Beamspace Channel Estimation for the Millimeter-wave Massive MIMO Systems

Dual Loops-based Iteration Reduction Algorithms

Zhu, Lijun; Li, Zheng; An, Zeliang; Chu, Zheng; Zhu, Zhengyu; Chen, Gaojie; Li, Yonghui

Published in:
IEEE Transactions on Vehicular Technology

DOI (link to publication from Publisher):
10.1109/TVT.2025.3612767

Publication date:
2025

Document Version
Accepted author manuscript, peer reviewed version

Beamspace Channel Estimation for the Millimeter-wave Massive MIMO Systems: Dual Loops-based Iteration Reduction Algorithms

Lijun Zhu, Zheng Li, Zeliang An, *Member, IEEE*, Zheng Chu, *Member, IEEE*, Zhengyu Zhu, *Senior Member, IEEE*, Gaojie Chen, *Senior Member, IEEE*, and Yonghui Li, *Fellow, IEEE*

***Abstract*—In the millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) channel estimation problem blue employing lens antenna arrays, conventional compressed sensing algorithms demand numerous matrix-vector multiplications per iteration, thereby incurring substantial computational complexity. To address this challenge, we propose dual-loop beamspace channel estimation strategies that leverage the sparsity of the mmWave beamspace channel, formulating the estimation problem as a sparse signal recovery task. First, we design an effective dual-loop algorithm based on the ℓ_1 minimization problem to tackle the channel estimation problem. In the outer loop, an ℓ_1-based iterative reduction algorithm (ℓ_1-IRA) reduces the large-scale channel estimation problem to a series of small-scale subproblems by exploiting the sparsity of the beamspace channel. In the inner loop, the fast iterative shrinkage thresholding algorithm with backtracking (FISTAB) algorithm is used to solve these subproblems efficiently. Furthermore, conventional compressed sensing algorithms exhibit favorable performance in weakly correlated systems but suffer from significant performance degradation in strongly correlated scenarios. To mitigate this limitation, we design an ℓ_{1-2} minimization problem-based IRA (ℓ_{1-2}-IRA) for the beamspace channel estimation problem. Finally, simulation results show that the proposed dual loop methods significantly reduce pilot overhead and improve beamspace channel estimation accuracy compared to conventional channel estimation techniques.**

***Index Terms*—mmWave, dual-loop, beamspace channel estimation, iterative reduction algorithm, FISTAB**

I. Introduction

With the emergence of disruptive applications such as high-definition network video, autonomous driving, virtual reality [1], and the Internet of Things (IoT) [2], the demand for high-rate, low-latency, and ultra-dense networks is rapidly growing. The millimeter-wave (mmWave) band, ranging from 30 to 300 GHz, provides abundant spectral resources and bandwidths to fulfill the demand for high-efficiency, low-latency, and high-rate transmissions in 5G and 6G [3], [4]. The natural complementarity of mmWave and massive massive multiple-input multiple-output (MIMO) technologies makes the mmWave band ideal for massive MIMO deployment [5]–[7]. Specifically, the short wavelength of mmWave enables for the integration of multiple antenna units into a small physical space. Meanwhile, massive MIMO technology forms narrow beams by precisely controlling the phase and amplitude of many antenna units. However, with each antenna element employing a dedicated radio-frequency (RF) chain, the traditional multi-antenna configuration results in high hardware costs and increased energy consumption for mmWave massive MIMO systems [8], [9].

To reduce the number of RF chains required for the base station (BS), lens antenna arrays in mmWave massive MIMO systems have attracted considerable attention. Lens antenna arrays consist of electromagnetic lenses and antenna arrays, where the electromagnetic lenses have power focusing capability, and the elements of the antenna arrays are located on the focal plane of electromagnetic lenses [10]. The lens antenna arrays convert the spatial domain channel into beamspace domain channel by focusing signals from different directions (beams) on different antennas. By utilizing the sparse nature of the beamspace channel, a subset of beams with significant power can be selected, significantly reducing MIMO dimensions and the number of RF chains, thus addressing high power consumption and hardware costs in mmWave massive MIMO systems. However, the beam selection techniques depend on accurate channel state information (CSI) in the beamspace channel [11],

This work was supported in part by Joint optimization and upgrade R&D project for liquidation robots under Grant 2025XX-JS-620. This work was supported in part by the National Natural Science Foundation of China under Grant 62571495, in part by Natural Science Foundation of Henan Province under Grant 232300421097, in part by Program for Science & Technology Innovation Talents in Universities of Henan Province under Grant 23HASTIT019, in part by State Key Laboratory of Integrated Services Networks under Grant ISN25-24, Xidian University. The work of Zheng Chu was supported in part by the Ningbo Natural Science Foundation under Grant 2024J233. This work of Zhengyu Zhu was supported in part by the National Natural Science Foundation of China under Grant 62571495, 62571182. This work was supported by Fundamental Research Funds for the Central Universities, Sun Yat-sen University, under Grant No.24hytd010.(*Corresponding author: Zhengyu Zhu, Zheng Chu.*)

L. Zhu is with Seedlight robot (Shenzhen) Co., Ltd., Shenzheng 518000, China, and also with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China. (e-mail: zlj@seedlight.cn).

Z. Li and Z. Zhu are with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China. z. zhu is also is also with State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, 710071, China (email: stones_li@outlook.com, iezyzhu@zzu.edu.cn).

Z. An is with the School of Computer and Information, AnHui Polytechnic University, WuHu, AnHui, 241000, China, and also with Department of Electronic Systems, Aalborg University, Aalborg, DK-9220, Denmark. (email: will_amc@163.com).

Z. Chu is with the Department of Electrical and Electronic Engineering, University of Nottingham Ningbo China, Ningbo 315100, China, (email: andrew.chuzheng7@gmail.com).

G. Chen is with the School of Flexible Electronics (SoFE), Sun Yat-sen University, Sun Yat-sen University, Shenzhen, Guangdong 518107, China (e-mail: gaojie.chen@ieee.org).

Y. Li is with the School of Electrical and Information Engineering, University of Sydney, Sydney, NSW 2006, Australia (e-mail: yonghui.li@sydney.edu.au).

[12]. Channel estimation with lens antenna arrays is challenging due to high-dimensional channels and limited RF chains. Therefore, conventional channel estimation algorithms based on all-digital and hybrid analog/digital processing techniques are no longer applicable.

Leveraging the sparse nature of the mmWave beamspace channel, the channel estimation problem can be formulated as a sparse signal recovery problem, rendering compressed sensing algorithms effective solutions [13]–[15]. Recently, several schemes for beamspace channel estimation have been proposed, such as support detection (SD) [16]–[18], orthogonal matching pursuit (OMP) [19]–[22], approximate message passing (AMP) [23]–[25], and deep learning (DL) [26]–[32]. With priori knowledge of the number of channel components, Gao *et al.* [16] proposed a mmWave channel estimation scheme based on the SD algorithm, transforming the total beamspace channel estimation problem into a series of subproblems, where each sub-problem contained a sparse channel component. Long *et al.* [17] developed a block support detection (BSD) beamspace channel estimation algorithm with block sparsity. Sharifi *et al.* [18] extended the support set of non-zero components and proposed an improved SD algorithm, which selected two previous smaller support sets for nonline-of-sight (NLoS) paths and a new peaked support set for line-of-sight (LoS) paths in each iteration. For the mmWave massive MIMO system, the authors of [19]–[21] proposed channel estimation solvers based on the off-grid OMP algorithm, which could trace the parameters of multipath components to reconstruct sparse channel matrix. By exploiting the sparsity of the mmWave channel, Li *et al.* [22] grouped the subcarriers and estimated the support set by the OMP algorithm. However, the greedy type-based approaches involve a large number of matrix-vector multiplications at each iteration, leading to a rise in the computational complexity of the algorithms.

Cheng *et al.* [24] proposed a two-stage wideband channel estimation scheme for mmWave beamspace MIMO systems. The first stage employed multi-task sparse Bayesian learning (MT-SBL) for rough channel estimation, while the second stage used expectation-maximization (EM) algorithms for the channel parameter estimation. By employing the Kmeans Gaussian mixture model (GMM) algorithm to estimate the priori parameters of the beamspace channel, Ruan *et al.* [25] proposed a simplified learning-AMP (SL-AMP) network, reducing the training complexity and improving the channel estimation accuracy. By utilizing the priori information from the Gaussian mixture distribution model, Wei *et al.* [33] derived a new shrinkage function and proposed the iterative-based learning AMP (LAMP) algorithm for more accurate beamspace channel estimation. However, the AMP-type algorithms [24], [25], [33], [34] were dependent on the prior distribution and noise variance of the channel, which are often difficult to obtain in practice. As further research, the recent works [35]–[38] utilized tensor-based signal processing techniques with mmWave channel estimation, in which the multidimensional parameters (i.e., time delay, angle or Doppler shift) for each mmWave channel path could be accurately estimated.

Gao *et al.* [27] proposed a model driven-based DL approach [28] for mmWave channel estimation, unfolding the SBL algorithm into a deep neural network (DNN) to iteratively update the channel's variance. To reduce the computational complexity of traditional SBL algorithm, Zhang *et al.* [29] proposed a multilayer sparse Bayesian channel estimation algorithm, which provided a flexible approach to implement different layers of channel estimation by using either OMP or SBL, thus achieving an optimal balance between channel estimation accuracy and computational complexity. Subsequently, Gao *et al.* [30] proposed a DL method for estimating the mmWave massive MIMO channel by unfolding SBL, where the expectation step (E-step) used the AMP algorithm to reduce the complexity, and the maximization step (M-step) was implemented by a DL network to improve the channel estimation accuracy. Zheng *et al.* [31] presented a DL method based on a trainable proximal gradient descent network (TPGD-Net) to solve the beamspace channel estimation problem, which unfolded a hierarchical network with a proximal gradient descent (PGD) algorithm and set the gradient descent step size as a trainable parameter [28]. However, the DL methods suffer from limited interpretability, making their theoretical performance analysis challenging.

Guo *et al.* [39] presented an iterative reduction algorithm (IRA) for solving the least absolute shrinkage and selection operator (LASSO) problem, reducing the large-scale LASSO problem to small-scale subproblems. As the number of antennas at the BS increases, conventional compressed sensing algorithms, i.e., OMP, fast iterative shrinkage thresholding algorithm (FISTA), and alternate direction multiplier method (ADMM), require a large number of matrix vector multiplications for each iteration, resulting in higher computational complexity. Therefore, motivated by the IRA [39], we design the ℓ_1-IRA to solve the beamspace channel estimation problem, respectively, leveraging the sparse structure of the beamspace channel. Moreover, traditional compressed sensing algorithms are suitable for non-coherent conditions, and these algorithms have difficulty in effectively distinguishing channel components and providing the sparsest solution for channel vectors in highly coherent scenarios. To tackle this challenge, we propose a difference of convex algorithm (DCA)-based ℓ_{1-2}-IRA framework for the beamspace channel estimation problem. The main contributions of this paper are as follows:

- By leveraging the inherent sparsity of the mmWave beamspace channel, we formulate the channel estimation problem into a sparse signal recovery problem.
- First, we propose an efficient ℓ_1-based dual-loop algorithm for beamspace channel estimation. By leveraging beamspace channel sparsity, the outer loop employs an ℓ_1-IRA framework to decompose the large-scale channel estimation problem into a series of smaller subproblems. The inner loop then applies FISTA with backtracking (FISTAB) algorithm to efficiently solve these subproblems. Since the outer loop significantly reduces the dimensions of the channel estimation problem, our approach mitigates the computational burden of large-scale matrix-vector multiplications in the inner loop, thus reducing computational complexity.

