Digit. Signal Process. 179 (2026) 106189

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Beamspace channel estimation for millimeter-wave MIMO-FBMC systems with lens antenna arrays

Ying Wang [a], Qiang Guo [a], Jianhong Xiang [a,b,*], Yu Zhong [b]

[a] College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001, China
[b] 10th Research Institute, China Electronics Technology Group Corporation, Chengdu, 610036, China

ARTICLE INFO

Keywords:
Millimeter-wave
MIMO-FBMC
Beamspace channel estimation
Multi-d-AMP
Lens antenna array

ABSTRACT

Millimeter-wave Multiple-input Multiple-output (MIMO) is a key technology for enabling ultra-high data rates in wireless communications. However, due to the limited number of RF chains compared to the large number of antennas, beamspace channel estimation faces challenges. In this paper, we propose a channel estimation method based on Multi-dimensional Approximate Message Passing (Multi-d-AMP), which enables accurate estimation of the beamspace channel through sparse signal recovery. Specifically, we first construct the lens antenna array-based millimeter-wave MIMO adaptive selection network model and formulate the beamspace channel estimation as a sparse channel recovery problem. Second, based on the orthogonality of pilots, a Multi-d-AMP algorithm with a nonlinear scaling mechanism is designed to enable parallel channel estimation for multiple users. Finally, based on minimizing the estimation error, a nonlinear rescaling mechanism for hyperparameters in the logarithmic scale is introduced to prevent performance degradation caused by fixed hyperparameters. Simulation results show that the proposed scheme achieves desirable robustness and estimation accuracy.

1. Introduction

Future beyond 5G or 6G networks are expected to support ultra-high data rates, massive connectivity, and other requirements [1]. Millimeter-wave Multiple-input Multiple-output (MIMO), with its abundant available bandwidth, has emerged as an effective technology [2,3]. In particular, integrated millimeter-wave MIMO with Filter Bank Multi-Carrier (FBMC) modulation enables higher data rates and better spectral properties [4].

In millimeter-wave MIMO-FBMC, achieving the desired array gain requires accurate channel estimation. The reason is that beamforming relies on Channel State Information (CSI) to accurately direct the beam towards the target user [3]. Although Wang et al. [4] provides a feasible channel estimator, the proposed scheme requires many pilots to obtain sufficient CSI observations, leading to excessive pilot overhead. In contrast, beamspace channel estimation is more interesting. The reason is that the beamspace channel exhibits sparsity, and the CSI can be sparsely reconstructed using a small number of pilots. Lee et al. [5] proposed an Orthogonal Matching Pursuit (OMP)-based millimeter-wave MIMO open-loop channel estimation method, which reduces the coherence of the redundant dictionary and improves angle resolution through non-uniform quantized angle grids. Gao et al. [6] proposed a beamspace channel estimation scheme based on Support Detection (SD), which exhibits high reliability and low pilot overhead. However, the scheme only considers the case of single-antenna users. Thereafter, beamspace channel estimation based on successive support detection (SSD) was proposed in Gao et al. [7], which compensates for the shortcomings of the SD scheme. That is, both single-antenna and multi-antenna users are considered. Wei et al. [8] proposed a beamspace channel estimation method based on Gaussian Mixture Learning Approximate Message Passing (GM-LAMP). The proposed scheme assumes that the beamspace channel follows a Gaussian Mixture (GM) distribution and derives a GM-based shrinkage function, thereby the GM-LAMP network is constructed to achieve more accurate beamspace channel estimation. In addition, a low-pilot-overhead channel estimation scheme based on hybrid-field OMP has been proposed [9]. The proposed scheme improves the estimation accuracy by constructing a hybrid-field channel model to capture the hybrid-field characteristics. Zheng et al. [10] proposed a deep learning-assisted FBMC-MTC system that supports diversified machine-type data transmission while effectively mitigating the intrinsic imaginary interference of FBMC. Wang et al. [11] proposed a channel estimation network based on global self-attention weighting (TR-crcCAMnet). The proposed deep learning network extracts feature vectors of the channel response through parallel convolutional branches and embeds

* Corresponding author.
E-mail addresses: wangyingstu@163.com (Y. Wang), guoqiang@hrbeu.edu.cn (Q. Guo), xiangjianhong@hrbeu.edu.cn (J. Xiang), jade.zhong@hotmail.com (Y. Zhong).

https://doi.org/10.1016/j.dsp.2026.106189

Available online 27 April 2026

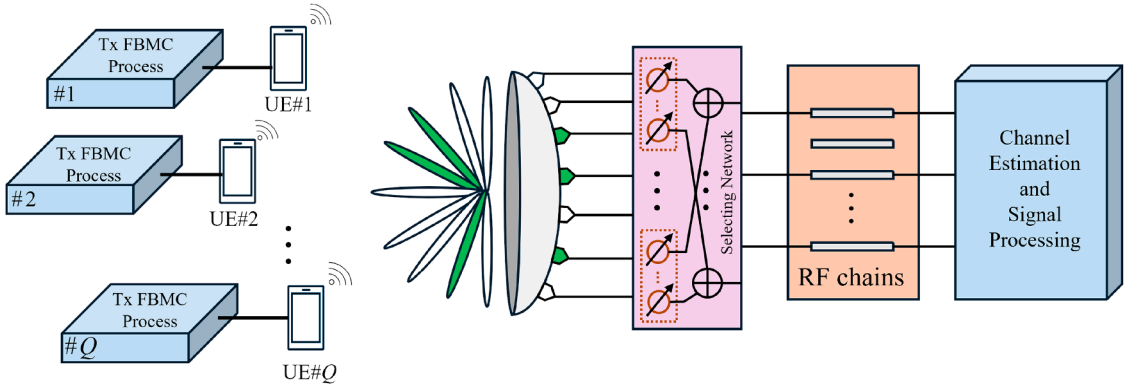

Fig. 1. Uplink millimeter-wave MIMO-FBMC system architecture based on lens antenna array.

a channel-attention module into the residual blocks to suppress redundant channel information, thereby improving the estimation accuracy. Cheng et al. [12] proposed a deep-learning-based channel estimation and equalization method for FBMC-OQAM. In this scheme, a deep neural network learns both the CSI and the constellation demapping strategy and then performs implicit equalization on the distorted frequency-domain sequence to directly recover the binary bits. However, the absence of explicit CSI output in the network deprives downlink users of the prior information required for beamforming. For FBMC systems, Ren et al. [13] proposed a joint channel estimation and preamble-based channel estimation scheme. Chen et al. [14] proposed an interference-utilization channel estimation method together with pilot optimization. Liu et al. [15] proposed a preamble-based channel estimation scheme for FBMC systems with delay diversity. However, these schemes are difficult to apply directly to MIMO-FBMC systems.

It is worth noting that the design of channel estimation methods for MIMO-FBMC systems must account for the intrinsic imaginary interference, which undermines the complex orthogonality relied upon in MIMO processing. However, by applying a linear transformation to the modulated QAM symbols, we can construct complex-valued symbols suitable for FBMC-OQAM transmission. This preprocessing operation enables FBMC-OQAM to be equivalent to a complex-orthogonal system in the baseband model, thereby allowing the direct application of sparse recovery algorithms. Inspired by the assumption of prior distributions, we assume that the beamspace channel follows a Laplacian distribution, which is commonly employed in Bayesian compressive sensing [16]. By introducing nonlinear scaling with hyperparameters and constructing a Multi-dimensional AMP (Multi-d-AMP), synchronous channel estimation is achieved for multi-user (or single-user equipped with multiple antennas) scenarios. Our contribution can be summarized below:

- For the problem of multi-user parallel channel estimation, we propose a beamspace channel estimation method based on Multi-d-AMP. The core of the proposed method lies in the construction of a multidimensional soft-thresholding decision function. Specifically, the residual with the Onsager correction term is first calculated to characterize the implicit effective noise present during the estimation process. Second, based on this effective noise, a pseudo-observation model is constructed for estimating the beamspace channel. Finally, by independently solving the maximum a posteriori estimation closed-form solution of the pseudo-observation model for each user, a multi-dimensional soft-thresholding decision function is jointly constructed.
- To address the limitation of fixed hyperparameters that cannot balance sparsity and noise, this paper introduces a logarithmic scale-based hyperparameter nonlinear scaling mechanism, enabling noise-aware hyperparameter adjustment. Specifically, the maximum a posteriori estimation of the beamspace channel is derived based on the Laplace distribution, and the analytical relationship between the scale parameter and the soft threshold is established. Second, a nonlinear scaling mechanism in the logarithmic scale is introduced for the hyperparameter to suppress extreme thresholds and ensure scaling stability. Finally, an empirical balancing formulation is derived to minimize channel estimation error, offering a mathematical basis for scenario-specific hyperparameter design.
- To address the problems that the effective time interval of channel reciprocity is not yet clearly defined and the empirical dependence of beamspace channel sparsity, we propose a theoretical basis for calculating the effective reciprocity interval (Lemma 1) and a mathematical criterion for determining beamspace channel sparsity (Lemma 2), respectively. These results offer analytical foundations for quantifying the temporal validity of channel reciprocity and reliably identifying sparse structures in beamspace channel representations.
- We adopt the Saleh-Valenzuela channel model, which is widely used in millimeter-wave communications, to validate the feasibility of the proposed scheme. The simulation results demonstrate that the performance of Multi-d-AMP exhibits advantages across the entire signal-to-noise ratio range.

