Channel Estimation for DMA-based MIMO Communications: Model-based Learning Approach

Xiangyu Zhang, *Student Member, IEEE,* Haiyang Zhang, *Member, IEEE,* Nir Shlezinger, *Member, IEEE,* Luxi Yang, *Senior Member, IEEE,* and Yonina Eldar, *Fellow, IEEE,*

Abstract

Dynamic Metasurface Antennas (DMA) is a cutting-edge antenna technology that allows for the creation of large antenna arrays with low energy consumption and cost. The effectiveness of DMA-based wireless communication is reliant on precise channel knowledge, which can be challenging to obtain due to the complex structure of DMA. In this study, we propose using model-based learning algorithms to overcome the challenging channel estimation problem in DMA-based Multiple Input Multiple Output (MIMO) communications. Our approach starts by casting the channel estimation problem as a compressed sensing issue through the use of a fixed spatial gridding dictionary. We then employ the Learning Iterative Shrinkage and Thresholding Algorithm (LISTA) to recover the sparse channel parameters when the DMA weighting matrix (DWM) is randomly configured. However, the random DWM and the sparse representation hinder the accuracy of LISTA recovery. To overcome this challenge, we introduce a novel deep learning algorithm called LISTA-sensing matrix optimization (LISTA-SMO), which jointly optimizes both DWM and the sparsifying dictionary. The LISTA-SMO transforms the DMA signal processing and sparse representation into trainable layers, whose parameters are defined as the DWM and sparsifying dictionary, respectively. These layers are integrated into the LISTA, allowing for the optimization of both matrices along with the sparse recovery task during the training stage of LISTA. Furthermore, we propose a self-supervised learning technique to tackle the difficulty of acquiring noise-free data. Our numerical results demonstrate that the proposed model-based deep learning methods outperform traditional sparse recovery methods in terms of the normalized mean-square error (NMSE) of the channel estimation error. As anticipated, our LISTA-SMO approach achieves better channel estimation results compared to the standard LISTA.

I. Introduction

To meet the ambitious requirements of future sixth-generation (6G) networks, such as terabits per second peak rate, large-scale/extremely large-scale antenna arrays are regarded as one of the critical physical-layer technologies [1], [2]. However, in reality, constructing an antenna array with hundreds or even thousands of elements presents several challenges, including high fabrication costs, increased power consumption, limited physical size and shape, and deployment restrictions. To address these challenges, an innovative technology known as Dynamic Metasurface Array (DMA) has been proposed for the implementation of large-scale MIMO antenna arrays [3]. Specifically, the DMA is a unique type of metasurface antenna array that consists of substantial subwavelength metamaterial radiators excited via a waveguide or microstrip. DMA allows for the alteration of the metasurface radiator's characteristics, enabling the adjustment of incoming signal amplitudes and phase shifts [4]. As a result, signal processing techniques, like compression and analog combination, are integrated into the physical antenna construction, significantly reducing cost and power consumption. Additionally, DMAs allow for the fabrication of large-scale planar antenna arrays while maintaining a tiny hardware footprint [5]–[7].

In recent years, there has been a growing interest in the use of DMA in massive MIMO communication systems [5], [6], [8], [9]. The literature [5]. Studies have shown that DMA offers several advantages over traditional fully-digital antenna arrays. For instance, the literature [5] presents a mathematical model for massive MIMO systems utilizing DMA as the receive antenna arrays and explores the associated constraints. The authors demonstrate that DMA is easier to achieve the fundamental theoretical rate limits of the massive MIMO system compared to traditional fully-digital antenna arrays. Furthermore, the literature [6] examines a DMA-based MIMO-OFDM system with limited ADC (Analog-to-Digital Converter) and finds that DMA has a higher level of spectral flexibility than conventional hybrid architectures, resulting in significant performance gains. Another study [8] focuses on a near-field large-scale MIMO system implemented using DMA, in which the weight of large-scale DMA is designed to steer the beam towards a near-field area. In [9], the authors consider a hybrid reconfigurable intelligent surface and DMA-assisted MIMO system. They claim that the DMA-assisted system reaches higher spectrum efficiency over conventional MIMO systems. In conclusion, these studies highlight the potential benefits of incorporating DMA in massive MIMO communication systems and suggest that it is a promising avenue for future research.

According to the literature above, the potential benefit of DMA highly relies on the availability of accurate channel state information (CSI). However, in reality, obtaining accurate CSI in DMA-assisted communications is quite challenging due to the DMA's inherent analog compression operation [5], [6]. Precisely, DMA consists of multiple independent microstrips that accommodate a huge number of radiating elements. During uplink channel estimation, pilots sent by users are received by the DMA elements. The pilots received by the same microstrip undergo analog compression by the microstrip and are then fed to a single RF chain, i.e., analog compression. This analog compression requires the channel to be estimated from the compressed signal at the end of the RF chain, which presents an underdetermination problem. Accurately solving this problem within the coherence time is a challenge for channel estimation algorithms.

The aforementioned challenges raised by analog compression are also shown in the hybrid analog-digital antenna array MIMO system [10]. Several works solve the channel estimation problem in the hybrid antenna array by employing compressed sensing (CS) theory [11]–[13]. CS is a successful signal processing technique capable of recovering sparsity signals from a few observation signals [14]. When applying the CS in the channel estimation, the original channel is first transformed into a sparse form, and the CS algorithm estimates the sparse channel instead. For example, the work [11] addresses the wideband millimeter wave (mmWave) MIMO channel estimation problem by representing the channel as a sparse signal in the time-frequency domain and using the orthogonal matching pursuit (OMP) algorithm to estimate the sparse signal. In [12], the mmWave channel is represented as a sparse spatial signal; then, sparse Bayesian learning was proposed to estimate sparse channels. Additionally, paper [13] utilizes the iterative soft thresholding (ISTA) algorithm in the estimation of the spatial covariance of a channel. We note that the above-mentioned conventional sparse recovery approaches, such as ISTA and OMP, require the sensing matrix and suffer from high calculation complexity. To address these issues, recently, model-based learning has been applied to estimate sparse channels of hybrid-architecture-based wireless communication systems [15]–[17]. The model-based learning is one of the deep learning algorithms that incorporate signal processing models into the algorithm. Utilizing the prior knowledge provided by signal processing models, the model-based learning algorithms take the advantages of a lower computational burden, a lower number of parameters, and higher interpretability, flexibility, versatility, and reliability, compared to conventional model-free learning algorithms [18]. Most model-based learning algorithms applied to the channel estimation problem adopt the idea of deep unfolding that converts an iterative CS algorithm into

a deep neural network. For example, the ADMM-net proposed in [19] unfolds an iteration of the alternating direction method of multipliers algorithm into a neural network layer. Besides the unfolded ISTA algorithm, the ISTA-net is built by unfolding the ISTA algorithm [20]. ISTA-net is leveraged to solve the channel estimation problem in wideband millimeter-wave scenarios in [15].

When solving channel estimation in DMA using Compressed Sensing (CS) algorithms, a major obstacle is the construction of the sensing matrix. The sensing matrix must meet the Restricted Isometry Property (RIP) to successfully recover sparse channels from a noisy signal [21]. Currently, a popular approach to meet the RIP constraint is using random matrix construction [22]. However, this method may be challenging to implement in the DMA channel estimation. The sensing matrix in DMA channel estimation is constituted by the DMA weighting matrix (DWM) and the sparsifying dictionary. Due to the nature of the DMA architecture, which only compresses signals received from its own radiating elements, the DWM is inherently sparse, resulting in the sensing matrix that fails to meet the RIP. On the other aspect, not only the DWM but also the conventional sparsifying dictionary may lower the channel estimation accuracy. The conventional methods employ a fixed spatial gridding dictionary, which assumes the channel incident from several discrete grids. This assumption inevitably introduces sparse representation errors and lowers the estimation accuracy [23], [24]. In the CS theory field, it is common sense that a sensing matrix optimally designed for a certain class of signals can further improve the reconstruction accuracy [25]–[27], [27], [28]. Following this idea, literature [29]–[31] addresses this issue by solving the channel estimation problem with a learned sparsifying dictionary from channel measurements rather than relying on a fixed over-complete Fourier matrix. Despite this progress, there is still a lack of research investigating the joint optimization of the weighting matrix and the sparsifying dictionary in channel estimation.