Traditional compressed sensing algorithms perform optimally in low-coherence systems but exhibit performance degradation in highly coherent environments. To address this, we propose a DCA-based ℓ_{1-2} minimization method to solve beamspace channel estimation, enhancing the mmWave channel's sparsity. To reduce the complexity of the ℓ_{1-2} minimization algorithm, we develop a dual-loop estimation framework. The outer loop employs an ℓ_{1-2}-IRA framework by leveraging beamspace channel sparsity, while the inner loop uses a fast ADMM algorithm to efficiently solve the subproblems of the outer loop.

- Compared with the conventional channel estimation methods, the proposed ℓ_1 and ℓ_{1-2} minimization-based dual-loop algorithms achieve lower computational complexity of $\mathcal{O}\left(pQ\ln\left(1/\varepsilon\right)\right)$ and $\mathcal{O}\left(pN^2\right)$, respectively. Simulation results demonstrate that dual-loop beamspace methods achieve accurate channel estimation with reduced pilot overhead.

The remainder of this paper is organized as follows. In Section II, we describes system model and problem formulation. In Ssection III, we suggest a ℓ_1 minimization problem-based dual loops beamspace channel estimation approache. In Section IV, we propose a ℓ_{1-2} minimization problem-based dual loops beamspace channel estimation approache. In Section V, simulation experiments verify beamspace channel estimation accuracy and robustness. Finally, this paper is summarized in Section VI.

Notation: a, $\mathbf{a}$ and $\mathbf{A}$ denote scalars, vectors and matrices, respectively. The operations $\mathbf{a}^T$, $\mathbf{a}^H$ and $\mathbf{a}^{-1}$ denote the transpose, conjugate transpose, and inverse of the matrix $\mathbf{a}$, respectively. $|\mathbf{a}|$ and $\nabla f\left(\mathbf{a}\right)$ denote the absolute value and gradient information of the vector $\mathbf{a}$, respectively; $\left\|\mathbf{a}\right\|_1$ and $\left\|\mathbf{a}\right\|_2$ denote the 1-norm and 2-norm of the vector $\mathbf{a}$, respectively. $\langle\mathbf{a},\mathbf{b}\rangle$ are the inner products of the vectors $\mathbf{a}$ and $\mathbf{b}$. $\mathcal{U}\left(\mathbf{h}\right)$ denotes the optimization function with respect to the vector $\mathbf{a}$, and $\partial\mathcal{U}\left(\mathbf{h}\right)$ is the subgradient of the function with respect to the vector $\mathbf{a}$. The symbol $\mathrm{sign}\left(\cdot\right)$ is the sign function, ∞ denotes infinity, $\det\left(\mathbf{A}\right)$ is determinant of the matrix $\mathbf{A}$, and $\mathbb{E}\left(\cdot\right)$ is the mean value operation.

II. System Model and Problem formulation

We consider a mmWave massive MIMO system operating in the time-division duplex (TDD) mode, where the BS is equipped with N lens antennas and N_{RF} RF links to communicate with K single-antenna users simultaneously. Due to the hybrid precoding structure at the BS, a limited number of RF links are sufficient for the mmWave massive MIMO system, satisfying the condition $K \leqslant N_{RF} < N$ [16]. Fig. 1 illustrates the multi-user mmWave massive MIMO system with lens antenna array.

A. Traditional mmWave massive MIMO channel model

To estimate the beam-domain channel, we first introduce a spatial-domain mmWave massive MIMO channel model.

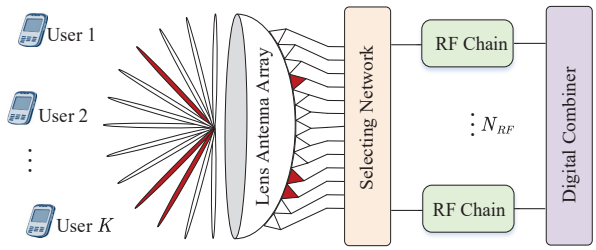


Fig. 1. The multi-user mmWave massive MIMO system based on lens antenna arrays.

Based on the Saleh-Valenzuela representation, the spatial-domain channel vector $\bar{\mathbf{h}}_k \in \mathbb{C}^{N\times 1}$ between the kth user and the BS is given by

$$\begin{aligned}\bar{\mathbf{h}}_k &= \sum_{i=0}^{L_k}\beta_{k,i}\mathbf{a}\left(\theta_{k,i}\right)\\ &= \beta_{k,0}\mathbf{a}\left(\theta_{k,0}\right)+\sum_{i=1}^{L_k}\beta_{k,i}\mathbf{a}\left(\theta_{k,i}\right)\\ &= \bar{\mathbf{c}}_{k,0}+\sum_{i=1}^{L_k}\bar{\mathbf{c}}_{k,i},\end{aligned} \tag{1}$$

where L_k denotes the number of paths, consisting of one LoS path and L_k NLoS paths. The LoS component is represented as $\bar{\mathbf{c}}_{k,0}=\beta_{k,0}\mathbf{a}\left(\theta_{k,i}\right)$, containing a singly entry, while the NLoS components are given by $\sum_{i=1}^{L_k}\bar{\mathbf{c}}_{k,i}=\sum_{i=1}^{L_k}\beta_{k,i}\mathbf{a}\left(\theta_{k,i}\right)$, comprising L_k entries [40]. Here, $\beta_{k,i}$ and $\theta_{k,i}$ represent the complex path gain and physical direction of the ith path for the kth user, respectively. For a typical uniform linear array (ULA), the array response vector $\mathbf{a}\left(\theta_{k,i}\right)\in\mathbb{C}^{N\times 1}$ can be determined from an angle [34]. Ignoring subscripts for simplicity, the normalized array response vector $\mathbf{a}\left(\theta\right)$ is denoted by

$$\mathbf{a}\left(\theta\right)=\frac{1}{\sqrt{N}}\left[e^{-j2\pi d\sin(\theta)m/\lambda}\right]_{m\in I(N)}, \tag{2}$$

where $I\left(N\right)=\left\{x-(N-1)/2, x=0,1,\cdots,N-1\right\}$ denotes an exponentially symmetric set with the center at zero. The antenna spacing between neighboring elements is $d=\lambda/2$, where λ is the carrier wavelength. Let $\varphi=d\sin\theta/\lambda$ be the spatial angle of the ULA, then (2) can be further transformed into

$$\mathbf{a}\left(\varphi\right)=\frac{1}{\sqrt{N}}\left[e^{-j2\pi\varphi m}\right]_{m\in I(N)}. \tag{3}$$

B. Lens antenna array-based mmWave wave massive MIMO system

Lens antenna arrays utilize the principle of optical lenses, which focus signals from the same direction to a focal point to form N orthogonal beams, effectively improving signal focusing and directionality [41]. The operation mechanism of the lens antenna arrays is similar to a discrete Fourier transform (DFT) matrix $\mathbf{U}\in\mathbb{C}^{N\times N}$, defined as follows

$$\mathbf{U}=\left[\mathbf{a}\left(\tilde{\varphi}_1\right),\mathbf{a}\left(\tilde{\varphi}_2\right),\cdots,\mathbf{a}\left(\tilde{\varphi}_m\right)\right]^H, \tag{4}$$

where $\bar{\varphi}_m = \frac{1}{N}\left(m - \frac{N+1}{2}\right) \in \left[-\frac{1}{2}, \frac{1}{2}\right]$,$m = 1, 2, \cdots, N$ denotes the predefined beam spatial direction of the discrete lens antenna, and for each $\mathbf{a}(\bar{\varphi})$ is denoted as

$$\mathbf{a}(\tilde{\varphi}) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\tilde{\varphi}m}\right]^T_{m\in I(N)}. \tag{5}$$

The discrete lens antenna array forms N orthogonal beams covering directions within $[-1/2, 1/2]$. Combined with (2)-(5), the spatial-domain channel vector $\bar{\mathbf{h}}_k$ is converted into beam-domain channel vector $\tilde{\mathbf{h}}_k$ through the lens antenna array, effectively mitigating multipath interference and signal attenuation during transmission. Thus, the beam-domain channel vector $\tilde{\mathbf{h}}_k$ is expressed as follows

$$\tilde{\mathbf{h}}_k = \mathbf{U}\bar{\mathbf{h}}_k = \sum_{i=0}^{L_k} \tilde{\mathbf{c}}_{k,i}, \tag{6}$$

where $\tilde{\mathbf{c}}_{k,i} = \mathbf{U}\bar{\mathbf{c}}_{k,i} = \mathbf{U}\beta_{k,i}\mathbf{a}(\theta_{k,i})$ is the multipath component of the ith path of the kth user, denoted by

$$\tilde{\mathbf{c}}_{k,i} = \left[\beta_{k,i}\Xi(\varphi_{k,i} - \bar{\varphi}_1), \cdots, \beta_{k,i}\Xi(\varphi_{k,i} - \bar{\varphi}_N)\right]^T, \tag{7}$$

where $\Xi(x) = \sin(\pi N x)/\sin(\pi x)$ is the Dirichlet sinc function, which has an energy focusing property causing most of the energy of $\tilde{\mathbf{c}}_{k,i}$ to be focused on a small number of antennas. In addition, due to the limited scattering in the mmWave channel, the signal energy is mainly concentrated on a small number of dominant paths. Therefore, the beam-domain channel $\tilde{\mathbf{h}}_k$ is sparse, with only a few nonzero values.

C. Problem Formulation

To obtain accurate CSI, all users transmit known pilot symbols to the BS within Q time slots, which are divided into M time slot blocks, each block containing K time slots, i.e., $Q = MK$. After beam selection, during the mth time slot block of pilot transmission, the received uplink signal matrix at the BS $\mathbf{Y}_m^U \in \mathbb{C}^{N\times K}$ is denoted as

$$\mathbf{Y}_m = \mathbf{U}\mathbf{H}\mathbf{X}_m + \mathbf{N}_m, \tag{8}$$

where $\mathbf{H} \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{X}_m \in \mathbb{C}^{K\times K}$ denotes K mutually orthogonal pilot matrices within the mth time slot block, satisfying $\mathbf{X}_m\mathbf{X}_m^H = \mathbf{I}_K$ and $\mathbf{X}_m^H\mathbf{X}_m = \mathbf{I}_K$. For simplify, we assume $\mathbf{X}_m = \mathbf{I}_K$. The noise matrix $\mathbf{N}_m \in \mathbb{C}^{N\times K}$ within the mth time slot block follows a Gaussian distribution $\mathcal{CN}\left(0, \sigma^2\right)$, where σ^2 represents the noise power.