Notation: Bold uppercase and lowercase letters denote matrices and vectors, respectively. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the complex conjugate, transpose, and conjugate transpose, respectively. $\mathcal{U}(-a,a)$ denotes a uniform distribution over the interval $(-a,a)$. $\mathbf{1}_M$ denotes an $M\times 1$ all-1 vector.

2. System model

In a millimeter-wave beamspace MIMO-FBMC system, the Base Station (BS) employs an M-element lens antenna array and N_{RF} (with $N_{RF} \ll M$) RF chains to serve Q single-antenna users (or a single user equipped with a Q-element antenna array) [4]. Based on the reciprocity of Time-Division Duplex (TDD) systems, the downlink channel can be directly obtained from the uplink channel [8]. Therefore, to efficiently acquire CSI, we consider an uplink TDD-based millimeter-wave MIMO-FBMC system. Fig. 1 shows the uplink millimeter-wave MIMO-FBMC system architecture based on lens antenna array. The overall system operation proceeds as follows: We consider Q single-antenna users, whose data symbols are processed by an FBMC modulator to generate the transmit signal. After propagation through the wireless channel, the FBMC signals arrive at the M-element lens antenna array at the BS. The core function of this array is to transform the incoming signals from the spatial domain into the beamspace, which focuses the physical paths onto a limited number of beam directions. The beam selection network, located after the lens antenna array, is controlled by N_{RF} RF chains and adaptively selects N_{RF} beams out of the M available beams for transmission. The N_{RF} selected signals are fed through the RF chains into the

baseband processor, where channel estimation is performed. Operating in TDD mode, this architecture exploits channel reciprocity, allowing the CSI estimated from the uplink to be used for downlink precoding and beamforming.

2.1. Benchmarking model

In a fixed frequency band, assuming that the FBMC signaling of a user contains L subcarriers and K time symbols, the transmitted signal $s_q(t)$ of the qth user can be expressed as Wang et al. [17]

$$s_q(t) = \sum_{l=1}^{L}\sum_{k=1}^{K} x_{l,k}\, p(t-KT) \times \exp\left(\mathrm{j}2\pi l F(t-KT)\right) \exp\left(\mathrm{j}\frac{\pi}{2}(l+k)\right), \tag{1}$$

where $p(t)$ denotes the prototype filter, for which we select PHYDYAS. T and F denote the symbol period and subcarrier spacing, respectively. Generally, $x_{l,k}$ denotes the real-valued Offset Quadrature Amplitude Modulation (OQAM) symbol. Due to the special structure of OQAM, FBMC inherently suffers from intrinsic imaginary interference, which leads to significant challenges for MIMO implementation [18]. However, by applying a precoding operation to the modulated QAM symbols, we can construct complex-valued symbols $x_{l,k}$ suitable for transmission. This scheme can effectively restore the complex orthogonality of FBMC, enabling the integration and application of MIMO techniques with FBMC to be as straightforward as in OFDM. Let all transmitted symbols be denoted by $\mathbf{x} = \left[x_{1,1}, \cdots, x_{L,K}\right]^T \in \mathbb{C}^{LK}$. Then, the precoding operation can be simply expressed as

$$\mathbf{x} = \mathbb{P}\left(\left[\tilde{x}_{1,1}, \cdots, \tilde{x}_{L/2\,,K}\right]^T\right), \tag{2}$$

where the function $\mathbb{P}(\cdot)$ accomplishes the precoding of $\tilde{x}_{l/2\,,k}$, and $\tilde{x}_{l/2\,,k}$ is the modulated QAM symbol. The time-domain precoding operation can be expressed as Wang et al. [18]

$$\begin{aligned} x_{l,k} &= \frac{1}{\sqrt{K}} \sum_{i=1}^{K/2} \tilde{x}_{l,i}\, Rwal(i,k) \\ st.\ Rwal(i,k) &= (-1)^{\sum_{j=1}^{\log_2(K/2)} \left\lfloor \frac{2i-1}{2^j} \right\rfloor \left\lfloor \frac{i}{2^j} \right\rfloor} \end{aligned}. \tag{3}$$

It is worth noting that during channel estimation, the pilot sequence should also undergo precoding. Using the frequency-domain precoding scheme presented in Wang et al. [19], FBMC can similarly support millimeter-wave MIMO transmission directly, as demonstrated in Wang et al. [4]. By stacking transmitted signals of all users into the vector $\mathbf{s}(t) = \left[s_1(t), \cdots, s_Q(t)\right]^T \in \mathbb{C}^{Q\times 1}$, the uplink millimeter-wave MIMO-FBMC model assisted by the lens antenna array can be expressed as Zeng and Zhang [20]

$$\mathbf{r}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t), \tag{4}$$

$$\mathbf{y}(t) = \mathbf{A}_t \mathbf{r}(t), \tag{5}$$

where $\mathbf{A}_t \in \mathbb{R}^{N_{RF}\times M}$ denotes the beam selecting network, $\mathbf{H} \in \mathbb{C}^{M\times Q}$ represents the beamspace channels of all users, and $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is the Gaussian white noise vector. Note that the precoding operation effectively eliminates the intrinsic imaginary interference of FBMC-OQAM, restoring the system's complex orthogonality and ensuring the validity of the equivalent complex baseband model in Eq. (4). In addition, if the selecting network is adaptive and implemented using low-cost 1-bit phase shifters as in Gao et al. [6,7], Wei et al. [8], then the elements of matrix $\mathbf{A}$ can be randomly selected from the set $\frac{1}{\sqrt{N_{RF}}}\{-1, 1\}$ with equal probability.

2.2. Beamspace channel

The use of lens antenna arrays enables a direct conversion of the spatial-domain channel into the beamspace channel [21]. Mathematically, the lens antenna array functions as a spatial discrete Fourier transform (DFT) matrix $\mathbf{U} \in \mathbb{C}^{M\times M}$. To formulate the beamspace channel, we begin our analysis from the spatial-domain channel. Considering the widely used clustered Saleh-Valenzuela (SV) channel model [22], the spatial channel vector $\bar{\mathbf{h}}_q \in \mathbb{C}^{M\times 1}$ of the qth user can be expressed as

$$\bar{\mathbf{h}}_q = \sqrt{\frac{M}{N_{\mathrm{p}}}} \sum_{n_{\mathrm{p}}=1}^{N_{\mathrm{p}}} c_{q,n_{\mathrm{p}}} \mathbf{a}\left(\phi_{q,n_{\mathrm{p}}}^{azi}, \phi_{q,n_{\mathrm{p}}}^{ele}\right), \tag{6}$$

where N_{p} is the number of resolvable paths. $c_{q,n_{\mathrm{p}}}$ and $\left(\phi_{q,n_{\mathrm{p}}}^{azi}, \phi_{q,n_{\mathrm{p}}}^{ele}\right)$ is the complex gain and spatial angle of the n_{p}th path, respectively. $\mathbf{a}\left(\phi_{q,n_{\mathrm{p}}}^{azi}, \phi_{q,n_{\mathrm{p}}}^{ele}\right) \in \mathbb{C}^{M\times 1}$ is the array steering vector, which depends on the array geometry. To simplify the expression, we omit the subscripts. For a Uniform Linear Array (ULA), the steering vector is expressed as Gao et al. [6]

$$\mathbf{a}_{\mathrm{ULA}}(\phi) = \frac{1}{\sqrt{M}}\left[\exp\left(-\mathrm{j}2\pi\phi\mathbf{m}\right)\right], \tag{7}$$

where $\mathbf{m} = [0, \cdots, M-1]^T$. ϕ can be calculated as $d\sin(\theta)/\lambda$, where d is the antenna spacing, λ is the wavelength, and θ is the angle. For a Uniform Planar Array (UPA), the steering vector is expressed as Wei et al. [8]

$$\mathbf{a}_{\mathrm{UPA}}\left(\phi^{azi}, \phi^{ele}\right) = \frac{1}{\sqrt{M}}\left[\exp\left(-\mathrm{j}2\pi\phi^{azi}\mathbf{m}_1\right)\right] \otimes \left[\exp\left(-\mathrm{j}2\pi\phi^{ele}\mathbf{m}_2\right)\right], \tag{8}$$

where $\mathbf{m}_1 = \left[0, \cdots, M_1-1\right]^T$ and $\mathbf{m}_2 = \left[0, \cdots, M_2-1\right]^T$. Note that $M = M_1 \times M_2$, where M_1 and M_2 denote the numbers of horizontal and vertical antennas, respectively. ϕ^{azi} and ϕ^{ele} can be calculated as $d\sin(\theta^{azi})\sin(\theta^{ele})/\lambda$ and $d\cos(\theta^{ele})/\lambda$, respectively. θ^{azi} and θ^{ele} denote the azimuth and elevation angles, respectively. In addition, for ULA and UPA, $\mathbf{U}$ can be calculated as Wang et al. [4]

$$\mathbf{U}_{\mathrm{ULA}} = \left[\mathbf{a}\left(\hat{\phi}_1\right), \cdots, \mathbf{a}\left(\hat{\phi}_M\right)\right]^H, \tag{9}$$

$$\mathbf{U}_{\mathrm{UPA}} = \left[\mathbf{a}\left(\hat{\phi}_1^{azi}, \hat{\phi}_1^{ele}\right), \cdots, \mathbf{a}\left(\hat{\phi}_{M_1}^{azi}, \hat{\phi}_{M_2}^{ele}\right)\right]^H, \tag{10}$$

where $\hat{\phi}_m = \frac{1}{M}\left(m - \frac{M+1}{2}\right)$. Thus, the beamspace channel $\mathbf{H} \in \mathbb{C}^{M\times Q}$ for all users can be expressed as

$$\mathbf{H} = \left[\mathbf{U}\bar{\mathbf{h}}_1, \mathbf{U}\bar{\mathbf{h}}_2, \cdots, \mathbf{U}\bar{\mathbf{h}}_Q\right]. \tag{11}$$

