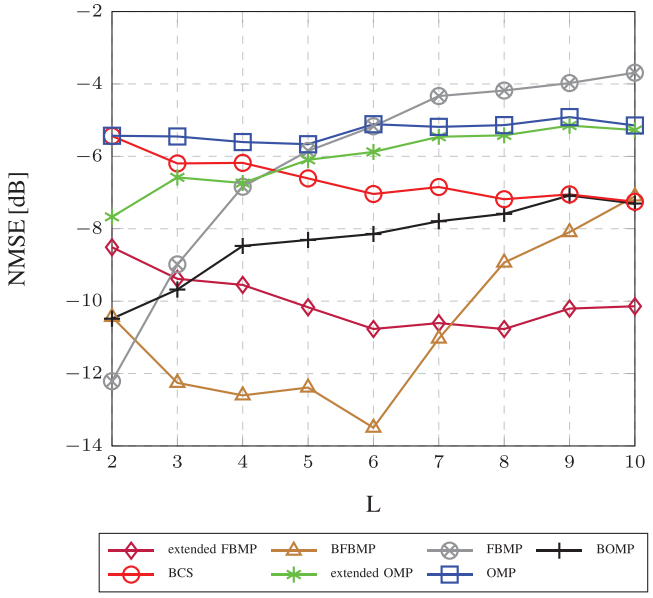

FIGURE 6 NMSE performances (without overlaps) at varying scattering paths

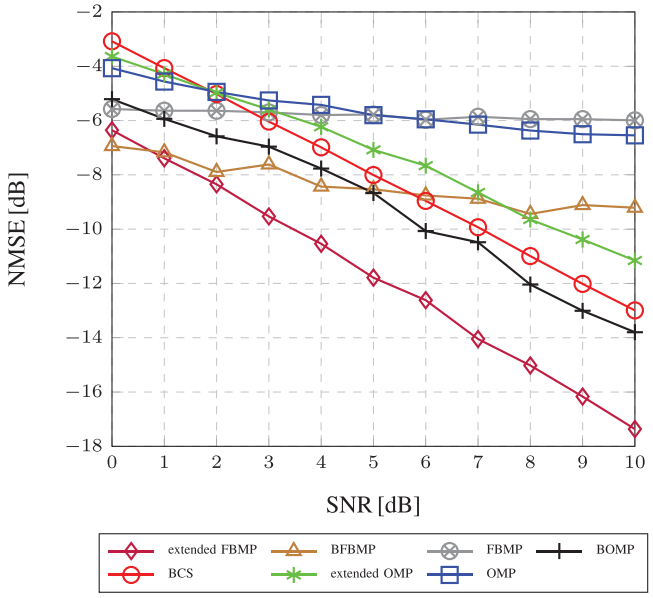

FIGURE 7 NMSE performances (with overlaps) at different SNR (dB)

roughness and the small wavelength of mmWave, the diffusion scattering effect is more obvious and results in non-negligible angular spread (or clustering) feature compared with microwave communications in the limited scattering environment. However, the scattering effect is based on the material properties, type of environment and physical parameters. Thus, a prior information of the angular spread is hard to be acquired. Therefore, in the following simulations, we consider some more complex scenarios (i.e. isotropic scattering environment) which have overlapped angular spreads and off-grid errors. In Figure 7, we consider the scenario that the AoA angular spreads may overlap with others. We use the same parameters as Figure 4 and show

the worst case for comparison. Compared with Figure 4, it can be found that only the performances of BFBMP and BOMP changes and other algorithms remain almost the same performance. In Figure 7, BOMP becomes unstable at all SNR and it has slightly worse NMSE performance compared with Figure 4. But BFBMP performance becomes more unstable and deteriorates seriously at all SNR. Compared with the coherence based support estimation method in BOMP, BFBMP utilizes a Bayesian based searching method which is high precision but more sensitive to the block pattern mismatch. However, BOMP is not able to achieve accurate channel estimation with large noises. As a result, BFBMP still performs better than BOMP when SNR<5 dB.

In summary, when the real sparsity is known and the SNR is high, BOMP is suitable to be employed for the mmWave channel estimation. Otherwise, the proposed BFBMP is more suitable to be employed to achieve better performance with affordable computational complexity.