At the mth time slot block, the BS employs a selecting network $\mathbf{W}_m \in \mathbb{C}^{K\times N}$ to merge the received signal matrix $\mathbf{Y}_m$, and right-multiple the pilot matrix $\mathbf{X}_m^H$, the measurement matrix $\mathbf{Y}_m \in \mathbb{C}^{K\times K}$ of the beamspace channel matrix $\tilde{\mathbf{H}}$ is represented by

$$\begin{aligned}\mathbf{Y}_m &= \mathbf{W}_m\mathbf{Y}_m\mathbf{X}_m^H \\ &= \mathbf{W}_m\mathbf{U}\mathbf{H}\mathbf{X}_m\mathbf{X}_m^H + \mathbf{W}_m\mathbf{N}_m\mathbf{X}_m^H \\ &= \mathbf{W}_m\tilde{\mathbf{H}} + \tilde{\mathbf{N}}_m,\end{aligned} \tag{9}$$

where $\tilde{\mathbf{H}} = \mathbf{U}\mathbf{H} = \left[\tilde{\mathbf{h}}_1, \cdots, \tilde{\mathbf{h}}_K\right] \in \mathbb{C}^{N\times K}$ is the beamspace channel matrix, $\tilde{\mathbf{N}}_m = \mathbf{W}_m\mathbf{N}_m\mathbf{X}_m^H \in \mathbb{C}^{K\times K}$ denotes effective noise matrix within the mth time slot block.

Without loss of generality, we estimate the beamspace channel vector $\tilde{\mathbf{h}}_k$ for the kth user, with a similar method applicable to other users to obtain the full beamspace channel matrix $\tilde{\mathbf{H}}$. Following the pilot transmission over M time slot blocks, the measurement vector $\tilde{\mathbf{y}}_k \in \mathbb{C}^{Q\times 1}$ of $\tilde{\mathbf{h}}_k$ for the kth user is given by

$$\tilde{\mathbf{y}}_k \triangleq \tilde{\mathbf{\Phi}}\tilde{\mathbf{h}}_k + \tilde{\mathbf{n}}_k, \tag{10}$$

where $\tilde{\mathbf{y}}_k = [y_{1,k}, \cdots, y_{M,k}]^T \in \mathbb{C}^{Q\times 1}$ denotes the kth column of $\mathbf{Y}_m$, $\tilde{\mathbf{h}}_k \in \mathbb{C}^{N\times 1}$ denotes the kth column of $\tilde{\mathbf{H}}$, and $\tilde{\mathbf{n}}_k = [n_{1,k}, \cdots, n_{M,k}]^T \in \mathbb{C}^{Q\times 1}$ denotes the kth column of $\tilde{\mathbf{N}}_m$. $\tilde{\mathbf{\Phi}} = [\mathbf{W}_1, \cdots, \mathbf{W}_M]^T \in \mathbb{C}^{Q\times N}$ is the total combination matrix containing M time slot blocks , with elements randomly selected from the set $\left(1/\sqrt{Q}\right)\{-1, 1\}$.

Due to the orthogonality of the pilot frequency, channel estimation for all K users is independent and identical. Therefore, the subscript k in problem (10) can be omitted and denoted as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{\Phi}}\tilde{\mathbf{h}} + \tilde{\mathbf{n}}. \tag{11}$$

Considering the real-valued counterpart $\mathbf{h} \in \mathbb{R}^{2N\times 1}$ of the beamspace channel vector $\tilde{\mathbf{h}}$, the linear measurement in the real domain is expressed as follows

$$\begin{aligned}\mathbf{y} &= \mathbf{\Phi}\mathbf{h} + \mathbf{n} \\ &= \begin{bmatrix} \Re\left(\tilde{\mathbf{\Phi}}\right) & -\Im\left(\tilde{\mathbf{\Phi}}\right) \\ \Im\left(\tilde{\mathbf{\Phi}}\right) & \Re\left(\tilde{\mathbf{\Phi}}\right) \end{bmatrix} \begin{bmatrix} \Re\left(\tilde{\mathbf{h}}\right) \\ \Im\left(\tilde{\mathbf{h}}\right) \end{bmatrix} + \begin{bmatrix} \Re(\tilde{\mathbf{n}}) \\ \Im(\tilde{\mathbf{n}}) \end{bmatrix},\end{aligned} \tag{12}$$

where $\mathbf{\Phi} \in \mathbb{R}^{2Q\times 2N}$, $\mathbf{y} \in \mathbb{R}^{2Q\times 1}$ denote the real-valued counterparts of $\tilde{\mathbf{\Phi}}$ and $\tilde{\mathbf{y}}$, respectively. It is worth noting that each element of the lens antenna arrays corresponds to a predefined beamspace direction, and the element of $\mathbf{h}$ has a maximum value when the actual angle formed by the antenna is close to the predefined beamspace angle of the lens antenna array. At the mmWave frequency, the mmWave beamspace channel is usually approximately sparse due to the limited number of scatterers, resulting in fewer transmission paths. Therefore, we can employ the sparse signal recovery algorithms to estimate the beamspace channel $\mathbf{h}$ with reduced pilot overhead. Specifically, by solving the ℓ_0 minimization problem to find the sparsest signal representation, the beamspace channel estimation problem is formulated as a sparse signal recovery problem, as follows

$$\begin{aligned}&\min \|\mathbf{h}\|_0, \\ &\text{s.t.} \|\mathbf{y} - \mathbf{\Phi}\mathbf{h}\|_2^2 \leqslant \varepsilon,\end{aligned} \tag{13}$$

where $\mathbf{\Phi}$ is the sensing matrix, $\|\mathbf{h}\|_0$ denotes the number of non-zero elements, and ε is the error tolerance parameter [42].

Due to the non-convexity of $\|\mathbf{h}\|_0$, the problem in (13) is NP-hard [33], making it computationally infeasible for high-dimensional channels. To obtain an approximate solution, a convex relaxation strategy is applied, transforming (13) into the following ℓ_1-minimization unconstrained form [43]

$$\mathbf{h}^* = \arg\min_{\mathbf{h}} \left\{\frac{1}{2}\|\mathbf{\Phi}\mathbf{h} - \mathbf{y}\|^2 + \lambda\|\mathbf{h}\|_1\right\}, \tag{14}$$

where λ is a tuning parameter that controls the sparsity of the solution. Traditional compressive sensing algorithms,

such as SOMP [8], SD [16], FISTA [44], and ADMM [45], have been applied to the sparse channel estimation problem. However, these algorithms often fail to achieve the required channel estimation accuracy. Let $f(\mathbf{h}) = 1/2 \left\|\mathbf{\Phi h} - \mathbf{y}\right\|_2^2$, it is worth noting that these algorithms, such as SD, FISTA and ADMM, require computing the gradient information $\triangledown f(\mathbf{h}) = \mathbf{\Phi}^T(\mathbf{\Phi h} - \mathbf{y})$, implying that $\mathbf{\Phi}^T\mathbf{z}$ and $\mathbf{\Phi h}$ need to be computed at each iteration, with $\mathbf{z} = \mathbf{\Phi h} - \mathbf{y}$ being some intermediate vector. As the number of antennas at the BS increases, these algorithms necessitate numerous matrix-vector multiplications per iteration, leading to higher computational complexity. Therefore, we propose low-complexity sparse beamspace channel estimation algorithms to enhance the channel estimation accuracy and reduce the pilot overhead.

III. ℓ_1 Minimization Problem-based Dual Loops Beamspace channel estimation approaches

In this section, we propose a ℓ_1 minimization problem-based dual-loop IRA algorithm to solve beamspace channel estimation problem, utilizing the clustering property of the non-sparse values of the channel. The outer loop adopts the ℓ_1-IRA framework to decompose the large-scale channel estimation problem into smaller subproblems. Meanwhile, the inner loop employs the FISTAB algorithm [44] to solve these subproblems.

A. The ℓ_1-IRA Framework

Since the beamspace channel is approximately sparse, with most elements of $\mathbf{h}$ being zero, it is unnecessary to optimize channel estimation for all elements. To simplify the channel estimation problem, we first define the estimated support set based on current Karush-Kuhn-Tucker (KKT) conditions

$$\begin{cases} W_1(\mathbf{h}^t) = \left\{j \left| \mathbf{h}_j^t \neq 0 \right.\right\}, \\ W_2(\mathbf{h}^t) = \left\{j \left| \mathbf{h}_j^t = 0 \text{ and } \right.\right. \\ \qquad\qquad \left.\left|\mathbf{\Phi}_j^T\left(\mathbf{\Phi h}^t - \mathbf{y}\right)\right| > (1-\tau)\lambda\right\}, \\ W(\mathbf{h}^t) = W_1(\mathbf{h}^t) \cup W_2(\mathbf{h}^t), \end{cases} \tag{15}$$

where τ is a small positive number, t denotes the tth external iteration index. $W_1(\mathbf{h}^t)$ denotes the index set of non-zero components of the tth iteration of $\mathbf{h}$, $W_2(\mathbf{h}^t)$ denotes the index set satisfying the zero component of the tth iteration of $\mathbf{h}$ and the current KKT condition, and $W(\mathbf{h}^t)$ denotes the estimated support set. For writing convenience, we abbreviate $W_1(\mathbf{h}^t)$, $W_2(\mathbf{h}^t)$, and $W(\mathbf{h}^t)$ as W_1^t, W_2^t, and W^t, respectively. In addition, the original channel estimation problem can be reduced to a series of small-scale subproblems over the index set of W^*, only if the support set W^* corresponding to the true solution is accurately obtained. However, in practice, it is difficult to obtain W^* completely, and the components corresponding to the approximate solution may contain only a portion of the indexes in W^*, which may be zero currently, and thus it is necessary to join them into the estimated support set W^t. To efficiently identify these index sets, it is a good strategy to adopt the KKT condition because, in the real solution, the components corresponding to these index sets are non-zero; therefore, if they are zero currently, the KKT condition must have been violated severely, so prioritizing the index sets violating the KKT condition can find these components quickly [39], [46].

In the mmWave channel estimation, we only need to perform proximal iterations on the estimated support set W^t. However, the computational cost of identifying the estimated support set is comparable to the computational cost of performing one proximal iteration on the complete index set. Therefore, we employ an efficient strategy to perform multiple proximal iterations after identifying the estimated support set W^t [39]. By identifying the estimated support set W^t, the large-scale problem (14) is reduced to the following small-scale subproblems

$$\mathbf{h}_{W^t}^{t+1} = \arg\min_{\mathbf{h}} \left\{ \frac{1}{2}\left\|\mathbf{\Phi}_{W^t}\mathbf{h} - \mathbf{y}\right\|_2^2 + \lambda\left\|\mathbf{h}\right\|_1 \right\}. \tag{16}$$

Let $\mathbf{h}_{(W^t)^c}^{t+1} = \mathbf{0}$, where $(W^t)^c = (W(\mathbf{h}^t))^c = \{1, 2, \cdots, N\} \setminus W(\mathbf{h}^t)$ denotes the complement of the estimated support W^t. To illustrate the benefit of this simplification, it is assumed that all nonzero elements of the channel vector $\mathbf{h}$ are in W^t. Therefore, it takes only one iteration from the initial point zero to accurately estimate $\mathbf{h}$ for the original channel estimation problem, and the dimension of the channel estimation problem is reduced from $2Q \times 2N$ to $2Q \times |2W^t|$.