Note that millimeter-wave high-frequency transmission primarily relies on line-of-sight paths and a limited number of scattered paths, resulting in only a few significant components in the beam domain. Thus, the beamspace channel is approximately sparse [23].

3. Beamspace channel estimation

The sparsity of the beamspace channel enables the application of compressed sensing techniques for channel estimation, effectively exploiting the inherent structural properties of the system while significantly reducing pilot overhead. It is worth noting that, in TDD systems, to fully exploit channel reciprocity, the pilot signaling and channel estimation must be completed within the valid reciprocity interval; otherwise, the downlink channel cannot be accurately inferred from the uplink estimation. Neglecting the effect of noise perturbations, the effective time of channel reciprocity can be determined by Lemma 1.

Lemma 1. *In lens antenna array-assisted mmWave MIMO-FBMC systems operating under TDD mode, assuming the channel remains approximately constant over a duration $\mathcal{T}$, the uplink and downlink channels can be regarded as reciprocal within this time interval. The maximum duration $\mathcal{T}$ over which this reciprocity holds is jointly determined by the channel coherence time within the beam angular range and the beam dwelling time of the lens array, given by*

$$\mathcal{T} = \min\left(\frac{0.382}{\nu_{\max}}, \mathcal{T}_{beam}\right), \tag{12}$$

where $\nu_{\max}$ denotes the maximum Doppler frequency. $\mathcal{T}_{beam}$ is the beam dwelling time, which depends on the lens aperture and beam switching mechanism.

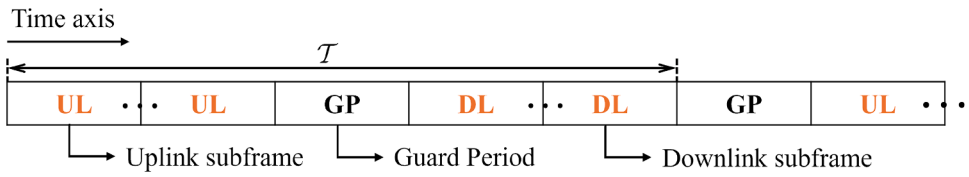

Fig. 2. Schematic diagram of TDD frame structure (simplified version).

Proof. The proof of Lemma 1 is provided in Appendix A. □

3.1. Problem formulation

The TDD frame structure is shown in Fig. 2. To obtain CSI, the uplink subframe lasts for $N_{\mathcal{P}}$ time instants during which pilots are transmitted to the BS. The guard period is employed to prevent cross-interference during uplink and downlink switching. After completion of channel estimation, downlink subframes are used to transmit data.

For now, we consider only the uplink. Considering the widely adopted orthogonal pilot transmission strategy [6], the uplink channel estimation for each user is performed independently. At the nth instant of pilot transmission, the received pilot vector $\mathbf{y}_{q,n} \in \mathbb{C}^{N_{RF}\times 1}$ of the qth user can be expressed as Gao et al. [6], Wei et al. [8]

$$\mathbf{y}_{q,n} = \mathbf{A}_n \mathbf{h}_q s_{q,n} + \mathbf{n}_{q,n}, \tag{13}$$

where $\mathbf{h}_q \in \mathbb{C}^{M\times 1}$ is the beamspace channel vector for the qth user. $s_{q,n}$ is the transmitted pilot signal. $\mathbf{n}_{q,n} = \mathbf{A}_{q,n}\bar{\mathbf{n}}_{q,n}$ is the effective noise vector. $\bar{\mathbf{n}}_{q,n} \sim \mathcal{CN}\left(0, P_{\mathrm{n}}\mathbf{I}_M\right)$ denotes the Gaussian noise vector with power P_{n}. After the transmission of $N_{\mathcal{P}}$ pilot instances, all the pilot vectors $\mathbf{y}_q = \left[\mathbf{y}_{q,1}^T, \cdots, \mathbf{y}_{q,N_{\mathcal{P}}}^T\right]^T \in \mathbb{C}^{N_{\mathcal{P}}N_{RF}\times 1}$ can be obtained, denoted as

$$\mathbf{y}_q = \mathbf{A}\mathbf{h}_q + \mathbf{n}_q. \tag{14}$$

In Eq. (14), without loss of generality, we assume $s_{q,n} = 1$ for $n = 1, \cdots, N_{\mathcal{P}}$, as in Gao et al. [7], Kim and Love [24]. In MIMO-FBMC, by designing comb-type pilot sequences, it can be ensured that the pilot signal satisfies $s_{q,n} = 1$ at the pilot transmission instant n. In this work, Zadoff-Chu sequences are employed to design constant-modulus transmit signals. In addition, the beam selection network $\mathbf{A} = \left[\mathbf{A}_1^T, \cdots, \mathbf{A}_{N_{\mathcal{P}}}^T\right]^T \in \mathbb{R}^{N_{\mathcal{P}}N_{RF}\times M}$ for all pilot instants is a matrix with elements $\pm\frac{1}{\sqrt{N_{\mathcal{P}}N_{RF}}}$. $\mathbf{n}_q = \left[\mathbf{n}_{q,1}^T, \cdots, \mathbf{n}_{q,N_{\mathcal{P}}}^T\right]^T \in \mathbb{C}^{N_{\mathcal{P}}N_{RF}\times 1}$ is the effective noise vector over $N_{\mathcal{P}}$ instances. It is worth noting that the subscript in Eq. (14) can be omitted due to the orthogonality of the pilots, which ensures that the channel estimation method is the same for all users. Moreover, for channel estimation, we adopt a random beam selection network implemented with low-cost 1-bit phase shifters. Specifically, at each of the $N_{\mathcal{P}}$ pilot instants, every element of the beam selection matrix $\mathbf{A}$ is randomly chosen with equal probability from the set $\frac{1}{\sqrt{N_{\mathcal{P}}N_{RF}}}\{+1, -1\}$. This design follows the random measurement principle of compressed sensing, aiming to construct a sensing matrix that satisfies the RIP, thereby ensuring that the sparse recovery algorithm can accurately estimate the channel with high probability. The random strategy avoids the mismatch between fixed or partially prior-based beam selection and the actual channel sparsity, enabling the channel energy in all possible beam directions to be captured with high probability. However, the random strategy is employed only during the channel estimation phase. Once the CSI is obtained, downlink data transmission adopts an interference-aware (IA) beam selection strategy based on the estimated values.

3.2. Sparse structure of the beamspace channel

Although Brady et al. [23] has demonstrated the sparsity of the beamspace channel, it remains essential to establish a rigorous mathematical foundation for this property, particularly in the context of millimeter-wave and high-frequency communications. We formalize the proof by Lemma 2.

Lemma 2. *Let* $\bar{\mathbf{h}} = \sqrt{\frac{M}{N_{\mathrm{p}}}}\sum_{n_{\mathrm{p}}=1}^{N_{\mathrm{p}}} c_{n_{\mathrm{p}}}\mathbf{a}(\phi)$, *where* $a(\phi)$ *denotes the array steering vector and* $N_{\mathrm{p}} \ll M$. *Then, on the DFT basis, the beamspace channel* $\mathbf{h} = \mathbf{U}\bar{\mathbf{h}}$ *is approximately sparse, with its significant nonzero entries predominantly concentrated on the* N_{p} *beam pairs.*

Proof. Note that the lens antenna array functions as a spatial DFT matrix $\mathbf{U}$. As a representative case, the array response vector $\mathbf{a}(\phi)$ for the ULA is denoted as

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{M}}\left[1, \exp\left(-\mathrm{j}2\pi\phi\right), \cdots, \exp\left(-\mathrm{j}2\pi\phi(M-1)\right)\right]^T, \tag{15}$$

where $\phi = d\sin(\theta)/\lambda$. Each column of the matrix $\mathbf{U}$ corresponds to the array response associated with a discrete direction $\hat{\phi}_m \in \Theta$, where $\Theta = \left[d\sin\left(\theta_1\right)/\lambda, \cdots, d\sin\left(\theta_M\right)/\lambda\right]$. Let

$$\tilde{\mathbf{a}}(\phi) = \mathbf{U}\mathbf{a}(\phi), \tag{16}$$

This transformation is equivalent to projecting the continuous angular response onto discrete directions. Mathematically, Eq. (16) has the following structure:

1) When $\phi \approx \hat{\phi}_m$, the corresponding beam exhibits its maximum main lobe.
2) For other values of $\hat{\phi}_m$, the magnitude decays rapidly.
3) For each path, only a few elements corresponding to angles close to the actual spatial angle exhibit maximum values.