In this work, we delve into the challenge of channel estimation for multi-user MIMO communications that utilize DMA technology. We leverage the spatial sparse feature of the channel and represent it in sparse form by spatial gridding dictionary. The channel estimation is then formulated as a CS problem where the goal is to estimate the sparse channel from a limited number of observation pilot signals using a randomizing constructed DWM. After that, we exploit the model-based learning approach, the Learning Iterative Shrinkage and Threshold Algorithm (LISTA), to estimate channels. In response to the challenges in the sensing matrix construction, we optimize the sensing matrix by proposing LISTA-sensing matrix optimization (SMO). LISTA-

SMO is a model-based algorithm whose neural network is constituted by LISTA and the sensing matrix optimized layer that is designed by unfolding the signal processing process into neural network layers. The new neural networks optimize the DWM and the sparsifying dictionary during LISTA training. Moreover, we introduce a self-supervised learning (SSL) approach to train the LISTA-SMO network because LISTA-SMO can be treated as a denoiser. SSL doesn't require the noiseless channels as labels, significantly reducing the training complexity and making it easier to implement in real systems.

The main contributions of this paper can be summarized as follows

- Firstly, we study the channel estimation problem for emerging DMA-based wireless communications. Using a sparsifying dictionary generated by spatial gridding, we first represent the channel as a sparse representation. Then we formulate the resulting channel estimation problem as a compressed sensing (CS) problem and leverage the Learning Iterative Shrinkage and Threshold Algorithm (LISTA) to solve it under the random DMA weighting matrix (DWM). In our algorithm, the architecture of our neural network is constructed following the CS algorithm, which greatly improves efficiency and interpretability.
- Secondly, we detail the low-accuracy reason raised by the sensing matrix construction in solving the DMA channel estimation with LISTA. Precisely, the sensing matrix in DMA channel estimation is constituted by the DWM and the sparsifying dictionary. Due to the high sparse characteristic of DWM, the randomized DWM induces LISTA has a low probability of recovering the channel from the observation. Meanwhile, the spatial gridding sparsifying dictionary introduces the sparse representation error into the LISTA and further lowers the estimation accuracy. In response, we formulate an optimization problem to optimize the sensing matrix.
- To solve the sensing matrix optimization problem, we propose a model-based learning algorithm, namely LISTA-sensing matrix optimization (LISTA-SMO). LISTA-SMO first builds the sensing matrix optimization layers by unfolding the DMA signal processing process and sparse representation process, in which the optimizing matrices are formulated as learnable parameters. After that, LISTA-SMO embeds the sensing matrix optimization layers into the LISTA and optimizes the sensing matrix at the LISTA training stage.
- The architecture of LISTA-SMO shows that LISTA-SMO is a denoiser with ℓ_1 regularization item. Then, we propose a self-supervised learning method for LISTA-SMO. The proposed self-supervised learning does not need noiseless channels as the label and reaches a similar

level of performance to supervised learning.

Besides, we provide rich numerical results to show the superiority of our algorithms. Particular, numerical results shows that the channel estimation accuracy of our proposed model-based learning methods is significantly greater than the conventional non-learning methods. Also, thanks to the jointly training of sensing matrix, LISTA-SMO outperforms LISTA in terms of the NMSE of channel estimation error.

The remaining part of this paper is organized as follows. Section II introduces the architecture of DMA and the channel model. Also, channel estimation is formulated as a CS problem. Then, in Section III, we introduce the model-based method, LISTA. Our proposed algorithm, LISTA-SMO, for optimizing the DWM and sparsifying dictionary, is presented in Section IV. In this Section, self-supervised learning is presented. The numerical results are presented in Section V, and Section VI concludes the whole paper.

Notations— The bold type $\mathbf{a}$, $\mathbf{A}$ denote the vector and matrix, respectively. $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$ denote the transpose and the conjugate transpose of $\mathbf{A}$. $\otimes$ represents the Kronick product. Besides, $\|\cdot\|_0, \|\cdot\|_1, \|\cdot\|_2$ represent the ℓ_0-norm, ℓ_1-norm and ℓ_2-norm, respectively. $\mathbb{E}\{\cdot\}$ represents the expectation operator. $\mathbf{A} = \mathrm{diag}(a_i)$ denotes the matrix $\mathbf{A}$ is a diagonal matrix, whose diagonal elements is a_i.

II. System Model and Problem Formulation

In this section, we present the DMA-based multi-user MIMO communication system and formulate the channel estimation problem. As shown in Fig. 1, we consider a multi-user MIMO system consisting of a DMA-based BS and K single-antenna users. We study the uplink channel estimation scenario, where users transmit orthogonal pilots to the BS. The pilot signals output of DMA will be used to estimate the wireless channels. In the following, we will first briefly introduce the Dynamic Metasurface Antennas in Subsection II-A. Then, the channel model will be presented in Subsection II-B. After that, we formulate the channel estimation problem as a CS problem in Subsection II-C.

A. Dynamic Metasurface Antennas

A typical DMA consists of N_d microstrips, each with N_e metamaterial radiations, for a total of $N = N_d \times N_e$ elements. DMA receives signals via the metamaterial radiating elements and modifies their amplitude and/or phase by adjusting the characteristic parameter of the elements.

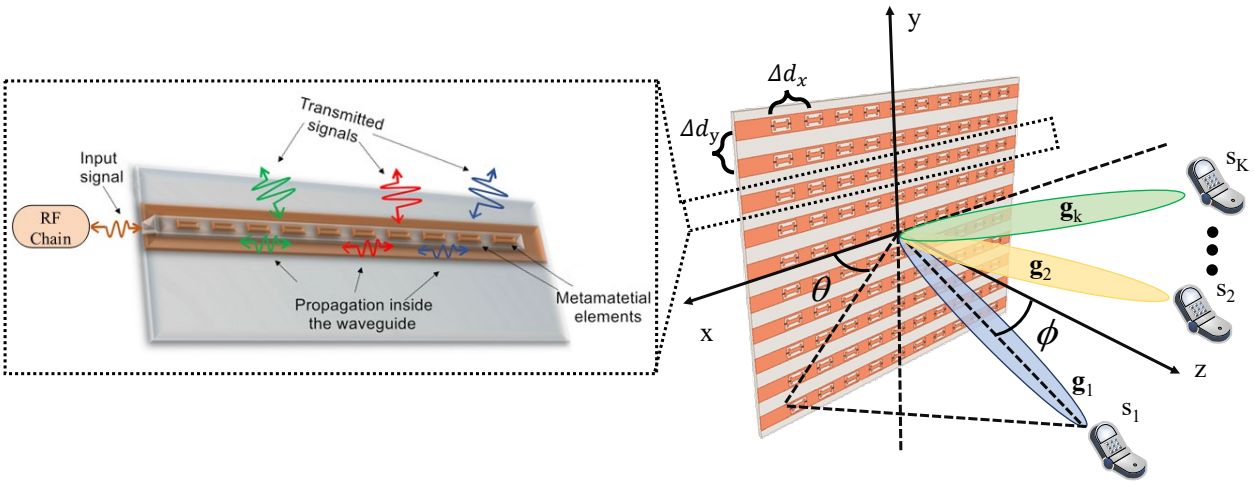


Fig. 1: The system model of multi-user channel estimation.