B. The FISTAB to Solve the Subproblems

Standard convex optimization software, PDG, FISTA, and ADMM algorithms are usually employed to solve (16). Calculating the Lipschitz constant $D^z = \left|\mathbf{\Phi}_{W^z}^T\mathbf{\Phi}_{W^z}\right|$ for each iteration is computationally expensive, where z denotes the zth internal iteration index. To achieve fast convergence and lower computational complexity, FISTAB [44] is used to solve the ℓ_1-IRA framework subproblems, enhancing channel estimation accuracy and convergence speed.

For ease of analysis, let

$$\begin{aligned} \mathscr{F}(\mathbf{h}) &= \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h} - \mathbf{y}\right\|_2^2 + \lambda\left\|\mathbf{h}\right\|_1 \\ &= f(\mathbf{h}) + \lambda g(\mathbf{h}), \end{aligned} \tag{17}$$

where $f(\mathbf{h}) = \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h} - \mathbf{y}\right\|_2^2$, and $g(\mathbf{h}) = \left\|\mathbf{h}\right\|_1$. First, at a given point $\mathbf{h}^z$, approximating $f(\mathbf{h})$ by the Taylor expansion, the second-order Taylor expansion of $f(\mathbf{h})$ is denoted by

$$\begin{aligned} f(\mathbf{h}) = f(\mathbf{h}^z) + (\mathbf{h} - \mathbf{h}^z)^T \triangledown f(\mathbf{h}^z) + \\ \frac{1}{2}(\mathbf{h} - \mathbf{h}^z)^T \mathcal{H}_f(\mathbf{h}^z)(\mathbf{h} - \mathbf{h}^z) + \cdots \end{aligned} \tag{18}$$

where $\mathcal{H}_f(\mathbf{h}^z)$ is the Hessian matrix of $f(\mathbf{h})$ at point $\mathbf{h}^z$. For the function $f(\mathbf{h})$, $\triangledown f(\mathbf{h}) = \mathbf{\Phi}_{W^z}^T(\mathbf{\Phi}_{W^z}\mathbf{h} - \mathbf{y})$ and $\mathcal{H}_f(\mathbf{h}) = \mathbf{\Phi}_{W^z}^T\mathbf{\Phi}_{W^z}$ can be computed separately, and then (18) can be transformed into

$$\begin{aligned} f(\mathbf{h}) = \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h}^z - \mathbf{y}\right\|_2^2 + \left\langle \triangledown f(\mathbf{h}^z), \mathbf{h} - \mathbf{h}^z \right\rangle \\ + \frac{1}{2}(\mathbf{h} - \mathbf{h}^z)^T \mathcal{H}_f(\mathbf{h}^z)(\mathbf{h} - \mathbf{h}^z) \end{aligned} \tag{19}$$

By approximating the Hessian matrix of $f(\mathbf{h})$ with the minimal Lipschitz constant D^z of the gradient $\triangledown f(\mathbf{h})$, (19) can be further transformed into

$$f(\mathbf{h}) \approx \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h}^z - \mathbf{y}\right\|_2^2 + \left\langle \triangledown f(\mathbf{h}^z), \mathbf{h} - \mathbf{h}^z \right\rangle + \frac{D^z}{2}\left\|\mathbf{h} - \mathbf{h}^z\right\|_2^2. \tag{20}$$

Then problem (17) can be further formulated as

$$\begin{aligned}
&\mathcal{Q}_{D^z}(\mathbf{h}, \mathbf{h}^z) \\
&= \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h}^z - \mathbf{y}\right\|_2^2 + \left\langle \triangledown f(\mathbf{h}^z), \mathbf{h} - \mathbf{h}^z \right\rangle + \\
&\quad \frac{D^z}{2}\left\|\mathbf{h} - \mathbf{h}^z\right\|_2^2 + \lambda\left\|\mathbf{h}\right\|_1 \\
&= \frac{1}{2}\left\|\mathbf{\Phi}_{W^z}\mathbf{h}^z - \mathbf{y}\right\|_2^2 + \frac{D^z}{2}\left\|\mathbf{h} - \mathbf{h}^z + \frac{1}{D^z}\triangledown f(\mathbf{h}^z)\right\|_2^2 \\
&\quad - \frac{1}{2D^z}\left\|\triangledown f(\mathbf{h}^z)\right\|_2^2 + \lambda\left\|\mathbf{h}\right\|_1 \\
&= \frac{D^z}{2}\left\|\mathbf{h} - \left(\mathbf{h}^z - \frac{1}{D^z}\triangledown f(\mathbf{h}^z)\right)\right\|_2^2 + B(\mathbf{h}^z) + \lambda\left\|\mathbf{h}\right\|_1
\end{aligned} \tag{21}$$

where the term $B(\mathbf{h}^z) = 1/2\left\|\mathbf{\Phi}_{W^z}\mathbf{h}^z - \mathbf{y}\right\|_2^2 - 1/(2D^z)\left\|\triangledown f(\mathbf{h}^z)\right\|_2^2$ is a constant in (21) with respect to the channel vector $\mathbf{h}$. By ignoring the constant term, then (21) is equivalent to the following optimization problem

$$\mathcal{P}_{D^z}(\mathbf{h}^z) = \arg\min_{\mathbf{h}}\left\{\lambda\left\|\mathbf{h}\right\|_1 + \frac{D^z}{2}\left\|\mathbf{h} - (\mathbf{h}^z - D^z\triangledown f(\mathbf{h}^z))\right\|_2^2\right\}. \tag{22}$$

To solve (22), the FISTAB technique then adopts the following iterative steps

$$\mathbf{h}^z = \mathcal{P}_{D^z}\left(\mathbf{h}^{z-1}\right), \tag{23}$$

$$\nu^{z+1} = \frac{1 + \sqrt{1 + 4\left(\nu^z\right)^2}}{2}, \tag{24}$$

$$\mathbf{h}^z = \mathbf{h}^z + \frac{\nu^z - 1}{\nu^{z+1}}\left(\mathbf{h}^z - \mathbf{h}^{z-1}\right), \tag{25}$$

where (23) is an iterative form of (22), ν is a positive sequence in (24), and (25) represents the updating of $\mathbf{h}^z$ with a linear combination of the two previous points $\left\{\mathbf{h}^z, \mathbf{h}^{z-1}\right\}$.

Let $S(\mathbf{\Phi}) = \sum_{i=1}^{|\mathbf{\Phi}|}\left\|\mathbf{\Phi}_i\right\|$, where $|\mathbf{\Phi}|$ and $\mathbf{\Phi}_i$ denote the number of columns and the jth column of $\mathbf{\Phi}$, respectively, then the initial value of the FISTAB algorithm is $D_0^{t+1} = S^t D_{\bar{\eta}^t}^t$ which can be estimated by using the previous information, where $S^t = (1-\alpha)S^{t-1} + (\alpha|S(\mathbf{\Phi}_{W^{t+1}})|)/(\eta|S(\mathbf{\Phi}_{W^t})|)$, $S^0 = 0$ and $\alpha > 0$. Since W^t tends to stabilise, then S^t tends to be a positive number.

Finally, **Algorithm** 1 summarizes the solution procedure for ℓ_1-IRA-FISTAB. Moreover, we have

$$\mathbf{g}_j(f, \mathbf{h}) = \begin{cases} \max\left\{\left|\frac{\partial f}{\partial \mathbf{h}_j}(\mathbf{h})\right| - \lambda, \min\left\{\left|\frac{\partial f}{\partial \mathbf{h}_j}(\mathbf{h}) + \lambda \text{sign}(\mathbf{h}_j)\right|, \right.\right. \\ \quad \left.\left.|\mathbf{h}_j|\right\}\right\}, \mathbf{h}_j \neq 0, \\ \max\left\{0, \left|\frac{\partial f}{\partial \mathbf{h}_j}(\mathbf{h})\right| - \lambda\right\}, \ \mathbf{h}_j = 0, \end{cases} \tag{26}$$

Algorithm 1 ℓ_1-IRA-FISTAB-based beamspace channel estimation method for resolving (14)

Initialization: The receive signal $\mathbf{y}$, sensing matrix $\mathbf{\Phi}$; parameter η, ε and ξ, the maximum iteration $T_{\max}$, $Z_{\max}$;

```
for t = 1, ..., T_max do
    Update the estimated support set W^t according to (15);
    Caculate the minimum Liplitz constant D^z;
    Let ĥ^0 = h^t_{W^t};
    for z = 1, ..., Z_max do
        Find the smallest nonnegative integers i_z such that with D̄ = η^{i^z} D_t^{z-1};
        F(P_D̄(h^z)) ⩽ Q_D̄(P_D̄(h^z), h^z)
        Repeat the iterative steps according to (23-25);
        if ||h^z − h^{z-1}||_2^2 / ||h^{z-1}||_2^2 ⩽ ε then
            Break;
        end if
        z = z + 1;
    end for
    h^{t+1}_{W^t} = h^z, h^{t+1}_{(W^t)^c} = 0,
    if ||g(f, h^t)|| < ξ then
        Break;
    end if
    t = t + 1;
    ĥ = h^{t+1};
end for
```

Output: The estimated channel vector $\hat{\mathbf{h}}$

where $\mathbf{g}(f, \mathbf{h})$ denotes the KKT residual of the ℓ_1-based optimization problem, and $\|\mathbf{g}(f, \mathbf{h}^t)\|$ is a measure [47] of the accuracy of the solution $\hat{\mathbf{h}}$ of (17).

IV. ℓ_{1-2} Minimization Problem-based Dual Loops Beamspace Channel Estimation Approaches

Traditional compressive sensing algorithms are primarily designed for non-coherent systems. In strongly coherence systems, these algorithms struggle to distinguish the channel components, resulting in suboptimal sparse solution for the channel vector. To address this issue, , we design a dual-loop beamspace channel estimation method based on the nonconvex ℓ_{1-2} minimization problem, which enhances the sparsity of the mmWave channel. In the outer loop, a nonlinear search algorithm, i.e., DCA, is used to solve the ℓ_{1-2} minimization problem. To reduce the algorithm complexity, we devise an IRA framework based on the ℓ_{1-2} minimization problem by exploiting the sparsity of beamspace channels. In the inner loop, a fast ADMM algorithm is utilized to solve the subproblems of the outer loop.

A. The coherent problem

In the mmWave massive MIMO systems, many problems are coherent; for instance, the coherence between users and BS varies with distance, exhibiting strong coherence when users are close to the BS. On the contrary, the coherence weakens as the distance between users and BS increases. However, the traditional compressive sensing approaches are suitable

for non-coherent conditions. In highly coherent scenarios, these algorithms struggle to distinguish channel components and provide sparse solutions. To tackle this challenge, Yin *et al.* [48] suggested an ℓ_{1-2} minimization algorithm based on nonconvex and Lipschitz continuous metrics, which has been widely applied in several fields, including magnetic resonance imaging (MRI), seismic attenuation compensation, phantom image-restoration problems.

To measure the coherence of the perception matrix $\mathbf{\Phi}$, we define the coherence coefficient [48] of $\mathbf{\Phi}$ as follows

$$\mu\left(\mathbf{\Phi}\right)=\max_{i\neq j}\frac{\left|\mathbf{\Phi}_i^T\mathbf{\Phi}_j\right|}{\left\|\mathbf{\Phi}_i\right\|_2\left\|\mathbf{\Phi}_j\right\|_2}, \tag{27}$$

where $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ are the arbitrary two columns of $\mathbf{\Phi}$.