This structure essentially corresponds to the property of the Dirichlet sinc function [20], i.e.,

$$|\tilde{\mathbf{a}}(\phi)| = \left|\left[\frac{\sin\left(M\pi\left(\phi-\hat{\phi}_1\right)\right)}{\sin\left(\pi\left(\phi-\hat{\phi}_1\right)\right)}, \cdots, \frac{\sin\left(M\pi\left(\phi-\hat{\phi}_M\right)\right)}{\sin\left(\pi\left(\phi-\hat{\phi}_M\right)\right)}\right]^T\right|. \tag{17}$$

Note that the above proof requires θ to be strictly aligned with the DFT angles. If θ is not exactly aligned with the DFT angles, a 'leakage effect' may occur [7]; however, the main energy remains concentrated. □

To characterize this sparsity, we desire to select a probability distribution to reflect the sparse structure of the beamspace channel. Lemma 2 demonstrates that most elements of the beamspace channel are close to zero, with only a few significant components, which is consistent with the characteristics of a Laplacian distribution. Therefore, we adopt the Laplace distribution to characterize the sparse prior, which is a well-established approach in Bayesian compressive sensing [16]. Fig. 3a and 3b show the beamspace channel amplitude and the corresponding Laplacian probability density function, respectively.

Assume that the elements of $\mathbf{h}$ are independent and identically distributed and follow a Laplace distribution, then

$$p(\mathbf{h}) = \prod_{m=1}^{M}\frac{\sigma}{2}\exp\left(-\sigma|h_m|\right) = \left(\frac{\sigma}{2}\right)^M\exp\left(-\sigma\|\mathbf{h}\|_1\right), \tag{18}$$

where σ is the scale parameter. Moreover, since the noise obeys a Gaussian distribution, thus

$$p(\mathbf{y}|\mathbf{h}) = \frac{1}{\left(2\pi P_{\mathrm{n}}\right)^{N_{\mathcal{P}}N_{RF}/2}}\exp\left(-\frac{1}{2P_{\mathrm{n}}}\|\mathbf{y}-\mathbf{A}\mathbf{h}\|_2^2\right). \tag{19}$$

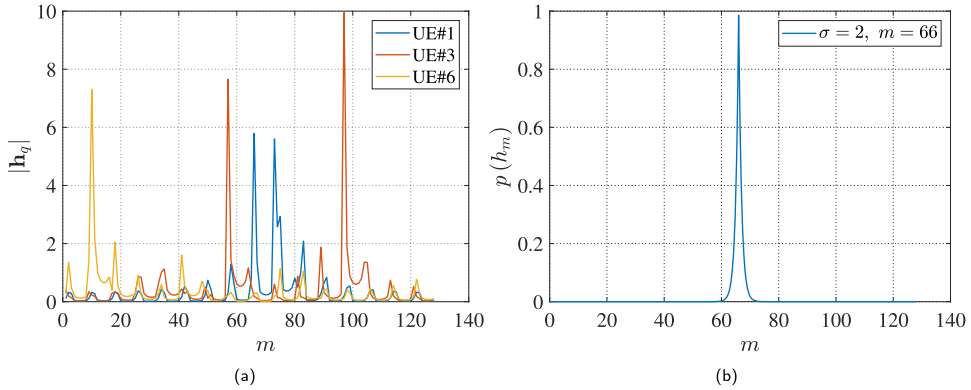

Fig. 3. (a) Beamspace channel amplitude for different users. (b) Laplace distribution with a scale parameter of 2.

According to Bayes' theorem, the maximum a posteriori estimate of $\mathbf{h}$ can be calculated as

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}) = \arg\min_{\mathbf{h}} p(\mathbf{y}|\mathbf{h})p(\mathbf{h}). \tag{20}$$

By applying the negative logarithmic equivalence transformation, we obtain

$$\begin{aligned}\hat{\mathbf{h}} &= \arg\max_{\mathbf{h}} p(\mathbf{y}|\mathbf{h})p((\mathbf{h})) \\ &= \arg\min_{\mathbf{h}} \left(-\log\left(p(\mathbf{y}|\mathbf{h})\right) - \log\left(p(\mathbf{h})\right)\right)\end{aligned}. \tag{21}$$

According to Eqs. (17) and (18), $-\log(p(\mathbf{h}))$ and $-\log(p(\mathbf{y}|\mathbf{h}))$ can be calculated as

$$-\log\left(p(\mathbf{h})\right) = -M\log\left(\frac{\sigma}{2}\right) + \sigma\|\mathbf{h}\|_1, \tag{22}$$

$$-\log\left(p(\mathbf{y}|\mathbf{h})\right) = \text{const} + \frac{1}{2P_{\text{n}}}\|\mathbf{y} - \mathbf{A}\mathbf{h}\|_2^2, \tag{23}$$

respectively. Neglecting constant terms, substituting Eqs. (22) and (23) into Eq. (21), we obtain

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \left(\|\mathbf{y} - \mathbf{A}\mathbf{h}\|_2^2 + 2P_{\text{n}}\sigma\|\mathbf{h}\|_1\right), \tag{24}$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ denote the Euclidean norm and the L1 norm, respectively. Eq. (23) can be formulated as a classical LASSO optimization, enabling the application of Compressed Sensing (CS) algorithms for accurate estimation. Selecting network $\mathbf{A}$ can be regarded as the sensing matrix in CS.

3.3. A nonlinear-scaled multi-d-AMP estimation scheme

When multiple users simultaneously request data transmission, resource contention among users may lead to queuing delays, which can degrade the overall transmission efficiency. Furthermore, as the number of users increases, the waiting time for some users may exceed the effective coherence time of channel reciprocity, resulting in channel mismatch and degraded quality of service. Therefore, the Approximate Message Passing (AMP) algorithm, which enables low-complexity signal reconstruction, emerges as a promising alternative [25]. However, the conventional AMP algorithm operating in a one-dimensional serial manner still suffers from waiting delays in multi-user scenarios, which limits its practical application efficiency. Due to the orthogonality of the pilots, a unified channel estimation scheme can be applied for all users. To enable the BS to estimate the channels of multiple users (or antennas) simultaneously, we propose the Multi-d-AMP algorithm. The Multi-d-AMP algorithm enables parallel channel estimation for multiple users, reducing latency without incurring additional computational complexity. The framework of the ith iteration of Multi-d-AMP is shown in Fig. 4.

It is worth noting that other compressed sensing-based estimators (e.g., OMP, SSD) typically perform channel estimation for each user independently in a sequential manner when handling multi-user scenarios. The sequential approach leads to a linearly increasing total computational delay as the number of users grows, which may cause the users' waiting time to exceed the effective reciprocity window of the channel. Although SSD accounts for multi-antenna users, its core estimation process remains sequential. In contrast, the proposed Multi-d-AMP algorithm achieves truly parallel estimation at the architectural level. By constructing a multidimensional soft-thresholding function, Multi-d-AMP processes the channels of all users in parallel, which not only reduces the waiting delay but also provides performance gains due to its parallel processing mechanism. Moreover, compared with deep learning-based schemes requiring offline training and large datasets (e.g., GM-LAMP and TR-crcCAMnet), the theoretically enhanced Multi-d-AMP offers better interpretability and environmental adaptability, without incurring any training overhead. Similar to that of AMP, the core update of the Multi-d-AMP framework is

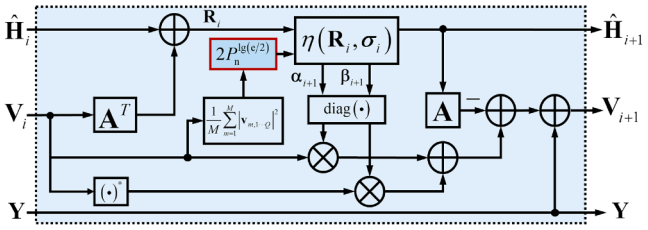

Fig. 4. The iteration framework of the ith step in Multi-d-AMP.