The signals will then flow through the microstrip and be gathered by the digital processing component. The signal output of DMA, denoted by $\mathbf{z}_{\mathrm{RF}} \in \mathbb{C}^{N_d \times 1}$, is given by [3], [5], [8]

$$\mathbf{z}_{\mathrm{RF}} = \mathbf{QHy}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{N \times 1}$ is the signal observed by N DMA elements. $\mathbf{H} \in \mathbb{C}^{N \times N}$ is the frequency weight that encapsulates the effect of signal propagation inside the microstrip. $\mathbf{H}$ is a diagonal matrix, with diagonal element $\mathbf{H}_{(n_d-1)*N_e+n_e,(n_d-1)*N_e+n_e} = h_{n_d,n_e} = e^{-\rho_{n_d,n_e}(\alpha_{n_e}+j\beta_{n_e})}$, which represents the propagation delay of n_e-th element in n_d-th microstrip. α_n is the waveguide attenuation coefficient, β_n is the wavenumber, and $\rho_{n,l}$ denotes the location of this element. $\mathbf{Q} \in \mathbb{C}^{N_d \times N}$ represents the configurable weight of each metamaterial element, which can be formulated as

$$\mathbf{Q}_{n_1,(n_2-1)N_d+l} = \begin{cases} q_{n_1,l} & n_1 = n_2 \\ 0 & n_1 \neq n_2 \end{cases}, \tag{2}$$

where $q_{n,l}$ follows the Lorentzian-constrained phase [32], given by

$$q_{n,l} \in \mathbb{Q} = \left\{ \frac{j + e^{j\varphi}}{2} \mid \varphi \in [0, 2\pi] \right\}. \tag{3}$$

As shown in Fig. 1, we set the DMA in XoY plane, where microstrips are settled parallelly with the x-axis. Δd_x and Δd_y represent the space between the microstrip and the elements. θ and ϕ denote the azimuth angle and elevation angle.

B. Channel Model

As in [33]–[35], we adopt the commonly used Saleh-Valenzuela channel model (also known as the multipath channel model) to characterize wireless channels. This model represents electromagnetic wave propagation by several propagation paths. Mathematically, the channel between the k-th user and DMA can be represented by

$$\mathbf{g}_k = \sqrt{\frac{N}{L_p}}\mathbf{A}_P\boldsymbol{\alpha}_P = \sqrt{\frac{N}{L_p}}\sum_{l_p=1}^{L_p} a_{l_p}\mathbf{a}\left(\theta_{l_p},\phi_{l_p}\right), \tag{4}$$

where L_p is the number of paths, following $L_p < \min(N_d, N_e)$. $\boldsymbol{\alpha}_P = \left[\cdots, a_{l_p}, \cdots\right]^{\mathrm{T}}$ is the channel gain vector, in which a_{l_p} is the channel gain of l_p-th path. $\mathbf{A}_P = \left[\cdots, \mathbf{a}\left(\theta_{l_p},\phi_{l_p}\right), \cdots\right]$ is the channel path matrix, in which $\mathbf{a}\left(\theta_{l_p},\phi_{l_p}\right) = \mathbf{a}_x\left(\theta_{l_p},\phi_{l_p}\right)\otimes\mathbf{a}_y\left(\theta_{l_p},\phi_{l_p}\right)$ represents the array response vector of l_p-th path, where $\mathbf{a}_x$ and $\mathbf{a}_y$ can be formulated as

$$\mathbf{a}_x\left(\theta_{l_p},\phi_{l_p}\right) = \frac{1}{\sqrt{N_e}}\left[1, e^{jd_x}, \ldots, e^{jd_x(N_e-1)}\right]^{\mathrm{T}}, \tag{5}$$

$$\mathbf{a}_y\left(\theta_{l_p},\phi_{l_p}\right) = \frac{1}{\sqrt{N_d}}\left[1, e^{jd_l}, \ldots, e^{jd_l(N_d-1)}\right]^{\mathrm{T}}, \tag{6}$$

where $d_x = 2\pi\frac{\Delta d_x}{\lambda}\cos\theta\sin\phi$ and $d_y = 2\pi\frac{\Delta d_y}{\lambda}\sin\theta\sin\phi$ represents the propagation delay caused by the position difference of two elements and two microstrips, respectively.

C. Problem Formulation

When performing channel estimation, the BS receives a sequence of pilot $\mathbf{Z}_{\mathrm{RF}} \in \mathbb{C}^{N_d\times T}$ from K users, given by

$$\mathbf{Z}_{\mathrm{RF}} = \mathbf{QHY} = \mathbf{QH}\left(\mathbf{GS} + \mathbf{N}_s\right), \tag{7}$$

where $\mathbf{Y} \in \mathbb{C}^{N\times T}$ is the DMA received signal and $\mathbf{S} \in \mathbb{C}^{K\times T}$ is the orthogonal pilot sequences of K users, which follow $\mathbf{s}_u\mathbf{s}_u^{\mathrm{H}} = 1$, $\mathbf{s}_u\mathbf{s}_v^{\mathrm{H}} = 0$, where $\mathbf{s}_u$ and $\mathbf{s}_v$ are u and v-th row of $\mathbf{S}$. $\mathbf{N}_s \in \mathbb{C}^{N_d\times T}$ represents the received additive noise. As the pilot is known at the BS side, the signal for estimating the channel of k-th user can be obtained by

$$\mathbf{z}_n = \mathbf{Z}_{\mathrm{RF}}\mathbf{s}^{\mathrm{H}} = \mathbf{QHg}_k + \mathbf{QHn}_s\mathbf{s}^{\mathrm{H}} = \mathbf{QHg}_k + \mathbf{n}_q, \tag{8}$$

where $\mathbf{g}_k$ is the channel of k-th single user, $\mathbf{n}_q$ is equivalent noise. We assume that the elements of $\mathbf{n}_q$ follows normal distribution $\mathcal{N}(0,\sigma^2)$. For easier expression, we drop the subscription of $\mathbf{g}_k$ and utilize $\mathbf{g}$ to represent $\mathbf{g}_k$ in the following.

Our main objective is to estimate $\mathbf{g}_k$ from $\mathbf{z}_n$, which is an underdetermined problem because of $N_d \ll N$. Fortunately, the channel can be represented by the parameter set $[\alpha_{l_p}, \theta_{l_p}, \phi_{l_p}], l_p \in [1, ..., L_p]$ as shown in (4). As $L_p \ll N_d \ll N$, it is more reasonable to estimate the parameter set $[\alpha_{l_p}, \theta_{l_p}, \phi_{l_p}]$ rather than $\mathbf{g}$.

To achieve this, we transform the channel into a sparse form with spatial gridding. The spatial gridding first divides the interesting areas by D_a and D_p azimuth angle and elevation angle, namely D grids, and then constructs a spatial gridding dictionary $\mathbf{A} \in \mathbb{C}^{N \times D}$ with the array response vectors on these grids, which can be formulated as

$$\mathbf{A} = \left[\mathbf{a}\left(\theta_1, \phi_1\right), \quad \cdots, \quad \mathbf{a}\left(\theta_D, \phi_D\right) \right], \tag{9}$$

where θ_i and ϕ_i are the azimuth and elevation angle of grid. Then, by assuming the channel path is incident from these grids, the channel can be represented in a sparse form, given by $\mathbf{g} \simeq \mathbf{A}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_D]^{\mathrm{T}} \in \mathbb{C}^{D \times 1}$ denotes the channel gain in the corresponding angle. Due to the limited number of transmitting paths in the practical environment, i.e., $\|\boldsymbol{\alpha}\|_0 \ll D$, which denotes the $\boldsymbol{\alpha}$ are highly sparse. Then, (8) can be rewritten as

$$\mathbf{z}_n \simeq \mathbf{QHA}\boldsymbol{\alpha} + \mathbf{n}_q. \tag{10}$$

The estimation of the sparse signal $\boldsymbol{\alpha}$ from $\mathbf{z}$ in (10) is a typical CS problem, which can be formulated as

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \|\mathbf{z} - \mathbf{QHA}\boldsymbol{\alpha}\|_2 \\ \text{s.t} \quad & \|\boldsymbol{\alpha}\|_0 < \epsilon, \end{aligned} \tag{11}$$

where ϵ is the sparse constraint constant. Since the constrain condition $\|\boldsymbol{\alpha}\|_0$ is not convex, finding a true minimum solution to Problem (11) may be quite challenging. In the next section, we will translate this problem into a more tractable problem and solve it by ISTA and LISTA.