B. ℓ_{1-2}-based IRA Framework

To enhance signal sparsity and recovery performance, we adopt the ℓ_{1-2} norm [48] instead of the ℓ_1 norm to better approximate the ℓ_0 norm term [39]. Hence, we employ a convex relaxation technique to convert (13) into an unconstrained ℓ_{1-2} minimization form

$$\mathbf{h}^*=\arg\min_{\mathbf{h}}\left\{\frac{1}{2}\left\|\mathbf{\Phi h}-\mathbf{y}\right\|^2+\lambda\left\|\mathbf{h}\right\|_1-\lambda\left\|\mathbf{h}\right\|_2\right\}, \tag{28}$$

where the objective function in (28) is a nonconvex optimization problem, and we employ the DCA for solving nonconvex optimization problems.

Specifically, the DCA minimizes the objective function $\mathcal{F}\left(\mathbf{h}\right)=\mathcal{G}\left(\mathbf{h}\right)-\mathcal{H}\left(\mathbf{h}\right)$, where $\mathcal{G}\left(\mathbf{h}\right)$ and $\mathcal{H}\left(\mathbf{h}\right)$ are lower semicontinuous proper convex functions, and $\mathcal{G}\left(\mathbf{h}\right)-\mathcal{H}\left(\mathbf{h}\right)$ is the DC decomposition of $\mathcal{F}\left(\mathbf{h}\right)$.

The DCA constructs two candidate sequences, $\{\mathbf{h}^s\}$ and $\{\mathbf{d}^s\}$, representing the optimal solutions of the primal and dual problems, respectively. To perform the DCA, one iteratively computes

$$\begin{cases}\mathbf{d}^s\in\partial\mathcal{H}\left(\mathbf{h}^s\right),\\ \mathbf{h}^{s+1}=\arg\min_{\mathbf{h}}\mathcal{G}\left(\mathbf{h}\right)-\left(\mathcal{H}\left(\mathbf{h}\right)+\left\langle\mathbf{d}^s,\mathbf{h}-\mathbf{h}^s\right\rangle\right),\end{cases} \tag{29}$$

where s denotes the sth external iteration index, $\mathbf{d}^s\in\partial\mathcal{H}\left(\mathbf{h}^s\right)$ denotes that $\mathbf{d}^s$ is the subgradient of $\mathcal{H}\left(\mathbf{h}\right)$ in $\mathbf{h}^s$ [49]. Because $\mathcal{F}\left(\mathbf{h}\right)$ is bounded from below [48], then the objective function value of the DCA is convergent, and the objective function in (28) is converted into the following two convex decompositions

$$\mathcal{F}\left(\mathbf{h}\right)=\left(\frac{1}{2}\left\|\mathbf{\Phi h}-\mathbf{y}\right\|_2^2+\lambda\left\|\mathbf{h}\right\|_1\right)-\alpha\lambda\left\|\mathbf{h}\right\|_2, \tag{30}$$

where $\mathcal{G}\left(\mathbf{h}\right)=\frac{1}{2}\left\|\mathbf{\Phi h}-\mathbf{y}\right\|_2^2+\lambda\left\|\mathbf{h}\right\|_1$ and $\mathcal{H}\left(\mathbf{h}\right)=-\alpha\lambda\left\|\mathbf{h}\right\|_2$ are two convex decompositions. $\left\|\mathbf{h}\right\|_2$ is differentiable and its subgradient is denoted as

$$\mathbf{d}^s=\begin{cases}\mathbf{0}, & \text{if } \mathbf{h}^s=\mathbf{0},\\ -\alpha\lambda\frac{\mathbf{h}^s}{\left\|\mathbf{h}^s\right\|_2}, & \text{otherwise.}\end{cases} \tag{31}$$

According to the DCA iteration formula in (31), each DCA iteration involves solving an ℓ_1-regularized convex subproblem of the following form

$$\mathbf{h}^{s+1}=\begin{cases}\arg\min_{\mathbf{h}}\frac{1}{2}\left\|\mathbf{\Phi h}-\mathbf{y}\right\|_2^2+\lambda\left\|\mathbf{h}\right\|_1, \text{if } \mathbf{h}^r=\mathbf{0},\\ \arg\min_{\mathbf{h}}\frac{1}{2}\left\|\mathbf{\Phi h}-\mathbf{y}\right\|_2^2+\lambda\left\|\mathbf{h}\right\|_1\\ \qquad -\left\langle\mathbf{h},\alpha\lambda\frac{\mathbf{h}^s}{\left\|\mathbf{h}^s\right\|_2}\right\rangle, \text{ otherwise.}\end{cases} \tag{32}$$

To simplify the beamspace channel estimation problem, with the identified estimation supports sets V^s, the large-scale problem (28) based on ℓ_{1-2} minimization is reduced to the following small-scale subproblems

$$\hat{\mathbf{h}}_{V^s}^{s+1}=\arg\min_{\mathbf{h}}\left\{\frac{1}{2}\left\|\mathbf{\Phi}_{V^s}\mathbf{h}-\mathbf{y}\right\|_2^2-\left\langle\mathbf{h},\lambda\alpha\frac{\mathbf{h}_{V^s}^s}{\left\|\mathbf{h}^s\right\|_2}\right\rangle+\lambda\left\|\mathbf{h}\right\|_1\right\}, \tag{33}$$

where the support set V^s is defined acording to the KKT condition, as follows

$$\begin{cases}V_1\left(\mathbf{h}^s\right)=\left\{i\left|\mathbf{h}_i^s\neq 0\right.\right\},\\ V_2\left(\mathbf{h}^s\right)=\left\{i\left|\mathbf{h}_i^s=0 \text{ and } \right|\mathbf{\Phi}_i^T\left(\mathbf{\Phi h}^s-\mathbf{y}\right)-\right.\\ \qquad\qquad \left.\alpha\lambda\mathbf{h}^s/\left\|\mathbf{h}^s\right\|_2\right|>\left(1-\tau\right)\lambda\right\},\\ V\left(\mathbf{h}^s\right)=V_1\left(\mathbf{h}^s\right)\cup V_2\left(\mathbf{h}^s\right),\end{cases} \tag{34}$$

where $V_1\left(\mathbf{h}^s\right)$ denotes the index set of the non-zero components in the sth iteration of $\mathbf{h}$, and $V_2\left(\mathbf{h}^s\right)$ denotes the index set that satisfies the zero component in the sth iteration of $\mathbf{h}$ and the current KKT condition, $V\left(\mathbf{h}^s\right)$ denotes the estimated support set. For simplicity, we abbreviate $V_1\left(\mathbf{h}^r\right)$, $V_2\left(\mathbf{h}^s\right)$, and $V\left(\mathbf{h}^s\right)$ as V_1^s, V_2^s, and V^s, respectively.

C. ADMM to Solve the Subproblems

ADMM effectively solve large-scale optimization problems by integrating the augmented Lagrangian method with dual decomposition. Therefore, we utilize the ADMM algorithm to solve small-scale subproblems of ℓ_{1-2} minimization after simplifying the ℓ_{1-2}-IRA framework. We introduce the auxiliary variable $\mathbf{z}$ to derive the ADMM solver [45], and (33) is equivalent to the following constrained minimization problem

$$\begin{aligned}\hat{\mathbf{h}}_{V^r}^{r+1}=&\\ &\arg\min_{\mathbf{h}}\left\{\frac{1}{2}\left\|\mathbf{\Phi}_{V^r}\mathbf{h}-\mathbf{y}\right\|_2^2-\left\langle\mathbf{h},\lambda\alpha\frac{\mathbf{h}_{V^r}^r}{\left\|\mathbf{h}^r\right\|_2}\right\rangle+\lambda\left\|\mathbf{z}\right\|_1\right\},\\ &\text{s.t. } \mathbf{h}-\mathbf{z}=0,\end{aligned} \tag{35}$$

where r denotes the rth internal iteration index. Then, the augmented Lagrangian function is expressed as

$$\begin{aligned}\mathcal{L}_\rho\left(\hat{\mathbf{h}},\mathbf{z},\mathbf{w}\right)=&\frac{1}{2}\left\|\mathbf{\Phi}_{V^r}\mathbf{h}-\mathbf{y}\right\|_2^2-\left\langle\mathbf{h},\lambda\alpha\frac{\mathbf{h}_{V^r}^r}{\left\|\mathbf{h}^r\right\|_2}\right\rangle\\ &+\lambda\left\|\mathbf{z}\right\|_1+\mathbf{w}^T\left(\mathbf{h}-\mathbf{z}\right)+\frac{\rho}{2}\left\|\mathbf{h}-\mathbf{z}\right\|_2^2,\end{aligned} \tag{36}$$

where $\mathbf{w}$ is the Lagrange multiplier and $\rho > 0$ is the penalty factor. ADMM iteratively updates the variables $\mathbf{h}$, $\mathbf{z}$, $\mathbf{w}$ by minimizing the generalized Lagrange function

$$\begin{cases} \mathbf{z}^{r+1} = \arg\min_{\mathbf{z}} \mathcal{L}_\rho \left(\hat{\mathbf{h}}^r, \mathbf{z}^r, \mathbf{w}^r\right), \\ \hat{\mathbf{h}}^{r+1} = \arg\min_{\mathbf{h}} \mathcal{L}_\rho \left(\hat{\mathbf{h}}, \mathbf{z}^{r+1}, \mathbf{w}^r\right), \\ \mathbf{w}^{r+1} = \mathbf{w}^r + \rho\left(\hat{\mathbf{h}}^{r+1} - \mathbf{z}^{r+1}\right). \end{cases} \tag{37}$$

The update variables $\mathbf{z}$ and $\mathbf{h}$ have closed-form solutions. Specifically, the update variable $\mathbf{z}$ is solved by soft thresholding

$$\mathbf{z}^{r+1} = \mathcal{S}\left(\hat{\mathbf{h}}^r + \frac{\mathbf{w}^r}{\rho}, \frac{\lambda}{\rho}\right), \tag{38}$$

where the soft thresholding operation is denoted as $\left(\mathcal{S}\left(\hat{\mathbf{h}}, \beta\right)\right)_i = \operatorname{sign}(h_i)\max\{|h_i| - \beta, 0\}$. The variable $\mathbf{h}$ is updated by the gradient method, as follows

$$\begin{aligned} \hat{\mathbf{h}}^{r+1} = &\left(\mathbf{\Phi}_{V^r}^T \mathbf{\Phi}_{V^r} + \rho \mathbf{I}_{N\times N}\right)^{-1} \left(\mathbf{\Phi}_{V^r}^T \mathbf{y} - \mathbf{d}^r \right. \\ &\left. + \rho \mathbf{z}^{r+1} - \mathbf{w}^r\right), \end{aligned} \tag{39}$$

where $\left(\mathbf{\Phi}_{V^r}^T \mathbf{\Phi}_{V^r} + \rho \mathbf{I}_{N\times N}\right)^{-1}$ requires computing the matrix inverse, increasing the computational complexity for large matrices. Therefore, we employ a Cholesky factorization to approximate $\left(\mathbf{\Phi}_{V^r}^T \mathbf{\Phi}_{V^r} + \rho \mathbf{I}_{N\times N}\right)^{-1}$. Let $\mathcal{F}^r = \mathbf{\Phi}_{V^r}^T \mathbf{y} - \mathbf{d}^r + \rho \mathbf{z}^{r+1} - \mathbf{w}^r$, then (36) can be converted to a more computationally efficient form

$$\hat{\mathbf{h}}^{r+1} = \mathbf{U}^{-1}\mathbf{L}^{-1}\left(\mathbf{\Phi}_{V^r}^T \mathbf{y} - \mathbf{d}^r + \rho \mathbf{z}^{r+1} - \mathbf{w}^r\right), \tag{40}$$

where $\mathbf{U}$ and $\mathbf{L}$ denote Cholesky factorization of $\mathbf{\Phi}_{V^r}^T \mathbf{\Phi}_{V^r} + \rho \mathbf{I}_{N\times N}$, i.e., $\mathbf{\Phi}_{V^r}^T \mathbf{\Phi}_{V^r} + \rho \mathbf{I}_{N\times N} = \mathbf{L}\mathbf{U}$.