$$\hat{\mathbf{H}}_{i+1} = \boldsymbol{\eta}(\mathbf{R}_i, \boldsymbol{\sigma}_i), \tag{25}$$

where $\boldsymbol{\eta}(\cdot)$ denotes a multi-dimensional soft-thresholding function. $\mathbf{R}_i \in \mathbb{C}^{M\times Q}$ is the pseudo-observation matrix with residuals, and $\boldsymbol{\sigma}_i \in \mathbb{R}^{1\times Q}$ is the threshold vector. Without loss of generality, we consider the ith iteration and omit the subscript for simplicity. The maximum a posteriori estimate of $\mathbf{h}$ under the pseudo-observation model has a closed-form solution, which can be expressed as

$$\hat{h}_m = \begin{cases}\left(1 - \frac{\sigma_q}{|r_m|}\right)r_m, & \text{if } |r_m| > \sigma_q \\ 0, & \text{otherwise}\end{cases},\ m = 1, \cdots, M. \tag{26}$$

Since $\frac{\mathbf{r}}{|\mathbf{r}|} = \exp(\mathrm{j}\boldsymbol{\omega})$, Eq. (26) can therefore be rewritten as

$$\hat{\mathbf{h}} = \exp(\mathrm{j}\boldsymbol{\omega})\max\left(|\mathbf{r}| - \sigma_q\mathbf{1}_M, 0\right). \tag{27}$$

Thereby, $\boldsymbol{\eta}(\mathbf{R}, \boldsymbol{\sigma}) \in \mathbb{C}^{M\times Q}$ can be expressed as

$$\boldsymbol{\eta}(\mathbf{R}, \boldsymbol{\sigma}) = \begin{bmatrix}\left(\exp(\mathrm{j}\boldsymbol{\omega}_1)\max\left(|\mathbf{r}_1| - \sigma_1\mathbf{1}_M, 0\right)\right)^T \\ \vdots \\ \left(\exp(\mathrm{j}\boldsymbol{\omega}_Q)\max\left(|\mathbf{r}_Q| - \sigma_Q\mathbf{1}_M, 0\right)\right)^T\end{bmatrix}^T. \tag{28}$$

The pseudo-observation matrix $\mathbf{R} \in \mathbb{C}^{M\times Q}$ can be calculated as

$$\mathbf{R} = \hat{\mathbf{H}} + \mathbf{A}^T\mathbf{V}, \tag{29}$$

where $\mathbf{V} \in \mathbb{C}^{M\times Q}$ denotes the residual term that includes the Onsager correction. $\mathbf{V}_i$ at the ith iteration can be expressed as

$$\mathbf{V}_i = \mathbf{Y} - \mathbf{A}\hat{\mathbf{H}}_i + \mathbf{V}_{i-1}\mathrm{diag}(\boldsymbol{\alpha}_i) + \mathbf{V}^*_{i-1}\mathrm{diag}(\boldsymbol{\beta}_i), \tag{30}$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_Q] \in \mathbb{C}^{N_P N_{RF}\times Q}$. The Onsager correction coefficients $\boldsymbol{\alpha} \in \mathbb{R}^{1\times Q}$ and $\boldsymbol{\beta} \in \mathbb{R}^{1\times Q}$ can be calculated as

$$\begin{aligned}\boldsymbol{\alpha} &= \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\hat{h}_{m,1\cdots Q} \neq 0)\left[\frac{\partial\boldsymbol{\eta}(\mathbf{R},\boldsymbol{\sigma})}{\partial(\mathbf{R})}\right]_m \\ &= \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\hat{h}_{m,1\cdots Q} \neq 0)\left(1 - \frac{\boldsymbol{\sigma}}{2|\mathbf{r}_{m,1\cdots Q}|}\right),\end{aligned} \tag{31}$$

$$\begin{aligned}\boldsymbol{\beta} &= \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\hat{h}_{m,1\cdots Q} \neq 0)\left[\frac{\partial\boldsymbol{\eta}(\mathbf{R},\boldsymbol{\sigma})}{\partial\mathbf{R}^*}\right]_m \\ &= \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\hat{h}_{m,1\cdots Q} \neq 0)\left(1 - \frac{\boldsymbol{\sigma}\mathbf{r}^2_{m,1\cdots Q}}{2|\mathbf{r}_{m,1\cdots Q}|^3}\right),\end{aligned} \tag{32}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. According to Eqs. (31) and (32), the effective noise is not explicitly present but is reflected by the residual. Thus, the threshold vector $\boldsymbol{\sigma} \in \mathbb{R}^{1\times Q}$ can be calculated as

$$\boldsymbol{\sigma} = \varsigma\frac{1}{M}\sum_{m=1}^{M}|\mathbf{v}_{m,1\cdots Q}|^2, \tag{33}$$

where ς is a hyperparameter, which typically takes a fixed value of 1.1402 [26]. However, according to Eq. (24), we observe a correspondence between $2P_{\mathrm{n}}\sigma$ and Eq. (33), where P_{n} corresponds to $\frac{1}{M}\sum_{m=1}^{M}|\mathbf{v}_{m,1\cdots Q}|^2$. Therefore, the scale parameter σ should simultaneously match both the noise power and sparsity. The design of the threshold requires a balance between sparsity and noise suppression. From a Bayesian perspective, the threshold should be proportional to the equivalent noise power. However, a fixed proportionality coefficient cannot adapt to the dynamic range from low to high SNR. At low SNR, a larger threshold is needed to suppress noise, whereas at high SNR, a smaller threshold is required to recover weak signal components. Assuming that ς satisfies $f(P_{\mathrm{n}})$, then the following constraints should be met:

1) When P_{n} is relatively large, ς should also be correspondingly large;
2) As P_{n} approaches zero, ς should appropriately decrease.

To balance noise adaptability, suppress extreme thresholds, and ensure numerical stability, we introduce nonlinear scaling in the logarithmic domain (i.e., $\lg((\mathrm{e})/2)$) to achieve optimal matching. Thus, ς can be calculated as

$$\varsigma \triangleq f(P_{\mathrm{n}}) = 2P_{\mathrm{n}}^{\lg(\mathrm{e}/2)}, \tag{34}$$

where $\mathrm{e} = \exp(1)$. Note that Eq. (34) is an empirical balancing formula designed to minimize the channel estimation error. The hyperparameter balancing formula is an empirical function derived from extensive numerical experiments, aiming to achieve empirical NMSE minimization when channel sparsity is unknown. Although a closed-form theoretical proof is currently lacking, its superior performance validates its effectiveness in practice, as shown in Fig. 5. Multi-d-AMP no longer uses fixed hyperparameters; instead, it dynamically adjusts them based on P_{n}. The suboptimality of a fixed threshold in varying noise scenarios is avoided. Moreover, once the practical communication scenario is determined, Eq. (34) can serve as a mathematical basis for hyperparameter calculation, enabling estimation error minimization and enhancing channel estimation accuracy.

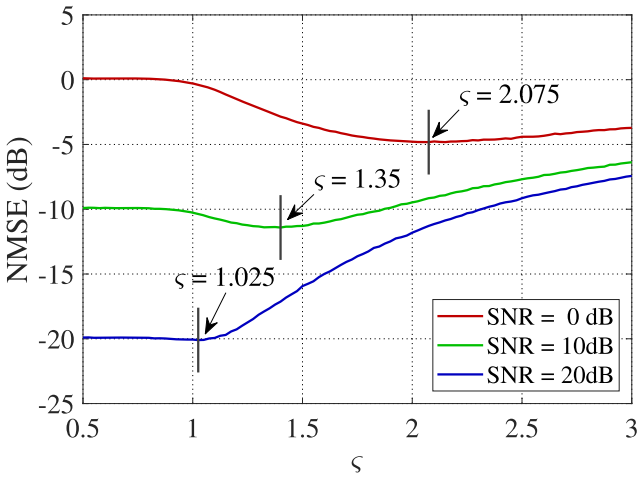

Fig. 5. Hyperparameter vs. NMSE of Channel Estimation.

Table 1
Simulation parameters.

Parameter name	Expression	Value
Carrier frequency	f_c	60GHz
Number of BS antennas	$M = M_1 \times M_2$	128
Number of users	Q	8
Number of RF chains	N_{RF}	16
Number of resolvable paths	N_p	3
Number of pilot instants	N_P	16
Antenna spacing	d	0.5

4. Numerical results

To validate the effectiveness of our work, we conduct a series of simulations based on the widely used SV channel model. Unless otherwise specified, the parameters are listed in Table 1. In addition, the Signal-to-Noise Ratio (SNR) for uplink channel estimation is defined as $P_s/P_{\mathrm{n}} = 1/P_{\mathrm{n}}$. For the SV channel model described in Eq. (5), the channel parameters of each user q follow the distribution [9]:

1) $c_{q,n_p} \sim \mathcal{CN}(0,1)$;
2) θ, θ^{ele} *and* $\theta^{azi} \sim \mathcal{U}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$.

The accuracy of channel estimation is evaluated using the Normalized Mean Square Error (NMSE), which is mathematically defined as

$$\bar{e}^2 = 10\lg\left(\left\|\hat{\mathbf{H}} - \mathbf{H}\right\|^2_{\mathrm{m2}}/\|\mathbf{H}\|^2_{\mathrm{m2}}\right), \tag{35}$$

where $\|\cdot\|_{\mathrm{m2}}$ denotes the m2-norm of a matrix.