III. DMA Channel Estimation via Model-Based Learning

In this section, we estimate the sparse channel gain $\boldsymbol{\alpha}$ from the received pilot $\mathbf{z}$ by solving the Problem (11) with the model-based learning method, LISTA.

Due to the non-convex constrain condition $\|\cdot\|_0$, directly solving the Problem (11) is challenging. By relaxing $\|\cdot\|_0$ to $\|\cdot\|_1$, we have a more tractable problem, given by:

$$\min_{\boldsymbol{\alpha}} \quad \|\mathbf{z} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|_2 + \xi\|\boldsymbol{\alpha}\|_1, \tag{12}$$

where $\mathbf{\Psi} = \mathbf{P}^{-1}\mathbf{QHA}$ and ξ is a regularization parameter. It has been proved that (11) and (12) share the same solution when the sensing matrix $\mathbf{\Psi}$ satisfies the RIP of order 2k with $\delta_{2k} < \sqrt{2} - 1$ [14].

One way to solve (12) is to apply the Iterative Shrinkage and Thresholding Algorithm (ISTA) [36], [37]. Standard ISTA starts from an initial solution $\boldsymbol{\alpha}^{(0)}$ and iteratively performs proximal gradient descent with respect to the objective function in (12). Specifically, the iteration can be formulated as

$$\boldsymbol{\alpha}^{(t+1)} = h_{(\eta)}\left((I - \frac{1}{\lambda_{\max}}\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Psi})\boldsymbol{\alpha}^{(t)} + \frac{1}{\lambda_{\max}}\mathbf{\Psi}^{\mathrm{H}}\mathbf{z}\right), \tag{13}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi}$; $h_{(\eta)}$ is the soft shrinkage function with threshold η, given by $[h_{(\eta)}]_g = \mathrm{sign}([\boldsymbol{\alpha}]_g)(|[\boldsymbol{\alpha}]_g| - \eta)_+$, where $[\cdot]_n$ is the n-th elements of vector; $\mathrm{sign}(\cdot)$ returns the sign of a scalar; $(\cdot)_+$ means $\max(\cdot, 0)$; η is a constant threshold.

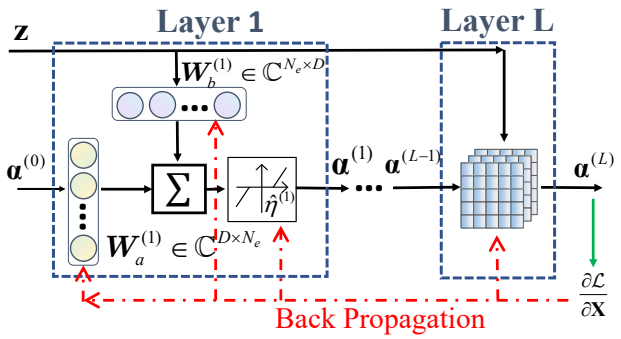


Fig. 2: The block diagram of LISTA algorithm

The recovering accuracy of ISTA heavily depends on the parameter of η and the iterative time. Selecting the proper parameter for each $\boldsymbol{\alpha}$ is challenging and ISTA can also be computationally intensive, requiring hundreds or thousands of iterations to converge. In response to these limitations, authors in [38] proposed a model-based learning method, Learning ISTA (LISTA). LISTA unfolds the iteration of ISTA as a layer of a neural network by formulating three learnable parameters $\mathbf{W}_a \in \mathbb{C}^{D \times D}$, $\mathbf{W}_b \in \mathbb{C}^{N_d \times D}$ and $\hat{\eta}$ to learn $(I - \frac{1}{L_e}\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Psi})$, $\frac{1}{L_e}\mathbf{\Psi}^{\mathrm{H}}$ and η in each iteration. Mathematically, it is given by

$$\boldsymbol{\alpha}^{(l+1)} = h_{(\hat{\eta}^{(l)})}\left(\mathbf{W}_a^{(l)}\boldsymbol{\alpha}^{(l)} + \mathbf{W}_b^{(l)}\mathbf{z}\right), \tag{14}$$

where $\boldsymbol{\alpha}^{(l)}$ is the hidden layer feature, and $\mathbf{z}$ is the weighted biased items added to each layer. By stacking L layers, we have the LISTA neural networks $\mathcal{N}(\mathbf{z}|\mathbf{W}_a, \mathbf{W}_b, \hat{\eta})$, which resembles a

recurrent neural network (RNN), a L-layer feed-forward neural network with side connections. We illustrate the neural network in Fig. 2. In Fig. 2, we represent the forward estimation process with the dark line and the backward update process with the red line. The forward process feeds the received pilot $\mathbf{z}$ to the neural network and outputs the estimated sparse channel $\boldsymbol{\alpha}^{(L)}$. Then, the DMA channel $\mathbf{g}$ can be obtained from $\mathbf{g} = \mathbf{A}\boldsymbol{\alpha}^{(L)}$ accordingly. The backward process updates the parameters of LISTA by minimizing the loss function

$$\mathcal{L} = \sum_{s=1}^{S} \|\boldsymbol{\alpha}_s^* - \boldsymbol{\alpha}_s^{(L)}\|_2, \tag{15}$$

on the training data set $\mathbb{D} = \{[\mathbf{z}_s, \boldsymbol{\alpha}_s^*] | \mathbf{z}_s \simeq \boldsymbol{\Psi}\boldsymbol{\alpha}_s^* + \mathbf{n}, s = 1, ..., S\}$, where $\boldsymbol{\alpha}_s^{(L)}$ is the output corresponding to $\mathbf{z}_s$; $\mathbf{z}_s$ and $\boldsymbol{\alpha}_s^*$ are the sample pair representing the received pilot and corresponding actual channel.

In contrast to ISTA, which utilizes an experiential and identical parameter in each iteration, LISTA fixes the number of layers and learns parameter set $\{\mathbf{W}_a, \mathbf{W}_b, \hat{\eta}\}$ from data. The learned parameter will fit the data and ensure recovery performance. Empirical evidence has shown that a trained LISTA can generalize better to unseen samples that follow the same distribution and recover $\boldsymbol{\alpha}$ from $\mathbf{z}$ within two orders of magnitude fewer iterations than the original ISTA.

Algorithm 1 LISTA in DMA channel estimation

Training Process

Initialize: Initialize the parameter $\mathbf{W}_a$, $\mathbf{W}_b$ and $\hat{\eta}$.

Input: the data set $\mathbb{D} = [\{\mathbf{z}_s, \boldsymbol{\alpha}_s^*\} | \mathbf{z}_s = \boldsymbol{\Psi}\boldsymbol{\alpha}_s^* + \mathbf{n}]\,(s = 1, ..., S)$.

While

    Sample a batch of data from the data set $\mathbb{D}$;

    Calculate the output of neural network $\boldsymbol{\alpha}^{(l)}$ and calculate the loss $\mathcal{L}$;

    Update the parameter $\mathbf{W}_a$, $\mathbf{W}_b$ and $\hat{\eta}$.

Until $\mathcal{L}$ converge.

Output: The well-trained neural network $\mathcal{N}$.

Inference Process

Feed the received pilot $\mathbf{z}$ into the well-trained neural network $\mathcal{N}$ and obtain the estimated sparse channel $\boldsymbol{\alpha}^{(L)}$. Then, obtain the DMA channel $\mathbf{g} = \mathbf{A}\boldsymbol{\alpha}^{(L)}$ accordingly.