6 | CONCLUSION

In this paper, we utilized the AoA angular spread feature by formulating the mmWave channel estimation as a block sparse signal recovery problem. BOMP method was first employed to validate the efficiency of the formulation. BFBMP was proposed to further improve the channel estimation performance efficiently without the real sparsity. Simulation results demonstrated that through utilizing angular spreads, BOMP and BFBMP are able to achieve superior channel estimation performance with less computational complexity compared with the conventional methods such as OMP and FBMP.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

You You https://orcid.org/0000-0002-9584-733X
Li Zhang https://orcid.org/0000-0002-4535-3200

REFERENCES

1. Rappaport, T.S., Sun, S., Mayzus, R., Zhao, H., Azar, Y., Wang, K., Wong, G.N., Schulz, J.K., Samimi, M., Gutierrez, F.: Millimeter wave mobile communications for 5G cellular: It will work!. IEEE Access 1, 335–349 (2013)
2. Heath, R.W., Gonzalez-Prelcic, N., Rangan, S., Roh, W., Sayeed, A.M.: An overview of signal processing techniques for millimeter wave MIMO systems. IEEE J. Sel. Top. Signal Process. 10(3), 436–453 (2016)
3. Smulders, P.F., Correia, L.: Characterisation of propagation in 60 GHz radio channels. Electron. Commun. Eng. J. 9(2), 73–80 (1997)
4. Rossi, M., Haimovich, A.M., Eldar, Y.C.: Spatial compressive sensing for MIMO radar. IEEE Trans. Signal Process. 62(2), 419–430 (2014)
5. Alkhateeb, A., El Ayach, O., Leus, G., Heath, R.W.: Channel estimation and hybrid precoding for millimeter wave cellular systems. IEEE J. Sel. Top. Signal Process. 8(5), 831–846 (2014)
6. Sun, S., Rappaport, T.S.: Millimeter wave MIMO channel estimation based on adaptive compressed sensing. In: Proceedings of IEEE International Conference on Communications, pp. 47–53. IEEE, Piscataway (2017)
7. Lee, J., Gil, G.-T., Lee, Y.H.: Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. IEEE Trans. Commun. 64(6), 2370–2386 (2016)
8. Mishra, A., Rajoriya, A., Jagannatham, A.K., Ascheid, G.: Sparse bayesian learning-based channel estimation in millimeter wave hybrid MIMO systems. In: Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications, pp. 1–5. IEEE, Piscataway (2017)
9. You, Y., Zhang, L.: Bayesian matching pursuit based channel estimation for millimeter wave communication. IEEE Commun. Lett. 24(2), 344–348 (2020)
10. Zhang, R., Zhang, J., Zhao, T., Zhao, H.: Block sparse recovery for wideband channel estimation in hybrid mmwave MIMO systems. In: Proceedings of IEEE Global Communication Conference (GLOBECOM), pp. 1–6. IEEE, Piscataway (2018)
11. Liu, J., Li, X., Fang, K., Fan, T.: Millimeter wave channel estimation based on clustering block sparse bayesian learning. In: Proceedings of International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–5. IEEE, Piscataway (2019)
12. Akdeniz, M.R., Liu, Y., Samimi, M.K., Sun, S., Rangan, S., Rappaport, T.S., Erkip, E.: Millimeter wave channel modeling and cellular capacity evaluation. IEEE J. Sel. Areas Commun. 32(6), 1164–1179 (2014)
13. Wang, P., Pajovic, M., Orlik, P.V., Koike-Akino, T., Kim, K.J., Fang, J.: Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread. In: Proceedings of IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, Piscataway (2019)
14. Li, X., Fang, J., Li, H., Wang, P.: Millimeter wave channel estimation via exploiting joint sparse and low-rank structures. IEEE Trans. Wireless Commun. 17(2), 1123–1133 (2018)
15. Liu, K., Li, X., Fang, J., Li, H.: Bayesian mmwave channel estimation via exploiting joint sparse and low-rank structures. IEEE Access 7, 961–970 (2019)
16. Wang, P., Pajovic, M., Orlik, P.: System and method for channel estimation in mmwave communications exploiting joint aod-aoa angular spread. US Patent 10,382,230, 2019
17. You, Y., Zhang, L., Liu, M.: IP aided omp based channel estimation for millimeter wave massive mimo communication. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE, Piscataway (2019)
18. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory 53(12), 4655–4666 (2007)
19. Eldar, Y.C., Kuppinger, P., Bolcskei, H.: Block-sparse signals: Uncertainty relations and efficient recovery. IEEE Trans. Signal Process. 58(6), 3042–3054 (2010)
20. Schniter, P., Potter, L.C., Ziniel, J.: Fast bayesian matching pursuit. In: Proceedings of Information Theory and Applications Workshop, pp. 326–333. IEEE, Piscataway (2008)
21. Poor, H.V.: An Introduction to Signal Detection and Estimation. Springer-Verlag, New York (1994)

How to cite this article: You, Y., Zhang, L.: Exploiting angular spread in channel estimation of millimeter wave MIMO system. IET Commun. 16, 195–205 (2022). https://doi.org/10.1049/cmu2.12329