4.1. NMSE performance analysis

Considering the theoretical foundation of CS in millimeter-wave MIMO systems, we provide a performance comparison of channel estimation using the OMP algorithm [5], the AMP algorithm, the SSD algorithm [7], and our proposed Multi-d-AMP algorithm. At the same time, we also compared with the deep learning-based GM-LAMP and TR-crcCAMnet.

Fig. 6a and 6b respectively show the NMSE performance of a 128×1 ULA and a 16×8 UPA under different SNR conditions. Based on the numerical results in Fig. 6, we can observe that Multi-d-AMP outperforms other methods across the entire SNR range. This advantage stems from the logarithmic scale nonlinear scaling mechanism of the algorithm's hyperparameters. The noise accumulation effect is suppressed through the dynamic adjustment of hyperparameters. The performance of Multi-d-AMP and traditional AMP overlaps at SNR = 20dB for ULA

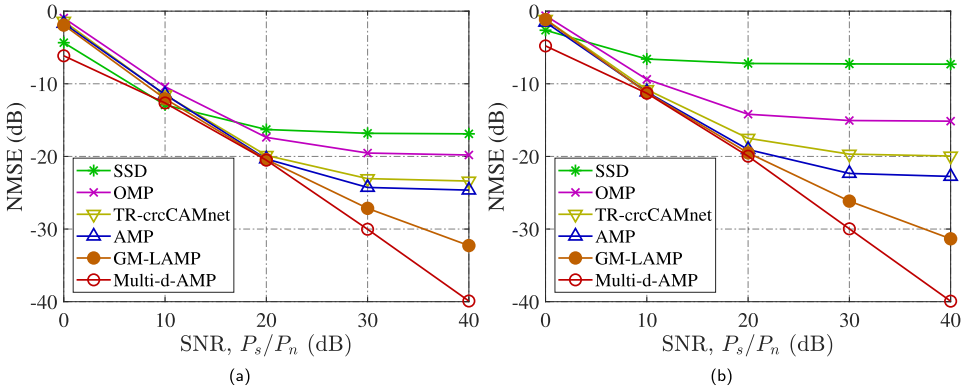

Fig. 6. (a) NMSE comparison of different algorithms in 128×1 ULA. (b) NMSE comparison of different algorithms in 16×8 UPA.

and SNR = 10dB for UPA. The reason is that the hyperparameters for nonlinear scaling adjustment (1.0826 and 1.4715, calculated according to Eq. (34)) are close to the fixed hyperparameter (1.1402). As the SNR increases, the convergence rate of Multi-d-AMP accelerates. At SNR = 25 dB, the NMSE of Multi-d-AMP reaches −25 dB, achieving the same reconstruction accuracy as AMP 5dB to 10dB earlier. This demonstrates the advantage of Multi-d-AMP in signal reconstruction efficiency. GM-LAMP performs worse than Multi-d-AMP in high SNR ranges, primarily because GM-LAMP assumes the spatial channel follows a fixed GM distribution. The GM prior is typically learned offline from training data and cannot adapt to the real-time channel conditions at the current instant. In addition, the one-dimensional parameters of ULA exhibit translational invariance, while the two-dimensional parameters of UPA experience coupling effects. Thus, channel estimation for UPA is more challenging than that for ULA. However, Multi-d-AMP shows almost the same performance in ULA and UPA, addressing the challenge of parameter coupling.

It is worth noting that in high-quality signal scenarios with SNR > 20 dB, the NMSE curve of Multi-d-AMP exhibits a larger slope, indicating its superior ability to utilize strong signal characteristics for accurate estimation. In contrast, traditional AMP is constrained by fixed hyperparameters, leading to performance saturation at high SNR.

4.2. BER performance analysis

We consider an indoor communication scenario with a carrier frequency of 60 GHz and a mobile speed $V \leq 5$ km/h. According to Lemma 1, the effective time interval for channel reciprocity is $\mathcal{T} \geq 1.37$ ms. Furthermore, we assume that the beam dwell time is equal to the duration required for the complete data transmission. In the lens antenna array-assisted millimeter-wave MIMO-FBMC system, the duration of each frame is set to 1 ms, referring to the LTE standard. Due to the adoption of a TDD structure, each frame is divided into an uplink pilot transmission phase (0.2ms), a downlink data transmission phase (0.7ms), and an uplink-downlink switching guard period (0.1ms). The design of this frame structure can fully exploit the channel reciprocity of the TDD system within the effective time interval of channel reciprocity.

Figs. 7a and 7b respectively show the BER performance under different SNR conditions for channel estimation using the Multi-d-AMP algorithm with a 128×1 ULA and a 16×8 UPA. Note that we consider various receiving strategies, including Zero-Forcing (ZF) detection, Maximum Likelihood (ML) detection, and Alamouti's space-time block coding technique, as in Wang et al. [4]. This allows for a comprehensive evaluation of system performance under different detection schemes. Based on the numerical results shown in Fig. 7, we can observe that the proposed Multi-d-AMP algorithm exhibit superior performance when used for channel estimation and subsequent channel compensation. Particularly in the high SNR regime, the algorithm effectively suppresses channel distortion, enabling the system to achieve nearly error-free reliable transmission. Moreover, due to the presence of multipath propagation, Alamouti's space-time block coding technique can provide diversity gain. Accordingly, depending on the communication scenario, the optimal receiving strategy should be selected to enhance reliability.

4.3. Achievable sum-rate analysis

To evaluate the impact of beamspace channel estimation on beam selection, we adopt IA beam selection scheme [27] and use the achievable sum rate as the performance metric.

Fig. 8a and 8b respectively show the achievable sum rate of the different estimated beamspace channels for the 128×1 ULA and 16×8 UPA. Based on the numerical results in Fig. 8, we can observe that Multi-d-AMP consistently outperforms across the entire SNR range. At SNR = 30dB, the sum rate of Multi-d-AMP approaches 85bit/s/Hz. Compared with GM-LAMP (approximately 81bit/s/Hz), Multi-d-AMP improves by about 7%, and compared with AMP (approximately 75bit/s/Hz), improves by about 12%, narrowing the gap with Perfect CSI (approximately 90bit/s/Hz) to around 6%. At SNR = 10dB, the performance of all estimation methods is limited, and the performance of Multi-d-AMP is nearly identical to that of traditional AMP. However, its sum rate (approximately 32bit/s/Hz) is higher than that of OMP (approximately 29bit/s/Hz) and SSD (approximately 25bit/s/Hz), which indicates that high noise beamspace channel estimation significantly impacts beam selection. When SNR > 30dB, the sum rate slope of Multi-d-AMP closely approaches that of the perfect CSI case, indicating that its estimation error decays rapidly with increasing SNR. However, the traditional methods tend to be saturated, which shows the potential of Multi-d-AMP in high-quality signal scenarios.

4.4. Computational complexity analysis

The computational complexity of each iteration in Multi-d-AMP is primarily dominated by two matrix multiplications: $\mathbf{A}^T\mathbf{V}_i$ and $\mathbf{A}\hat{\mathbf{H}}_i$. Assuming the algorithm iterates for N_{IT} times, the computational complexity is

$$C_{\text{Multi-d-AMP}} = \mathcal{O}\big(2N_{IT}N_{\mathcal{P}}N_{RF}Q(M+1)\big). \tag{36}$$

The computational complexity of the traditional AMP algorithm is

$$C_{\text{AMP}} = \mathcal{O}\Big(2QN_{IT}N_{\mathcal{P}}N_{RF}\Big(M\Big(\tfrac{1}{N_{\mathcal{P}}N_{RF}}+1\Big)+1\Big)\Big). \tag{37}$$

For the SSD algorithm, assuming it iterates N_{IT} times, the complexity is

$$C_{\text{SSD}} = \mathcal{O}\big(QM\big(N_{IT}MN_{\mathcal{P}}N_{RF} + N_{\mathcal{P}}N_{RF}N_{IT}^2 + N_{IT}^3\big)\big). \tag{38}$$

The complexity of the OMP is higher due to the presence of matrix inversion in the algorithm. Assuming the sparsity is $S = N_{IT}$, the compu-

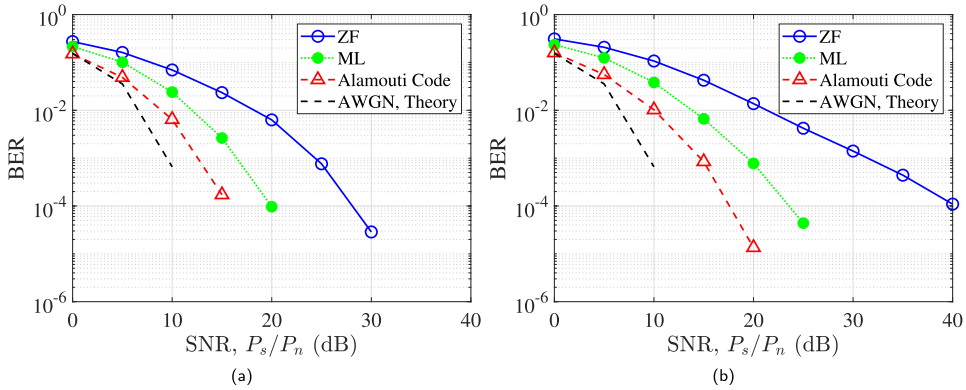

Fig. 7. 4QAM BER performance based on Multi-d-AMP channel estimation. (a) BER performance comparison of different detection algorithms in a 128×1 ULA. (b) BER performance comparison of different detection algorithms in a 16×8 UPA.