IV. Sensing Matrix Optimization in DMA channel estimation

In the previous section, we address the Problem (12) with LISTA. However, the channel estimation accuracy of LISTA is limited by the random $\mathbf{Q}$ and the spatial gridding sparsifying dictionary $\mathbf{A}$. With the aim of improving the channel estimation accuracy, in this section, we propose LISTA-SMO to optimize the DMA weighting matrix $\mathbf{Q}$ and the sparsifying dictionary $\mathbf{A}$.

Specifically, in Subsection IV-A, we first detail the channel estimation accuracy loss raised by the random $\mathbf{Q}$ and the spatial gridding sparsifying dictionary $\mathbf{A}$, and then formulate the optimization problem to optimize $\mathbf{Q}$ and $\mathbf{A}$. In Subsection IV-B, we propose our model-based learning algorithms, the LISTA-sensing matrix optimization(SMO) that jointly optimizes $\mathbf{Q}$ and $\mathbf{A}$. Besides, considering the difficulty in obtaining the noise-less data, we propose the self-supervised learning method for our proposed methods in Subsection IV-C. Finally, we analyze the computational complexity in Subsection IV-D and conclude with a discussion in Subsection IV-E. For easy expression, we utilize the description, optimizing $\mathbf{Q}$, to represent optimizing the φ on each element of DMA.

A. Problem Formulation for Optimizing Sensing Matrix

As described in [21], the sensing matrix significantly impacts the CS algorithms' recovery performance. In DMA channel estimation, the sensing matrix that consists of the random $\mathbf{Q}$ and the spatial gridding sparsifying dictionary $\mathbf{A}$ may not support LISTA to fully recover the channel. In the following, we first theoretically proved that the random $\mathbf{Q}$ and the spatial gridding sparsifying dictionary $\mathbf{A}$ decrease the channel estimation accuracy. Then, we formulate the optimization problem to optimize these two matrices.

We utilize Theorem 1 to prove the accuracy loss raised by the random weighting matrix $\mathbf{Q}$.

Theorem 1: When Problem (12) adopts a uniform random DMA weighting matrix, i.e., $\varphi \sim \mathcal{U}[0, 2\pi]$, the CS algorithm can recover the channel from the received pilot with the probability

$$p = \frac{\mathcal{X}_{2L_p}}{C_N^{2L_p}} (\frac{\sqrt{2}}{2})^{2L_p}, \tag{16}$$

where $\mathcal{X}_{2L_p} = C_{N_d}^{2L_p}(C_{(2L_p)N_e}^{2L_p} - \sum_{i=1}^{2L_p-1} C_{2L_p}^{i}\mathcal{X}_i)$; $\mathcal{X}_1 = C_{N_e}^{2L_p}$; L_p is the maximum path defined in (4).

Proof: Please refer to Appendix A. ■

Theorem 1 demonstrates that using a random DMA weighting matrix in the sensing matrix results in a low probability of successfully recovering the sparse channel. For example, for a scenario in which the array is $N_d = 20$ and $N_e = 20$ and the incident path is $L = 6$, the probability of totally recover the channel is 0.002. This leads us to seek a more appropriate DMA weighting matrix.

Additionally, the loss of accuracy can also be attributed to sparse representation error. The sparse channel is assumed to result from pilot signals that are discretized onto grids, as expressed in (9). This discretization leads to inevitable sparse representation error, resulting in a significant decrease in estimation accuracy.

Theorem 2: When the incident angle follows a uniform distribution, the expectation of estimation accuracy loss is $\varrho_{\text{mean}} = 2 - 2\exp\left(1 - \cos\frac{\Delta\theta}{4}\cos\frac{\Delta\phi}{4}\right)$, where $\Delta\theta$ and $\Delta\phi$ are grid intervals in the azimuth angle and elevation angle dimension.

Proof: Please refer to Appendix B. ■

From Theorems 1 and 2, we can conclude that it is essential to optimize $\mathbf{Q}$ and $\mathbf{A}$ to achieve higher channel estimation accuracy before solving the problem (12). Therefore, we represent the input and output relationship of CS algorithms as $\boldsymbol{\alpha} = \mathcal{F}(\mathbf{z}|\mathbf{Q}, \mathbf{A})$ and propose the matrix optimization problem as

$$\begin{aligned} \min_{\mathbf{Q},\mathbf{A}} \ & \|\mathbf{g}^* - \mathbf{A}\mathcal{F}(\mathbf{z}|\mathbf{Q}, \mathbf{A})\|_2 \\ \text{s.t } & \mathbf{z} = \mathbf{Q}\mathbf{H}\mathbf{g}^* + \mathbf{n}_q, \\ & q_{n,l} \in \mathbb{Q} = \left\{ \frac{j + e^{j\varphi}}{2} \mid \varphi \in [0, 2\pi] \right\}, \end{aligned} \tag{17}$$

where $\mathbf{g}^*$ is the actual channel.

Solving this problem is challenging. The reason has two aspects. Firstly, the formulation of $\mathcal{F}(\mathbf{z}|\mathbf{Q}, \mathbf{A})$ may be excessively complex. For example, in the case of ISTA, which is an iterative algorithm, $\mathcal{F}(\mathbf{z}|\mathbf{Q}, \mathbf{A})$ is a higher-order composite function composed of soft shrink functions, making it difficult to express due to its high number of iterations. Secondly, the solving process is computationally intensive. For example, in the case of LISTA, the function $\mathcal{F}(\mathbf{z}|\mathbf{Q}, \mathbf{A})$ can be replaced by the neural network $\mathcal{N}(\mathbf{z}|\mathbf{W}_a, \mathbf{W}_b, \hat{\eta})$. However, retraining the LISTA neural network is required every time $\mathbf{Q}$ or $\mathbf{A}$ is updated, resulting in a lengthy solving process that cannot be completed within a single coherent channel slot.

To respond to the above challenges, we borrow the idea of deep unfolding and propose a new model-based algorithm named LISTA-sensing matrix optimization (SMO). LISTA-SMO

builds the learnable matrices representing $\mathbf{Q}$ and $\mathbf{A}$ and embeds them into LISTA. Then the two matrices can be optimized during the LISTA training. Namely, LISTA-SMO aims to optimize the two matrices under the data set of LISTA $\mathbb{D}$, which can be formulated as

$$\begin{aligned} \min_{\mathbf{Q},\mathbf{A}} \ & \sum_{s=1}^{S} \|\mathbf{g}_s^* - \mathbf{A}\mathcal{F}(\mathbf{z}_s|\mathbf{Q},\mathbf{A})\|_2 \\ \text{s.t } & \mathbf{z}_s = \mathbf{Q}\mathbf{H}\mathbf{g}_s^* + \mathbf{n}, \\ & q_{n,l} \in \mathbb{Q} = \left\{ \frac{j + e^{j\varphi}}{2} \mid \varphi \in [0, 2\pi] \right\}, \end{aligned} \tag{18}$$

where $\mathbf{z}_s$ and $\mathbf{g}_s^*$ are the samples pairs from the data set $\mathbb{D}$.

LISTA-SMO solves Problem (18) to obtain a statistically optimized solution for channels in the LISTA dataset, in contrast to solving Problem (17) for an instantaneous channel. This reduces online computation significantly and allows for concurrent optimization and training processes by treating the optimization matrices as the learning parameter. As a result, LISTA-SMO simplifies the solving process and overcomes challenges faced by existing algorithms.

However, it's important to assume that the incident angle of the channel paths follows a specific distribution, commonly used in wireless communication [39], [40], in order to ensure the validity of the statistically optimized solution. Additionally, the LISTA data samples must be unbiasedly taken from this distribution.