Finally, Algorithm 2 summarizes the solution procedure for the ℓ_{1-2}-IRA-ADMM algorithm, and the stopping iteration criterion for the internal loop is denoted by

$$\begin{cases} \|\varsigma^r\|_2 \leqslant \sqrt{2N}\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\left\{\|\mathbf{h}^r\|_2, \|\mathbf{z}^r\|_2\right\}, \\ \|\eta^r\|_2 \leqslant \sqrt{2N}\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \|\mathbf{y}^r\|_2, \end{cases} \tag{41}$$

where $\varsigma^r = \mathbf{h}^r - \mathbf{z}^r$ and $\eta^r = \rho\left(\mathbf{z}^r - \mathbf{z}^{r-1}\right)$ denote primal and dual residuals at the rth iteration, respectively. $\epsilon^{\text{abs}} > 0$ and $\epsilon^{\text{rel}} > 0$ are absolute and relative tolerance, respectively.

D. Computational Complexity Analysis

The computational complexity of Algorithms 1 and 2 impacts the processing power consumption and hardware cost. Therefore, it is crucial to analyze the computational complexity of Algorithms 1 and 2, especially for the beamspace channel in massive MIMO systems. In Algorithm 1, since the ℓ_1-IRA framework in the outer loop significantly reduces the problem scale, and FISTAB with the inner loop avoids massive matrix-vector multiplications, which reduces the computational complexity of Algorithm 1. Therefore, the computational complexity of Algorithm 1 is $\mathcal{O}\left(pQ\ln(1/\varepsilon)\right)$, where ε denotes the beamspace channel vector estimation accuracy, and p is the sparsity level of $\mathbf{h}$. The computational complexity of the subproblem of Algorithm 2 is mainly derived from $\mathbf{h}$. The computational complexity of forming $\mathbf{h}$ is $\mathcal{O}\left(N^2\right)$ through the two back-solving steps. Therefore, the computational complexity of Algorithm 2 is $\mathcal{O}\left(pN^2\right)$.

Algorithm 2 ℓ_{1-2}-IRA-ADMM-based beamspace channel estimation method for solving (28)

Initialization: The receive signal $\mathbf{y}$, sensing matrix $\mathbf{\Phi}$; parameter ρ, λ, the maximum iteration $S_{\max}$, $R_{\max}$;

for $s = 1, \cdots, S_{\max}$ **do**
  Update the estimated support set V^r according to (34);
  for $r = 1, \cdots, R_{\max}$ **do**
    Update the auxiliary vector $\mathbf{z}$ according to (38);
    Update the beamspace channel vector $\hat{\mathbf{h}}$ according to (40);
    Update the Lagrangian multipliers vector $\mathbf{w}$ according to (37);
    if satisfy the stop iteration condition **then**
      Break;
      $r = r + 1$;
    end if
  end for
  $\mathbf{h}_{V^s}^{s+1} = \mathbf{h}^s, \mathbf{h}_{(V^s)^c}^{s+1} = 0$,
  if $\|\mathbf{g}(f, \mathbf{h}^s)\| < \xi$ **then**
    Break;
  end if
  $s = s + 1$;
  $\hat{\mathbf{h}} = \mathbf{h}^{s+1}$;
end for

Output: The estimated channel vector $\hat{\mathbf{h}}$

TABLE I
COMPUTATIONAL COMPLEXITY OF THE BEAMSPACE CHANNEL ESTIMATION

Algorithm	Computational complexity in each iteration
ℓ_1-IRA-FISTAB	$pQ\ln(1/\varepsilon)$
ℓ_{1-2}-IRA-ADMM	pN^2
FISTA	QN
ℓ_{1-2}	N^3
ADMM	N^3
OMP	$pQN + p^3Q$
SD	pQN

Table I compares the computational complexity of Algorithms 1 and 2 with five comparison algorithms, i.e., FISTA [44], SD [16], ADMM [45], OMP [8], and ℓ_{1-2} minimization algorithm [48]. From Table I, it can be observed that the computational method complexity of algorithms 1 and 2 is lower than that of FISTA and ℓ_{1-2} minimization algorithms, respectively, with Algorithm 1 exhibiting the lowest complexity among all the methods.

V. SIMULATION RESULTS

In this section, we perform extensive simulations to evaluate the performance of multiuser mmWave channel estimation with lense antenna arrays. Algorithms 1 and 2 are compared with various channel estimation techniques, including OMP [8], SD [16], ADMM [45], FISTA [44], and ℓ_{1-2} minimization algorithm [48]. The system parameters are configured as follows: the number of BS with lens antenna arrays is

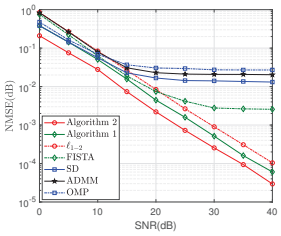


Fig. 2. Comparison of the NMSE performance with different SNRs at $Q = 160$.

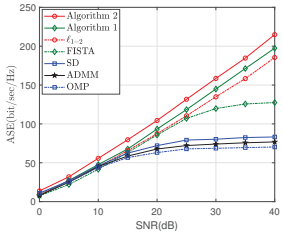


Fig. 3. Comparison of the ASE performance with different SNRs at $Q = 160$.

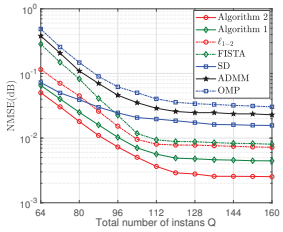


Fig. 4. Comparison of the NMSE performance with various number of time slots at SNR $= 20$ dB.

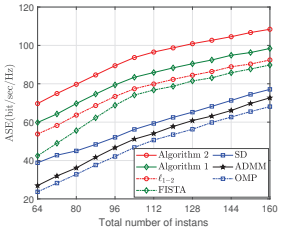


Fig. 5. Comparison of the ASE performance versus various number of time slots at SNR $= 20$ dB.

$N = 256$, the number of RF chains $N_{RF} = 16$, and the number of single-antenna users $K = 16$. The beamspace channel multipath components consist of one LoS path and two NLoS paths, and the amplitudes for the LoS and NLoS paths satisfy $\beta_{k,0} \sim \mathcal{CN}(0,1)$ and $\beta_{k,i} \sim \mathcal{CN}\left(0, 10^{-2}\right), i = 1, 2$, respectively. $\varphi_{k,0}$ and $\varphi_{k,i}$ follow the uniform distribution of independent and identically distributed (i.i.d.). $[-1/2, 1/2]$. In the implementation of Algorithms 1 and 2, the ℓ_1-IRA and ℓ_{1-2}-IRA frameworks are parameterized as $\eta = 1.2$, $\tau = 10^{-3}$, $\lambda = \lambda_c \left\|\mathbf{\Phi}^T \mathbf{y}\right\|_\infty$, $0 < \lambda_c < 1$, and $T_{\max} = \min\left(\max\left(1000, 2N\right), 10^4\right)$ [50]. For the ℓ_{1-2} minimization algorithm, we adopt an adaptive strategy to update the parameter α with an initial value of $\alpha_0 = 0.1$, and then use this strategy to adjust $\alpha^{r+1} = \begin{cases} \alpha^r, & \text{if } \mod(r,5) \neq 0 \\ \min\left\{1.5\alpha^r, 1\right\}, & \text{if } \mod(r,5) = 0 \end{cases}$, the other parameters are set to $\rho = 10^{-4}$ and $\lambda = 10^{-5}$. To evaluate the performance of the beamspace channel estimation, we compute the normalized mean square error (NMSE) and average spectral efficiency (ASE) over 100 Monte Carlo simulations, defined as follows, respectively

$$\text{NMSE} = \mathbb{E}\left[\left(\sum_{k=1}^{K} \frac{\left\|\mathbf{h}_k - \hat{\mathbf{h}}_k\right\|_2^2}{\left\|\mathbf{h}_k\right\|_2^2}\right)/K\right], \tag{42}$$

$$\text{ASE} = \mathbb{E}\left\{\log_2 \det\left(\mathbf{I}_K + \left(N_{BS} K\left(\sigma^2 + \text{NMSE}\right)\right)^{-1} \mathbf{H}\mathbf{H}^H\right)\right\} \tag{43}$$

where $\mathbf{H} = \sum_{k=1}^{K} \hat{\mathbf{h}}_k \in \mathbb{C}^{N \times K}$ denotes the estimated beamspace channel matrix.

Fig. 2 compares the NMSE performance of beamspace channel estimation of Algorithms 1 and 2 with five benchmark algorithms at different signal-to-noise ratios (SNRs). The NMSE performance of all algorithms gradually improves as the SNR increases from 0 dB to 40 dB, indicating that the influence of noise on the channel estimation accuracy gradually decreases. Algorithms 1 and 2 show significant performance advantages and robustness, especially at high SNR. Algorithms 1 and 2 consistently outperform the other benchmark algorithms, because they take full advantage of the sparse structure of the mmWave beamspace channel to quickly identify the non-zero components of the channel, thus improving the channel estimation accuracy. Compared with Algorithm 1, Algorithm 2 further enhances the sparsity of the beamspace channel in the strong coherence scenarios, i.e., $\mu(\mathbf{\Phi}) = 0.9245$. The NMSE performance of the SD, ADMM, and OMP algorithms saturates at SNR $\geqslant 25$, because the non-zero elements of the beamspace channel can be accurately estimated at high SNR, but the error that may be caused by treating the low-power components as zeros does not disappear. In contrast, the NMSE performance of Algorithms 1 and 2 and ℓ_{1-2} minimization algorithm is not affected. Specifically, at SNR $= 20$, the NMSE performance of Algorithm 2 improves by 50.43% over Algorithm 1. Additionally, Algorithm 1 achieves a 44.68% NMSE improvement over FISTA, while Algorithm 2 outperforms the ℓ_{1-2} minimization algorithm algorithm by 69.44%, demonstrating the effectiveness of the proposed approach in enhancing channel estimation accuracy.