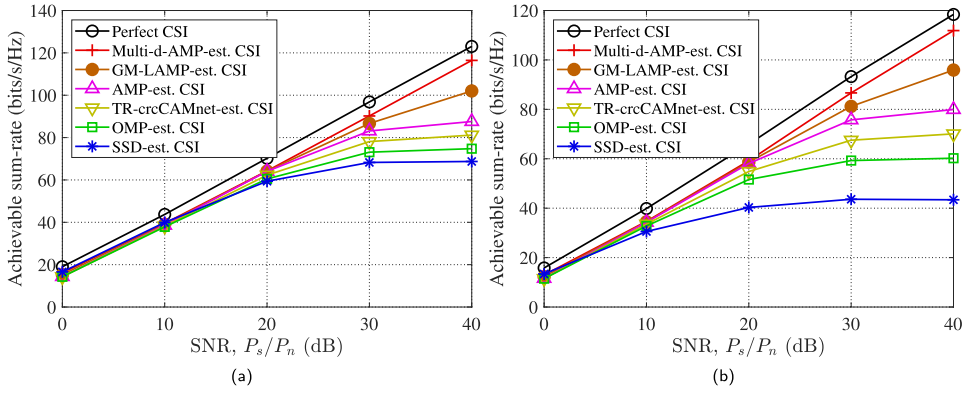

Fig. 8. (a) Sum-rate performance of IA beam selection for 128×1 ULA. (b) Sum-rate performance of IA beam selection for 16×8 UPA.

tational complexity of OMP is

$$C_{\mathrm{OMP}} = \mathcal{O}\Big(Q\big(N_P N_{RF}\big(N_{IT}M + N_{IT}^2 \quad + N_{IT}^3 + \tfrac{1}{N_P N_{RF}} N_{IT}^4\big)\big)\Big). \tag{39}$$

GM-LAMP has the same per-iteration complexity as AMP, but the introduction of the Gaussian mixture shrinkage function adds an additional $\mathcal{O}\big(N_{IT}QN_{RF}\mathcal{G}\big)$ computational overhead, where $\mathcal{G}$ is the number of components in the Gaussian mixture model. In contrast, the Multi-d-AMP algorithm avoids the additional overhead associated with model learning. The computational complexity of TR-crcCAMnet is

$$C_{TR-crcCAMnet} = \mathcal{O}\big(C^2\big(N_P M + L_c Q^2 \mathcal{C}\big)\big), \tag{40}$$

where $\mathcal{C}$ denotes the number of network channels. During online inference, the network involves forward propagation through multiple convolutional layers and attention mechanisms, and the large number of network channels results in a high absolute computational load. In contrast, Multi-d-AMP offers lower computational complexity and better interpretability.

Compared to traditional AMP, Multi-d-AMP reduces the complexity of $\mathcal{O}\big(2QN_{IT}M\big)$. The advantage is the absence of higher-order terms, making it suitable for large-scale MIMO or high-dimensional signal processing. OMP and SSD, due to the presence of higher-order terms, are more suitable for handling small-scale problems. Thus, Multi-d-AMP provides a more efficient solution for large-scale MIMO and high-dimensional signal recovery.

5. Conclusion

For the millimeter-wave MIMO-FBMC channel estimation problem, this letter proposes a beamspace channel estimation method based on Multi-d-AMP. The proposed method improves estimation efficiency by constructing a multi-dimensional soft-thresholding decision function, enabling parallel estimation of multi-dimensional channels. In addition, a logarithmic-scale-based nonlinear scaling mechanism for hyperparameters is introduced to enhance noise adaptability, suppress extreme threshold effects, and ensure stable parameter scaling. Specifically, the proposed algorithm exhibits lower computational complexity, making it more suitable for large-scale MIMO and high-dimensional signal processing scenarios. To validate the practicality of the proposed scheme, we analyzed the NMSE and sum-rate performance of ULA and UPA over the SV channel. The experimental data provided show the feasibility and effectiveness of the scheme. Although the proposed scheme has demonstrated its superiority through simulations, a key limitation is the lack of experimental validation on practical platforms. We plan to focus on practical platform testing of Multi-d-AMP in our future work. Specifically, we plan to deploy Multi-d-AMP on an FPGA-based platform to evaluate its performance in real wireless environments, addressing practical challenges such as hardware impairments, phase noise, and more complex channel non-stationarities. Moreover, developing a rigorous theoretical derivation of the hyperparameter balancing formula is also an important direction.

CRediT authorship contribution statement

Ying Wang: Conceptualization, Methodology, Software, Validation, Writing – original draft, Supervision; **Qiang Guo:** Writing – review & editing; **Jianhong Xiang:** Data curation, Project administration; **Yu Zhong:** Funding acquisition.

Data availability

Available Code can be downloaded at https://github.com/WangYeeng/CEmmWaveMIMOsystems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by Fundamental Research Funds for the Central Universities, China [3072022CF0801], Collaborative Research on Key Technologies for Maritime Search and Positioning Based on the BeiDou Navigation System, China [2018YFE0206500], and Sichuan Provincial Key Laboratory of Agile Intelligent Computing Project.

Appendix A. Proof of Lemma 1

According to Eq. (5), we assume that the channel is time-varying over time within the beam angular range, where each beam contains a cluster of paths that satisfy coherent superposition. Moreover, although scattering is limited in the millimeter-wave band, the beamspace channel still comprises multiple propagation paths [8,23]. Thus, the beamspace channel can be approximately regarded as a "reduced-order Jakes" model. The time correlation function $R(\mathcal{T})$ of the channel can be calculated as

$$R(\mathcal{T}) = \int_{-\nu_{\max}}^{\nu_{\max}} S(\nu_i) \exp\left(\mathrm{j}2\pi\nu_i\mathcal{T}\right)\mathrm{d}\nu_i, \tag{A.1}$$

where ν_i denotes the Doppler shift of the ith path. $S(\nu_i)$ denotes Jakes spectrum, defined as Hoeher [28]

$$S(\nu_i) = \begin{cases} \frac{1}{\pi\nu_{\max}\sqrt{1-(\nu_i/\nu_{\max})^2}}, & |\nu_i| \le \nu_{\max} \\ 0, & \text{otherwise} \end{cases}. \tag{A.2}$$

Thereby, $R(\mathcal{T})$ can be calculated as

$$R(\mathcal{T}) = \frac{2}{\pi\nu_{\max}} \int_0^{\nu_{\max}} \frac{\cos\left(2\pi\nu_i\mathcal{T}\right)}{\sqrt{1-(\nu_i/\nu_{\max})^2}}\mathrm{d}\nu_i. \tag{A.3}$$

Assuming that ν_i obeys a uniform angular distribution (i.e., $\nu_i = \nu_{\max}\cos(\theta_i)$ and θ_i obeys a uniform distribution), then

$$\begin{aligned} &R(\mathcal{T}) \\ &= \frac{2}{\pi\nu_{\max}} \int_{\frac{\pi}{2}}^{0} \frac{\cos\left(2\pi\nu_{\max}\mathcal{T}\cos(\theta_i)\right)}{\sqrt{1-\cos^2(\theta_i)}}\left(-\nu_{\max}\sin(\theta_i)\right)\mathrm{d}\theta_i \\ &= \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \cos\left(2\pi\nu_{\max}\mathcal{T}\cos(\theta_i)\right)\mathrm{d}\theta_i \\ &= J_0\left(2\pi\nu_{\max}\mathcal{T}\right) \end{aligned}. \tag{A.4}$$

We consider the first-zero method, i.e., the time instant when the channel first completely loses coherence, which corresponds to $J_0(2\pi\nu_{\max}\mathcal{T}_c) = 0$. Thereby, $2\pi\nu_{\max}\mathcal{T}_c \approx 2.4048$. Thus, the channel coherence time within the beam angular range can be calculated as

$$\mathcal{T}_c = \frac{2.4048}{2\pi\nu_{\max}} \approx \frac{0.382}{\nu_{\max}}. \tag{A.5}$$

Assuming the beam dwelling time is $\mathcal{T}_{beam}$, the channel reciprocity holds throughout the entire transmission process if $\mathcal{T}_c \ge \mathcal{T}_{beam}$. However, if $\mathcal{T}_c < \mathcal{T}_{beam}$, the channel reciprocity is valid only within the duration of $\mathcal{T}_c$. Therefore, the effective time interval of channel reciprocity is

$$\mathcal{T} = \min\left(\frac{0.382}{\nu_{\max}}, \mathcal{T}_{beam}\right). \tag{A.6}$$