B. LISTA with Sensing Matrix Optimization

This subsection presents the simultaneous solution of (12) and (18) using our proposed LISTA-SMO. LISTA-SMO is a model-based learning algorithm that combines LISTA with two matrix optimization layers: the DMA layer and the sparse representation layer. The DMA layer unfolds the DMA signal processing process into a neural network layer, where $\mathbf{Q}$ serves as the learning parameter. Similarly, we build the sparse representation layer by unfolding the sparse representation process, using $\mathbf{A}$ as the learnable parameter. During the training phase, the LISTA-SMO network updates both the LISTA parameter and the learnable parameters of $\mathbf{Q}$ and $\mathbf{A}$. After training, the well-trained network can recover the channel from the received pilot. Therefore, the training process corresponds to the solving process of (18), while the interference process corresponds to the solving process of (12). In the following sections, we will first introduce the DMA layer and the sparse representation layer, and then propose LISTA-SMO by embedding the two layers into LISTA.

1) DMA layer: To unfold the signal processing process of DMA, we construct the learning parameters $\mathbf{W}_{\varphi} \in \mathbb{R}^{N\times 1}$ to represent $\boldsymbol{\varphi}$, which is the phase of each antenna element in (2). Then, we construct the learning parameter $\mathbf{W_Q} \in \mathbb{C}^{N_d\times N}$, which is

$$\mathbf{W_Q}_{n_1,(n_2-1)N_d+l} = \begin{cases} \frac{j+e^{j2\pi\mathcal{S}(\mathbf{W}_{\varphi(n_2-1)N_d+l})}}{2} & n_1 = n_2 \\ 0 & n_1 \neq n_2 \end{cases}, \tag{19}$$

where $\mathcal{S}(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function for constraining the definition domain of $\boldsymbol{\varphi}$, as shown in (3). The DMA layer is built as

$$\hat{\mathbf{z}} = \mathbf{W_Q}\mathbf{H}\mathbf{y}, \tag{20}$$

where we utilize $\hat{\mathbf{z}}$ to represent the received pilot simulated DMA layer.

2) Sparsifying Representation Layer: To unfold the sparse representation process, we construct a learnable sparsifying dictionary $\mathbf{W_A} \in \mathbb{C}^{N\times D}$ that can transform the channel $\mathbf{g}$ into sparse domain $\boldsymbol{\alpha}$ and $(\mathbf{W_A})^{\mathrm{H}}$ can recover the channel $\mathbf{g}$ from $\boldsymbol{\alpha}$, which can be formulated as

$$\begin{aligned} \mathbf{g} &= \mathbf{W}_{\mathbf{A}}^{\mathrm{H}}\boldsymbol{\alpha}, \\ \boldsymbol{\alpha} &= \mathbf{W_A}\mathbf{g}. \end{aligned} \tag{21}$$

Then, we can build the sparse representation layer as

$$\mathbf{g} = \mathbf{W}_{\mathbf{A}}^{\mathrm{H}} \cdot h_{(\hat{\eta}^{(l)})}(\mathbf{W_A}\mathbf{g}). \tag{22}$$

3) LISTA-sensing matrix optimization: To embed $\mathbf{Q}$ and $\mathbf{A}$ into LISTA, we adopt a new form of LISTA, named Analytic Weight LISTA (ALISTA) [41], [42], which explicitly utilize $\mathbf{Q}$ and $\mathbf{A}$ in each layer to reduce the parameter of LISTA. The layer of ALISTA is given by

$$\boldsymbol{\alpha}^{(l+1)} = h_{(\hat{\eta}^{(l)})}\left(\boldsymbol{\alpha}^{(l)} - \kappa^{(l)}\mathbf{W}(\mathbf{QHA}\boldsymbol{\alpha}^{(l)} - \mathbf{z})\right), \tag{23}$$

where $\mathbf{W} \in \mathbb{C}^{D\times N_d}$, $\hat{\eta}$ and κ are learnable parameters and $\mathbf{W}$ is same for all layer.

Then, by integrating the DMA layer (20) and the sparse represent layer (22) into the (23), we have the LISTA-sensing matrix optimization (SMO), whose neural network layer can be naturally formulated as

$$\mathbf{g}^{(l+1)} = (\mathbf{W_A})^{\mathrm{H}} \cdot h_{(\hat{\eta}^{(l)})}\left(\mathbf{W_A}\left(\mathbf{g}^{(l)} - \kappa^{(l)}\mathbf{W}(\mathbf{W_Q}\mathbf{H}\mathbf{g}^{(l)} - \hat{\mathbf{z}})\right)\right). \tag{24}$$

To better illustrate (24), we show the block diagram of LISTA-SMO in Fig. 3. As shown in Fig. 3, we embed the DMA layer into LISTA by adding a DMA layer before the LISTA neural network, which replaces $\mathbf{z}$ with $\hat{\mathbf{z}}$. Besides, we replace $\mathbf{Q}$ in the (23) with DMA layer

$\mathbf{W_Q}$, which adds the DMA layer into each layer of LISTA. To embed the sparse representation layer into LISTA (23), we add $\mathbf{W_A}$ and $(\mathbf{W_A})^{\mathrm{H}}$ before and after the soft shrinking function and replace $\boldsymbol{\alpha}^{(l)}$ to $\mathbf{g}^{(l)}$.

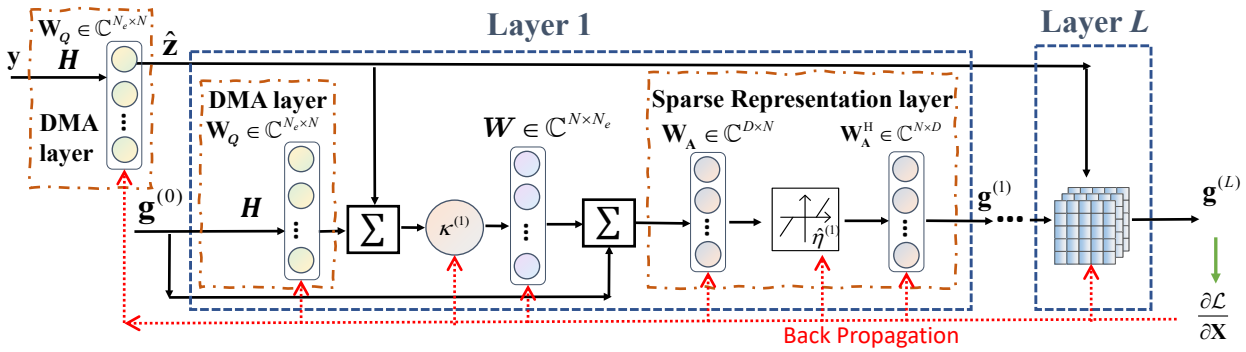


Fig. 3: The block diagram of LISTA-SMO

Similar to LISTA, the LISTA-SMO also includes the forward process and the backward process, which are represented with the dark line and red dashed line, respectively. The forward process feeds $\mathbf{y}$ to the neural network and directly outputs the estimated channel $\mathbf{g}^{(L)}$. The backward process updates the parameters by minimizing the loss function

$$\mathcal{L} = \sum_{s=1}^{S} \|\mathbf{g}_s^* - \mathbf{g}_s^{(L)}\|_2, \tag{25}$$

on the training data set is $\mathbb{D}_{\text{SMO}} = \{[\mathbf{y}_s, \mathbf{g}_s^*] | \mathbf{y}_s = \mathbf{g}_s^* + \mathbf{n}, s = 1, ..., S\}$, where $\mathbf{g}_s^{(L)}$ is the output corresponding to $\mathbf{y}_s$. $\mathbf{y}_s$ and $\mathbf{g}_s^*$ are the sample pair of $\mathbb{D}_{\text{SMO}}$.