Fig. 3 presents the ASE performance of Algorithms 1 and 2 compared to benchmark algorithms at various SNRs. As the SNR increases, the ASE performance of FISTA, SD, ADMM, and OMP algorithms increases slowly, while the ASE performance of Algorithms 1 and 2 increases linearly; this is mainly because, at high SNR, the signal is less affected by the noise and can transmit more information during the effective communication time; whereas at low SNR, the user terminals transmit fewer signals to avoid loss, resulting in lower spectral efficiency for all algorithms. Specifically, at SNR=20, the ASE performance of Algorithm 1 is improved by 8.38% and 27.66% compared to FISTA and SD, respectively. The ASE performance of Algorithm 2 improves by 10.17% and 17.32% compared to Algorithm 1 and ℓ_{1-2} minimization algorithm, respectively. This suggests that Algorithms 1 and 2 have higher information transfer capability per unit time.

Fig. 4 illustrates the channel estimation performance across different numbers of time slots, where the length of the pilot sequence equals the number of time slots. As the number of time slots increases, the estimation accuracy of all algorithms gradually improves. However, Algorithms 1 and 2 consistently achieve superior performance due to their enhanced ability to accurately identify sparse channel components. Moreover, they require the lowest pilot overhead to attain a given estimation accuracy under identical system conditions, demonstrating their efficiency. Notably, the NMSE performance of Algorithms 1 and 2 at $Q \geqslant 144$ consistently outperforms that of the other benchmark algorithms at $Q = 256$, highlighting their capability to achieve high-accuracy channel estimation with reduced pilot overhead.

Fig. 5 compares the ASE performance of different algo-

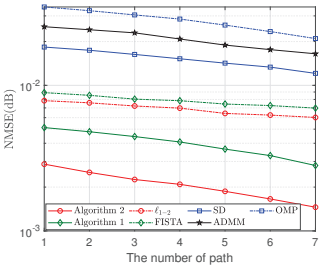


Fig. 6. Comparison of the NMSE performance versus the number of path at SNR = 20 dB.

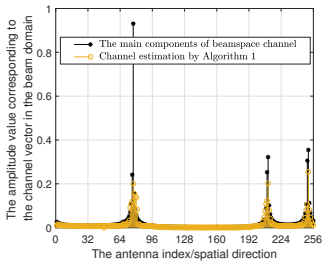


Fig. 7. Channel reconstruction performance of Algorithm 1 at SNR = 30 dB.

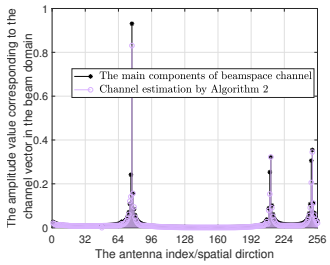


Fig. 8. Channel reconstruction performance of Algorithm 2 at SNR = 30 dB.

rithms across various time slots. As the number of time slots increases, ASE performance improves for all algorithms, with Algorithms 1 and 2 consistently achieving superior performance. Specifically, at $Q = 128$, the ASE performance of Algorithm 2 improves by 12.61% and 21.81% compared to Algorithm 1 and ℓ_{1-2} minimization algorithm, respectively, and the ASE performance of Algorithm 1 is enhanced by 21.24% compared with FISTA algorithm.

Fig. 6 presents the NMSE performance of Algorithms 1 and 2 compared to other five benchmark algorithms for different channel path numbers. As the number of paths increases, NMSE performance improves for all algorithms, with Algorithms 1 and 2 consistently outperforming the others. This is because the increasing number of channel paths increases the uncertainty in the channel, and the grid-based SD and OMP algorithms allocate more effective channel paths on a predefined angular grid, partially eliminating the effect of more paths and increasing the channel estimation complexity and possible interference. Algorithms 1 and 2 are less affected by the increase in the number of paths, because the IRA framework simplifies the dimension of the channel estimation problem and identifies the support set of the channel, while the ℓ_{1-2} minimization algorithm enhances the sparsity of the channel. Specifically, the NMSE performance of Algorithm 1 improves by 48.58% compared to FISTA at $L_k = 5$. The NMSE performance of Algorithm 2 is improved by 48.81% and 70.83% compared to Algorithm 1 and ℓ_{1-2} minimization algorithm, respectively.

Figs. 7 and 8 present the channel reconstruction performance of Algorithms 1 and 2, where Algorithm 2 more accurately detects channel components than Algorithm 1. In addition, Algorithms 1 and 2 identify the effective support for each transmission path (LoS and NLoS), determine the coordinates of the most effective beam (the one with the maximum power) in each path, and then detect the remaining non-zero components of the channel. This allows the coordinates of the non-zero components to be correctly identified, thus reducing the NMSE value of the channel estimation.

VI. Conclusion

This paper addressed multi-user mmWave channel estimation problem using the lens antenna array. First, we transformed the beamspace channel estimation problem into a sparse signal recovery problem by leveraging the sparse structure of the beamspace channel. We then developed a ℓ_1-IRA-FISTAB-based beamspace channel estimation method, which exploited the ℓ_1-IRA framework to simplify the large-scale channel estimation problem into a series of small-scale subproblems in the outer loop; and leveraged FISTAB algorithm to solve the subproblems in the inner loop. Additionally, we proposed a ℓ_{1-2}-IRA-ADMM-based dual loops method for beamspace channel estimation in the strongly coherent systems. Simulation results demonstrated that these dual-loop methods improved higher channel estimation accuracy with a small pilot overhead compared with traditional beasmspace channel estimation techniques.

References

[1] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami, "Cellular-connected wireless virtual reality: Requirements, challenges, and solutions," *IEEE Commun Mag.*, vol. 58, no. 5, pp. 105–111, May 2020.

[2] A. Dhar Dwivedi, R. Singh, K. Kaushik, R. Rao Mukkamala, and W. S. Alnumay, "Blockchain and artificial intelligence for 5G-enabled internet of things: Challenges, opportunities, and solutions," *Transactions on Emerging Telecommunications Technologies*, vol. 35, no. 4, p. e4329, 2024.

[3] Z. Gao, M. Ke, Y. Mei, L. Qiao, S. Chen, D. W. K. Ng, and H. V. Poor, "Compressive sensing-based grant-free massive access for 6G massive communication," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7411–7435.

[4] Y. Zhou, L. Liu, L. Wang, N. Hui, X. Cui, J. Wu, Y. Peng, Y. Qi, and C. Xing, "Service-aware 6g: An intelligent and open network based on the convergence of communication, computing and caching," *Digital Communications and Networks*, vol. 6, no. 3, pp. 253–260, 2020.

[5] Z. Wang, J. Zhang, H. Du, E. Wei, B. Ai, D. Niyato, and M. Debbah, "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," *IEEE Wirel Commun.*, vol. 31, no. 3, pp. 117–124, 2024.

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

[7] B. Sun, Y. Zhou, J. Yuan, Y.-F. Liu, L. Wang, and L. Liu, "High order psk modulation in massive mimo systems with 1-bit adcs," *IEEE Trans Wirel Commun.*, vol. 20, no. 4, pp. 2652–2669, 2020.

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

[9] B. Sun, Y. Zhou, J. Yuan, and J. Shi, "Interference cancellation based channel estimation for massive mimo systems with time shifted pilots," *IEEE Trans Wirel Commun.*, vol. 19, no. 10, pp. 6826–6843, 2020.

[10] R. J. Jayakumar and A. C. A. Arul, "A survey on beamspace millimeter-wave massive mimo systems: An overview of open issues, challenges, and future research trends," *International Journal of Sensors Wireless Communications and Control*, vol. 14, no. 1, pp. 1–20, Mar 2024.

[11] D. Shi, L. Song, X. Gao, J. Wang, M. Bengtsson, G. Y. Li, and X.-G. Xia, "Beam structured channel estimation for HF skywave massive MIMO-OFDM communications," *IEEE Trans Wirel Commun.*, vol. in press, 2024.

[12] P.-C. Chen and P. Vaidyanathan, "Channel estimation for mmWave using the convolutional beamspace approach," *IEEE Trans Signal Process.*, pp. 1–16, 2024.

[13] L. Zhu, Y. Xiong, Z. Li, Y. Guan, Z. Chu, Z. Zhu, P. Xiao, and C.-L. Wang, "A novel gridless uplink/downlink channel estimation method for millimeter wave MIMO-OFDM systems," *IEEE Trans Wirel Commun.*, 2025.

[14] L. Zhu, Z. Li, R. Zhang, Z. Chu, D. Mi, Z. Zhu, G. Chen, and L. Zhen, "New gridless method-based channel estimation for millimeter wave MIMO-OFDM systems," *IEEE Commun Lett.*, vol. 28, no. 9, pp. 2166–2170, 2024.

[15] L. Zhu, K.-H. Liu, L. Wan, and L. Sun, "Active user detection and channel estimation via fast ADMM," in *2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom*. IEEE, May 2023, pp. 1–6.

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

[17] X. Long, K. Song, Y. Luo, Y. Liu, J. Ma, and T. Qiu, "Beamspace channel estimation based on block support detection for millimeter-wave massive MIMO systems," in *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China*. IEEE, 2022, pp. 1–6.

[18] E. Sharifi, M. M. Feghhi, G. Azarnia, S. Nouri, D. Lee, and M. J. Piran, "Channel estimation based on compressed sensing for massive MIMO systems with lens antenna array," *IEEE Access*, vol. 11, pp. 79 016–79 032, July 2023.

[19] R. Zhang, L. Yang, M. Tang, W. Tan, and J. Zhao, "Channel estimation for mmwave massive MIMO systems with mixed-ADC architecture," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 606–613, Feb 2023.

[20] J. Tao, C. Qi, and Y. Huang, "Regularized multipath matching pursuit for sparse channel estimation in millimeter wave massive MIMO system," *IEEE Wirel Commun Lett.*, vol. 8, no. 1, pp. 169–172, Aug 2018.

[21] N.-S. Duong, Q.-T. Nguyen, and T.-M. Dinh-Thi, "OMP-based channel estimation with dynamic grid for mmwave MIMO positioning systems," *IEEE Commun Lett.*, vol. 27, no. 10, pp. 2623–2627, 2023.

[22] Y. Li and A. Hu, "Channel estimation via subcarrier grouping for wideband mmwave hybrid massive MIMO-OFDM systems," *IEEE Trans Green Commun Netw.*, vol. 8, no. 1, pp. 64–78, 2024.

[23] L. Qiao, J. Zhang, Z. Gao, D. Zheng, M. J. Hossain, Y. Gao, D. W. K. Ng, and M. Di Renzo, "Joint activity and blind information detection for UAV-assisted massive IoT access," *IEEE J Sel Areas Commun.*, vol. 40, no. 5, pp. 1489–1508, 2022.

[24] X. Cheng, J. Deng, and S. Li, "Wideband channel estimation for millimeter wave beamspace MIMO," *IEEE Trans Veh Technol.*, vol. 70, no. 7, pp. 7221–7225, May 2021.

[25] C. Ruan, Z. Zhang, H. Jiang, H. Zhang, J. Dang, and L. Wu, "Simplified learned approximate message passing network for beamspace channel estimation in mmwave massive MIMO systems," *IEEE Trans Wirel Commun.*, vol. 23, no. 5, pp. 5142–5156, 2024.

[26] L. Qiao, Z. Gao, M. B. Mashhadi, and D. Gündüz, "Massive digital over-the-air computation for communication-efficient federated edge learning," *arXiv preprint arXiv:2405.15969*, 2024.