References

[1] Y. Li, Y. Zhan, L. Zheng, X. Wang, Device activity detection and channel estimation for millimeter-wave massive MIMO, IEEE Trans. Commun. 72 (2) (2024) 1062–1074. https://doi.org/10.1109/TCOMM.2023.3325472
[2] K. Shi, X. Fang, X. Sha, Partial DFT codebook-based hybrid precoding in limited feedback millimeter wave massive MIMO systems, IEEE Wirel. Commun. Lett. 13 (7) (2024) 1788–1792. https://doi.org/10.1109/LWC.2024.3387926
[3] P. Wu, J. Cheng, Y.C. Eldar, J.M. Cioffi, Learned trimmed-ridge regression for channel estimation in millimeter-wave massive MIMO, IEEE Trans. Commun. 73 (2) (2025) 1128–1141. https://doi.org/10.1109/TCOMM.2024.3440875
[4] Y. Wang, Q. Guo, J. Xiang, Y. Zhong, Millimeter-wave MIMO transmission for FBMC systems with lens antenna arrays, IEEE Signal Process. Lett. 32 (2025) 1141–1145. https://doi.org/10.1109/LSP.2025.3547268
[5] J. Lee, G.-T. Gil, Y.H. Lee, Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications, IEEE Trans. Commun. 64 (6) (2016) 2370–2386. https://doi.org/10.1109/TCOMM.2016.2557791
[6] X. Gao, L. Dai, S. Han, C.-L. I, X. Wang, Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array, IEEE Trans. Wirel. Commun. 16 (9) (2017) 6010–6021. https://doi.org/10.1109/TWC.2017.2718502
[7] X. Gao, L. Dai, S. Zhou, A.M. Sayeed, L. Hanzo, Wideband beamspace channel estimation for millimeter-wave MIMO systems relying on lens antenna arrays, IEEE Trans. Signal Process. 67 (18) (2019) 4809–4824. https://doi.org/10.1109/TSP.2019.2931202
[8] X. Wei, C. Hu, L. Dai, Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems, IEEE Trans. Commun. 69 (1) (2021) 182–193. https://doi.org/10.1109/TCOMM.2020.3027027
[9] X. Wei, L. Dai, Channel estimation for extremely large-scale massive MIMO: far-field, near-field, or hybrid-field?, IEEE Commun. Lett. 26 (1) (2022) 177–181. https://doi.org/10.1109/LCOMM.2021.3124927
[10] X. Zheng, G. Liu, S. Li, T. Jiang, Deep learning-aided FBMC machine-type communication systems: design, simulation, and experimental test, IEEE Trans. Wirel. Commun. 23 (6) (2024) 5726–5739. https://doi.org/10.1109/TWC.2023.3328291
[11] L. Wang, J. Li, J. Xiang, W. Liu, TR-crcCAMnet: an attention-aided parallel weighted channel estimation based network, Phys. Commun. 72 (2025) 102746. https://doi.org/10.1016/j.phycom.2025.102746
[12] X. Cheng, D. Liu, C. Wang, S. Yan, Z. Zhu, Deep learning-based channel estimation and equalization scheme for FBMC/OQAM systems, IEEE Wirel. Commun. Lett. 8 (3) (2019) 881–884. https://doi.org/10.1109/LWC.2019.2898437
[13] D. Ren, J. Li, G. Lu, J. Ge, Joint channel estimation and equalization using new AFB output signal models for FBMC/OQAM systems, IEEE Trans. Commun. 69 (6) (2021) 4186–4201. https://doi.org/10.1109/TCOMM.2021.3064059
[14] D. Chen, R. Wang, T. Jiang, Channel estimation and pilot symbol optimization based on intrinsic interference utilization for OQAM/FBMC systems, IEEE Trans. Signal Process. 69 (2021) 4595–4606. https://doi.org/10.1109/TSP.2021.3101829
[15] W. Liu, S. Schwarz, M. Rupp, D. Chen, T. Jiang, Preamble-based channel estimation for OQAM/FBMC systems with delay diversity, IEEE Trans. Wirel. Commun. 19 (11) (2020) 7169–7180. https://doi.org/10.1109/TWC.2020.3008736
[16] S.D. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (1) (2010) 53–63. https://doi.org/10.1109/TIP.2009.2032894
[17] Y. Wang, Q. Guo, J. Xiang, Y. Zhong, Doubly selective channel estimation for FBMC and OFDM systems based on MIR correlation, IEEE Wirel. Commun. Lett. 14 (3) (2025) 851–855. https://doi.org/10.1109/LWC.2025.3525592
[18] Y. Wang, Q. Guo, J. Xiang, L. Wang, Y. Liu, Bi-orthogonality recovery and MIMO transmission for FBMC systems based on non-sinusoidal orthogonal transformation, Signal Process. 219 (2024) 109427. https://doi.org/10.1016/j.sigpro.2024.109427
[19] Y. Wang, Q. Guo, J. Xiang, Y. Zhong, Pruned DCT precoding-based FBMC modulation: an SC-FDMA inspired approach, Signal Process. 231 (2025) 109897. https://doi.org/10.1016/j.sigpro.2025.109897
[20] Y. Zeng, R. Zhang, Millimeter wave MIMO with lens antenna array: a new path division multiplexing paradigm, IEEE Trans. Commun. 64 (4) (2016) 1557–1571. https://doi.org/10.1109/TCOMM.2016.2533490
[21] X. Gao, L. Dai, A.M. Sayeed, Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications, IEEE Commun. Mag. 56 (4) (2018) 211–217. https://doi.org/10.1109/MCOM.2018.1600727
[22] A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath, Channel estimation and hybrid precoding for millimeter wave cellular systems, IEEE J. Sel. Top. Signal Process. 8 (5) (2014) 831–846. https://doi.org/10.1109/JSTSP.2014.2334278
[23] J. Brady, N. Behdad, A.M. Sayeed, Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements, IEEE Trans. Antennas Propag. 61 (7) (2013) 3814–3827. https://doi.org/10.1109/TAP.2013.2254442
[24] T. Kim, D.J. Love, Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels, in: 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2015, pp. 146–150. https://doi.org/10.1109/SPAWC.2015.7227017
[25] D.L. Donoho, A. Maleki, 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. https://doi.org/10.1109/ITWKSPS.2010.5503193

[26] M. Borgerding, P. Schniter, S. Rangan, AMP-inspired deep networks for sparse linear inverse problems, IEEE Trans. Signal Process. 65 (16) (2017) 4293–4308. https://doi.org/10.1109/TSP.2017.2708040
[27] T.Y. Elganimi, R.I. Elmajdub, G. Nauryzbayev, K.M. Rabie, IRS-assisted beamspace millimeter-wave massive MIMO with interference-aware beam selection, in: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), 2022, pp. 1–6. https://doi.org/10.1109/VTC2022-Fall57202.2022.10012839
[28] P. Hoeher, A statistical discrete-time model for the WSSUS multipath channel, IEEE Trans. Veh. Technol. 41 (4) (1992) 461–468. https://doi.org/10.1109/25.182598

(**Ying Wang**) (Graduate Student Member, IEEE) received the B.S. degree from Henan University of Engineering, Henan, China, in 2021, majoring in Communication Engineering. He passed the Graduate Exam for Harbin Engineering University (HUE) , Harbin, China in 2021 , and has been pursuing HEU's Ph.D. degree since 2022 , majoring in Information and Communication Engineering. He has authored/co-authored 10 technical papers in major international journals, and has received 2 Chinese invention patents. His main research interests include 4G/5G communications, massive MIMO, and multicarrier modulation, waveform design, and signal processing in wireless communications.

(**Qiang Guo**) received the M.S. and Ph.D. degrees from Harbin Engineering University (HEU), Harbin, China, in 2003 and 2007, respectively. He is an expert in accreditation for Chinese Education Ministry and Postgraduate Education, and a Fellow of the Ukrainian Engineering Academy. Currently, he is a Professor with Harbin Engineering University, Harbin, China, working as the Leader of HEU's Ocean Renaissance Academic Team. He has authored/co-authored more than 30 technical papers in major international journals and 1 published scholarly book by Science Publisher. His main research directions include radar detection and countermeasures, and satellite communications.

(**Jianhong Xiang**) , Member, IEEE, received the M.S. and Ph.D. degrees from Harbin Engineering University, Harbin, China, in 2006 and 2009, respectively. He is the vice president for the Digital Economy Research Association of Heilongjiang Province. Currently, he is a Professor with Harbin Engineering University, working as the Executive Deputy Director for the National Experimental Teaching Demonstration Center for Electrical and Electronics. He has authored/co-authored more than 10 + technical papers and has received more than 10 Chinese invention patents. His primary research interests include communications countermeasures, artificial intelligence, and satellite communications.

(**Yu Zhong**) received the B.S. degree from Guilin University of Electronic Science and Technology in 2001, majoring in Applied Electronic Technology. From 2002, he has worked at the China Electronics Technology Group Corporation's Tenth Research Institute. Currently, he is a Senior Engineer at the China Electronics Technology Group Corporation's 10th Research Institute, working as a Field Master. His research interests include signal processing, wireless communications.