During the training stage, LISTA-SMO will update the DMA weighting matrix and the sparsifying dictionary, as well as the LISTA parameter. After that, we can configure the optimized $\mathbf{W_Q}$ on the DMA and utilize the neural network without the first DMA layer to estimate the channel. We conclude LISTA-SMO in Algorithm 2.

Comment 1: The DMA layer optimizes $\mathbf{Q}$ to enhance the signal that incidents from the angle with high incident probability. Consequently, it will lower channel estimation error in the real world, although $\mathbf{Q}$ does not enable the sensing matrix to meet RIP.
Comment 1 illustrates the feasibility of optimizing the DWM.

Comment 2: The sparse representation layer is capable of achieving higher channel estimation accuracy than LISTA because the loss function for LISTA-SDO is directly calculated by the channel, which avoids introducing the sparse representation error at the beginning.

Algorithm 2 LISTA-SMO in DMA channel estimation

Training Process

Initialize: Initialize the parameter $\mathbf{W}$, $\mathbf{W_Q}$, $\mathbf{W_A}$, κ and $\hat{\eta}$ with random number.

Input: The data set $\mathbb{D}_{\text{SMO}} = \{[\mathbf{y}_s, \mathbf{g}_s^*] | \mathbf{y}_s = \mathbf{g}_s^* + \mathbf{n}, s = 1, ..., S\}$.

While

  Sample a batch of data from the data set $\mathbb{D}_{\text{SMO}}$;

  Calculate the output of neural network $\boldsymbol{\alpha}^{(l)}$ and calculate the loss $\mathcal{L}$ in (25);

  Update the parameter $\mathbf{W}$, $\mathbf{W_Q}$, $\mathbf{W_A}$, κ and $\hat{\eta}$.

Until $\mathcal{L}$ converge.

Inference Process

Configure the optimized $\mathbf{W}_\varphi$ on the DMA and feed the receive pilot $\mathbf{z}$ in to the well-trained neural network $\mathcal{N}$ to obtain the channel $\mathbf{g}$.

C. Self-supervised Learning for LISTA-SMO

Obtaining noiseless signal $\boldsymbol{\alpha}$ and $\mathbf{g}$ from the real wireless communication system will be complex, which is an obstacle in implementing our algorithm in the physical world. To solve this, we propose our Self-supervised Learning (SSL) train algorithm. As (25), the loss function indicates that the LISTA-SMO is a denoiser. LISTA-SMO adjusts the parameter to enhance the signal and filters noise by the soft shrink function. Hence, we can directly train the neural network with the objective function in (11), which is

$$\mathcal{L} = \sum_{s=1}^{S} \|\mathbf{y}_s - \mathbf{g}^{(L)}\|_2. \tag{26}$$

This loss function is directly obtained from (8).

Comment 3: The neural network training by loss function (26) depends on the data $\mathbf{y}_s$, rather than the noiseless data $\mathbf{g}^*$, which directly eases the pressure on the data collection.

D. Computation Analysis

We summarize the total number of parameters in LISTA and LISTA-SMO in Table I, L represents the total number of layers, and N_c represents the number of parameters for a convolutional neural network in ISTA-net.

It can be observed that LISTA-SMO has much fewer parameters than LISTA due to the shared parameter matrices across all layers. This shared structure also allows for the parameters

TABLE I: The number of parameters for different neural networks

LISTA	LISTA-SMO
$L(1+D^2+N_dD)$	$2L+N+DN+DN_d$

to be updated multiple times (up to L times) in a single gradient back-propagation, resulting in a more efficient optimization process compared to LISTA, which only updates once. Despite having more parameters, simulation results in Section IV show that LISTA-SMO and LISTA exhibit similar convergence times.

E. Discussion

In this paper, we investigate the channel estimation problem in DMA-based communication systems. We first present a solution by formulating the DMA channel estimation problem as a compressed sensing problem, which we then solve with a highly efficient model-based learning algorithm, LISTA. To enhance the estimation accuracy, we also focus on optimizing the DMA weighting matrix and the sparsifying dictionary. To achieve this, we propose a new model learning algorithm, LISTA-SMO, which combines LISTA with the DMA layer and the sparse representation layer. This approach allows the DMA weighting matrix and the sparsifying dictionary to be treated as learnable parameters and optimized during the LISTA training stage. In addition, the proposed self-supervised learning for LISTA-SMO significantly lessens the cost of data collection and the negative effects of inaccurate data.

LISTA-SMO is an effective solution for the DMA channel estimation problem, delivering high accuracy with low computational costs due to the following reason. Firstly, LISTA-SMO outperforms traditional LISTA due to its ability to deeply integrate optimization matrices with the distribution of the channel's incident direction. This leads to improved recovery probability for multiple incident directions rather than the whole space. Indeed, it has the ability to achieve more accurate estimation with a smaller dictionary size. Besides, LISTA-SMO calculates the loss function directly based on the channel rather than the sparse channel, minimizing potential sparse representation errors in the loss function. Finally, LISTA-SMO is a highly efficient solution for the DMA channel estimation problem due to the optimization of the two matrices at the training stage,

Compared to previous studies, our proposed LISTA-SMO is unique in its combination of matrix optimization with deep unfolding algorithms. Unlike prior deep unfolding algorithms that only focused on unrolling the iterative optimization algorithm into a neural network layer, our algorithm unrolls two parts of our problem - the DMA signal processing process and the sparse representation process. Hence, in our algorithm, the DMA weighting matrix and sparsifying dictionary are treated as learnable parameters with a specific physical meaning. Our optimization of the sparsifying dictionary, as opposed to fitting it to a pre-designed matrix, sets our algorithm apart from previous studies.

Optimizing the DMA weighting matrix and sparsifying dictionary for the instantaneous channel is difficult, so LISTA-SMO optimizes these matrices offline on a data set $\mathbb{D}$. To ensure the effectiveness of the optimization, the angle of a channel path must follow a specific distribution and remain unchanged, and the data set $\mathbb{D}$ must be unbiasedly sampled from this distribution. However, constructing such a data set is challenging, as it requires additional measuring equipment to observe the input $\mathbf{y}$, which can be costly. Moreover, due to environmental changes, the distribution of $\mathbf{g}$ may regularly shift, necessitating updating the learning algorithm's parameters following the changes of $\mathbf{g}$. This can also be costly. To address these two issues, we plan to rebuild the neural network to accept $\mathbf{z}$ while still being able to update $\mathbf{Q}$. This will be the subject of our future research. By doing so, we aim to improve the efficiency and effectiveness of our LISTA-SMO algorithm in dealing with dynamic changes in the distribution of $\mathbf{g}$.

V. Numerical Results

In this section, we provide numerical results to verify the channel estimation accuracy of our proposed algorithm, which is quantified by the normalized MSE (NMSE) given by $\mathrm{NMSE} = \mathbb{E}\left\{\left\|\hat{\mathbf{h}}-\mathbf{h}\right\|_2^2/\|\mathbf{h}\|_2^2\right\}$, where $\hat{\mathbf{h}}$ is the estimated channel and $\mathbf{h}$ is the actual channel. We first present the simulation setup and the data set in Subsection V-A. Then, we numerically evaluate the convergence and estimation accuracy of LISTA-SMO in Subsection V-B. Specifically, we analyze the performance of DMA weighting matrices, different sparsifying dictionaries, and different CS algorithms. Besides, we compare the performance in different array settings.

A. Simulation Setup

1) DMA parameter: We consider a planar DMA shown in Fig. 1 throughout the experiment. The DMA comprises $N_d = 20$ microstrips and $N_e = 20$ elements in each microstrip. The DMA

works with the carrier frequency 28 GHz (λ = 1.07 cm). The element spaces are $\Delta d_x = \lambda/2$ and $\Delta d_y = \lambda/2$. We use $\alpha = 0.6$ m^{-1} and $\beta = 827.67$ m^{-1} to represent the propagation inside the DMA waveguides, assuming a microstrip implemented in Duroid 5880 with 30 mill thickness.