[27] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Deep learning-based channel estimation for wideband hybrid mmwave massive MIMO," *IEEE Trans Commun.*, vol. 71, no. 6, pp. 3679–3693, 2023.

[28] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, "Beamspace channel estimation for wideband millimeter-wave MIMO: A model-driven unsupervised learning approach," *IEEE Trans Wirel Commun.*, vol. 22, no. 3, pp. 1808–1822, Sep 2022.

[29] Y. Zhang, M. El-Hajjar, and L.-l. Yang, "Multi-layer sparse bayesian learning for mmwave channel estimation," *IEEE Trans Veh Technol.*, vol. 73, no. 3, pp. 3485–3498, 2024.

[30] J. Gao, C. Zhong, and G. Y. Li, "AMP-SBL unfolding for wideband mmwave massive MIMO channel estimation," in *2023 IEEE International Conference on Communications Workshops (ICC Workshops), Rome, Italy*. IEEE, Oct 2023, pp. 60–65.

[31] P. Zheng, X. Lyu, and Y. Gong, "Trainable proximal gradient descent based channel estimation for mmwave massive MIMO systems," *IEEE Wirel Commun Lett.*, vol. 12, no. 10, pp. 1781–1785, 2023.

[32] Z. Li, Q. Xue, C. Dong, K. Niu, H. Wang, Q. Huang, Q. Gao, Y. Fei, and J. Zuo, "Deep learning beamspace channel estimation for mmWave massive MIMO with switch-based selection network," in *2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United*. IEEE, 2024, pp. 1–6.

[33] X. Wei, C. Hu, and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems," *IEEE Trans Commun.*, vol. 69, no. 1, pp. 182–193, Sep 2020.

[34] L. Qiao, A. Liao, Z. Li, H. Wang, Z. Gao, X. Gao, Y. Su, P. Xiao, L. You, and D. W. K. Ng, "Sensing user's activity, channel, and location with near-field extra-large-scale MIMO," *IEEE Trans Commun.*, vol. 72, no. 2, pp. 890–906, Oct 2023.

[35] R. Zhang, L. Cheng, S. Wang, Y. Lou, Y. Gao, W. Wu, and D. W. K. Ng, "Integrated sensing and communication with massive MIMO: A unified tensor approach for channel and target parameter estimation," *IEEE Trans Wirel Commun.*, vol. 23, no. 8, pp. 8571–8587, 2024.

[36] R. Zhang, X. Wu, Y. Lou, F.-G. Yan, Z. Zhou, W. Wu, and C. Yuen, "Channel training-aided target sensing for terahertz integrated sensing and massive MIMO communications," *IEEE Internet Things J.*, pp. 1–1, 2024.

[37] R. Zhang, L. Cheng, W. Zhang, X. Guan, Y. Cai, W. Wu, and R. Zhang, "Channel estimation for movable-antenna MIMO systems via tensor decomposition," *IEEE Wirel Commun Lett.*, vol. 13, no. 11, pp. 3089–3093, 2024.

[38] R. Zhang, L. Cheng, S. Wang, Y. Lou, W. Wu, and D. W. K. Ng, "Tensor decomposition-based channel estimation for hybrid mmwave massive MIMO in high-mobility scenarios," *IEEE Trans Commun.*, vol. 70, no. 9, pp. 6325–6340, 2022.

[39] G. Wang, W. Yu, X. Liang, Y. Wu, and B. Yu, "An iterative reduction FISTA algorithm for large-scale LASSO," *SIAM J Sci Comput.*, vol. 44, no. 4, pp. A1989–A2017, 2022.

[40] A. Garg, S. Srivastava, N. Yadav, A. K. Jagannatham, and L. Hanzo, "Angularly sparse channel estimation in dual-wideband Tera-Hertz (THz) hybrid MIMO systems relying on bayesian learning," *arXiv preprint arXiv:2402.12158*, 2024.

[41] Y. Li, Y. Luo, X. Wu, Z. Shi, S. Ma, and G. Yang, "Variational bayesian learning based localization and channel reconstruction in RIS-aided systems," *IEEE Trans Wirel Commun.*, vol. in press, 2024.

[42] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans Inf Theory.*, vol. 51, no. 12, pp. 4203–4215, 2005.

[43] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[44] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[45] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[46] G. Wang, X. Wei, B. Yu, and L. Xu, "An efficient proximal block coordinate homotopy method for large-scale sparse least squares problems," *SIAM J Sci Comput.*, vol. 42, no. 1, pp. A395–A423, 2020.

[47] Y. Xiong, N. Wu, Y. Shen, and M. Z. Win, "Cooperative localization in massive networks," *IEEE Trans Inf Theory.*, vol. 68, no. 2, pp. 1237–1258, 2021.

[48] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of ℓ_{1-2} for compressed sensing," *SIAM J Sci Comput.*, vol. 37, no. 1, pp. A536–A563, 2015.

[49] Y. Lou and M. Yan, "Fast $L_1 - L_2$ minimization via a proximal operator," *Journal of Scientific Computing*, vol. 74, no. 2, pp. 767–785, 2018.

[50] H. Ge and P. Li, "The dantzig selector: recovery of signal via $\ell_1 - \alpha\ell_2$ minimization," *Inverse Problems*, vol. 38, no. 1, p. 015006, 2021.

Lijun Zhu (Graduate Student Member, IEEE) received M.S. degree in system engineering from Shenyang University of Technology, China. She received Ph.D degree from Beijing University of Posts and Telecommunications in June 2025. Her research interests include millimeter wave communication, channel estimation, compressed sensing and performance optimization in massive MIMO systems. She is currently an algorithm engineer in Seedlight robot (Shenzhen) Co., Ltd., Shenzheng, China,

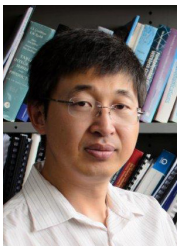
Zheng Li (Graduate Student Member, IEEE) received M.S. degree in communication engineering from Zhengzhou University, China, in 2023. He is currently pursuing the Ph.D degree with the School of Electronic and Information Engineering, Zhengzhou University. His research interests include Intelligent reflecting surface, wireless power communication network, integrated sensing and communication, compressed sensing, finite rate of innovation sampling, and array signal processing. He has served as a Technical Program Committee (TPC) member of 2025 IEEE ICC and 2025 IEEE WCSP.

Zeliang An (Member, IEEE) received the M.S. degree in Communications and Information Systems from Nanchang Aeronautical University, China in 2019 and the Ph.D. degree in the School of Communications and Information Engineering, obtaining the best doctoral dissertation award and President's honour award, from Chongqing University of Posts and Telecommunications (CQUPT), China in 2023. From 2022 to 2023, he was a joint supervision Ph.D. student with the Department of Electronic Systems, Aalborg University, Aalborg, Denmark. He is currently a Lecturer with the School of Computer and Information, Anhui Polytechnic University (AHPU), Anhui, China. His research interests include modulation recognition, deep learning, pattern recognition, and radio signal classification. Dr. An has been serving as a Reviewer of several international journals, such as the IEEE Transactions on industrial informatics, the IEEE Transactions on wireless communications, the IEEE Transactions on cognitive communications and networking, the IEEE Transactions on vehicular technology, the IEEE signal processing letters, the Elsevier signal processing, the Elsevier digital signal processing, and the IEEE communications letters.

Zheng Chu (Member, IEEE) is an Assistant Professor with Department of Electrical and Electronic Engineering (EEE), The University of Nottingham Ningbo China (UNNC). Prior to this, he held positions in University of Surrey, from 2017 to 2023, and Middlesex University, from 2016 to 2017, respectively. He received MSc and Ph.D. degrees from Newcastle University, Newcastle-upon Tyne, UK, in 2012 and 2016, respectively. His current research interests include smart radio environments/smart reflecting surface, 6G communication networks, IoT networks, AI driven future networks, wireless security, wireless powered networks. He has received the Exemplary Reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS in 2022, and the Best Paper Awards of UCOM, IEEE ICCT, and EAI CHINACOM in 2024.

Zhengyu Zhu (Senior Member, IEEE) received the Ph.D. degree in information engineering from Zhengzhou University, Zhengzhou, China, in 2017. From October 2013 to October 2015, he visited the Communication and Intelligent System Laboratory, Korea University, Seoul, South Korea, to conduct a collaborative research as a Visiting Student. He is currently an associate professor with Zhengzhou University. He served as an Associate Editor for theIEEE SENSOR JOURNAL, IEEE SYSTEMS JOURNAL, JOURNAL OF COMMUNICATIONS AND NETWORKS, the PHYSICAL COMMUNICATIONS. His research interests include information theory and signal processing for wireless communications such as B5G/6G, Intelligent reflecting surface, the Internet of Things, machine learning, millimeter wave communication, UAV communication, physical layer security, convex optimization techniques, and energy harvesting communication systems.

Gaojie Chen (Senior Member, IEEE) received the B.Eng. and B.Ec. Degrees in electrical information engineering and international economics and trade from Northwest University, China, in 2006, and the M.Sc. (Hons.) and PhD degrees in electrical and electronic engineering from Loughborough University, Loughborough, U.K., in 2008 and 2012, respectively. After graduation, he took up academic and research positions at DT Mobile, Loughborough University, University of Surrey, University of Oxford and University of Leicester, U.K. He is a Professor and Associate Dean of the School of Flexible Electronics (SoFE), at Sun Yat-sen University, China, and a visiting professor at the 5GIC & 6GIC, University of Surrey, UK. His research interests include wireless communications, flexible electronics, satellite communications, the Internet of Things and secrecy communications. He received the Best Paper Awards from the IEEE IECON 2023, and the Exemplary Reviewer Awards of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2018, the IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and the IEEE COMMUNICATIONS LETTERS in 2020 and 2021; and Exemplary Editor Awards of the IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS NETWORKING in 2021, 2022, 2023 and 2024, respectively. He served as an Associate Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS - Machine Learning in Communications from 2021 to 2022. He serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS NETWORKING, IEEE WIRELESS COMMUNICATIONS LETTERS, and a Senior Editor for the IEEE COMMUNICATIONS LETTERS, and a Panel Member of the Royal Society's International Exchanges, UK.



Yonghui Li (Fellow, IEEE) received his PhD degree in November 2002 from Beijing University of Aeronautics and Astronautics. Since 2003, he has been with the Centre of Excellence in Telecommunications, the University of Sydney, Australia. He is now a Professor and Director of Wireless Engineering Laboratory in School of Electrical and Information Engineering, University of Sydney. He is the recipient of the Australia Research Council (ARC) Queen Elizabeth II Fellowship in 2008, ARC Future Fellowship in 2012 and ARC Industry Laureate Fellowship in 2025. He is listed as a Clarivate highly cited researcher.

He is a Fellow of IEEE. His current research interests are in the area of wireless communications, with a particular focus on MIMO, millimeter wave communications, machine to machine communications, coding techniques and cooperative communications. He holds a number of patents granted and pending in these fields. He was an editor for IEEE transactions on communications and IEEE transactions on vehicular technology. He also served as the guest editor for several IEEE journals, such as IEEE JSAC, IEEE Communications Magazine, IEEE IoT journal, IEEE Access. He received the best paper awards from IEEE International Conference on Communications (ICC) 2014, IEEE PIRMC 2017 and IEEE Wireless Days Conferences (WD) 2014.