2) Channel Data Generation and Data Set: The channel $\mathbf{g}$ is randomly generated by following the Saleh-Valenzuela channel model in (4). In this model, the number of channel paths L_p is randomly chosen from 2 to 6. The channel path gain is generated by

$$a_l(d) = \sqrt{F(\phi_l)}\frac{\lambda}{4\pi d} \tag{27}$$

where d is the path length generated by a uniform distribution $U(15, 60)$. F is the radiation profile of each element, which is $F\left(\phi_{l_p}\right) = 2(b+1)\cos^b\left(\phi_{l_p}\right)$. b is the Boresight gain. We consider the dipole case, i.e., $b = 2$. The incident angle θ_{l_p} and ϕ_{l_p} are randomly sampled from a uniform distribution, which is $[(\varphi, \theta)|\varphi \sim \mathcal{U}(\frac{\pi}{12}, \frac{5\pi}{12}), \theta \sim \mathcal{U}(\frac{\pi}{6}, \frac{5\pi}{6})]$.

The sparse form channel $\boldsymbol{\alpha}$ is obtained by assuming the channel is incident from nearest grid.

We build the training set, test set, and validation, with the randomly generated the channel data. The training set, test set, and validation include 512000, 512, and 51200 channel data, respectively. The training set, test set are utilized in the training stage, whose sample have random SNR to increase its robustness to noise. The validation set is for testing the NMSE when the SNR is $[0, 3, \ldots, 21]$ dB. At the training stage, we extract 128 samples from the training set.

3) Comparing Algorithm and Parameter Setting: We compare our algorithm, LISTA-SMO, with the following algorithm:

- **OMP:** The algorithm was proposed in [43]. OMP was also utilized in hybrid MIMO channel estimation in [44]. OMP utilizes the random DMA matrix and the spatial gridding sparsifying dictionary.
- **FISTA:** The algorithm was proposed in [45], which is the modified ISTA. It was also utilized in hybrid MIMO channel estimation in [46]. FISTA utilizes the random DMA matrix and the spatial gridding sparsifying dictionary.
- **LISTA:** The algorithm is utilized to solve the DMA channel estimation in Section III. LISTA utilizes the random DMA matrix and the spatial gridding sparsifying dictionary.
- **ISTA-Net:** The algorithm was utilized to solve the hybrid antenna array channel estimation problem, as shown in the paper [15], [20]. ISTA-Net utilizes the random DMA matrix and the spatial gridding sparsifying dictionary.

- **LISTA with the DMA layer (LISTA+DMAL):** The neural network combines the LISTA and the DMA layer. We utilize it to show the promotion of the optimized DWM. LISTA+DMA utilizes the learned DMA weighting matrix and the spatial gridding sparsifying dictionary.
- **LISTA with the sparse representation layer (LISTA+SDL):** The neural network combines the LISTA and the sparse represent layer. We utilize it to show the promotion of the optimized sparsifying dictionary. LISTA+SDL utilizes the random DMA weighting matrix and the learned sparsifying dictionary.
- **LISTA-SMO:** The model-based algorithm that jointly optimizes the DWM and sparsifying dictionary in (24).

The sparsifying dictionaries are set as (9). The space grids are

$$[(\varphi,\theta)|\varphi = k_\varphi \frac{\pi}{60}, \theta = \frac{\pi}{18} + k_\theta \frac{2\pi}{225}, 10 \le k_\varphi \le 50, 1 \le k_\theta \le 40, k_\varphi \in \mathbb{Z}, k_\theta \in \mathbb{Z}]. \tag{28}$$

The number of grids is $D = 1600$. The LISTA+SDL and LISTA-SMO are designed to learn a sparsifying dictionary with $D = 400$.

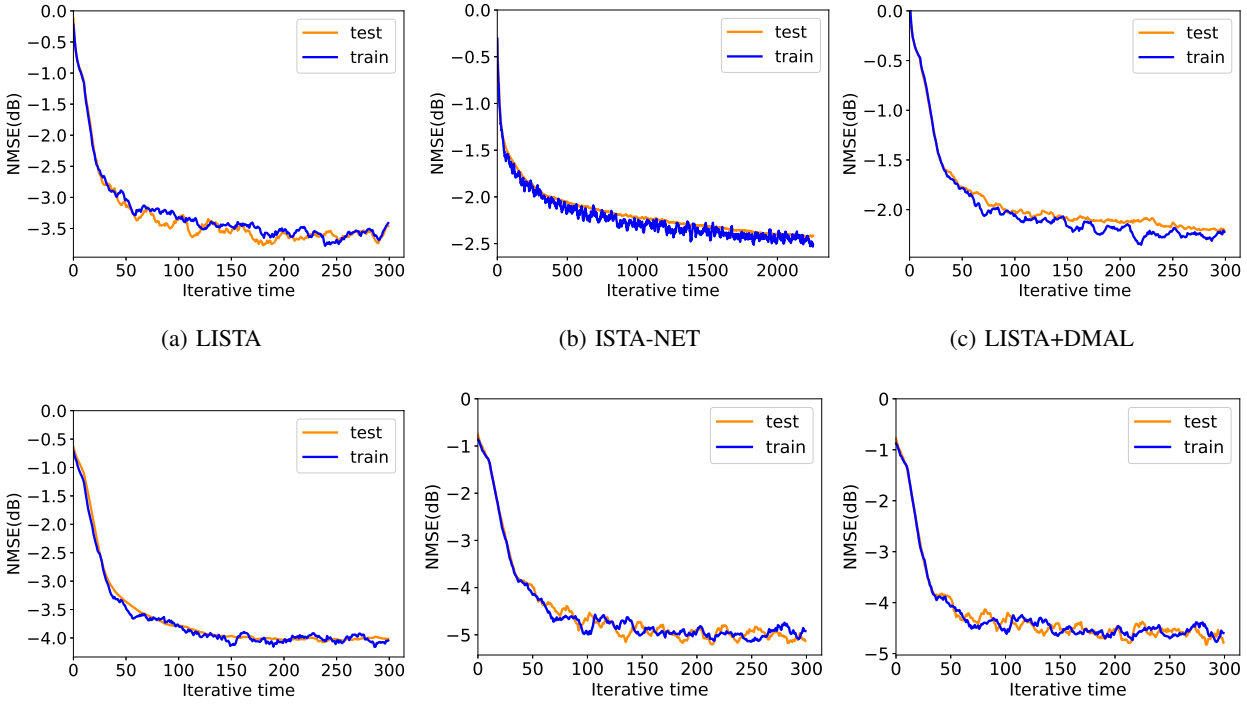
All the learning algorithms utilize the Adam optimizer with the same learning rate as 1e-4. The initial value of $\mathbf{W}_a, \mathbf{W}_b, \mathbf{W}_{\boldsymbol{\varphi}}, \mathbf{W}_{\mathbf{A}}, \mathbf{W}$ are random generated. The initial value of $\hat{\eta}$ and κ are 10^{-4} and 1, respectively.

B. Performance Analysis

We first show the efficiency of model-based learning algorithms by comparing the convergence performance of LISTA, ISTA-NET, LISTA+DMAL, LISTA+SDL, and LISTA-SMO in Fig. 4. During training, we sampled data from the training set without repetition, and defined an iterative time as all the samples being sampled once. We defined convergence as the point where the mean of NMSE of the test set in the last ten iterative times is no longer descending. In these figures, the blue and red lines represent the average NMSE on the training set and the test set, respectively.

The results showed that the algorithms based on LISTA converge in around 200 iterative times, while ISTA-NET needs more than 1500 iterative times to converge. This indicates that model-based algorithms have computational superiority. Moreover, all model-based learning algorithms show lower NMSE than ISTA-NET. These two conclusions indicate that model-based methods can extract the key feature more efficiently. Furthermore, as shown in Fig. 4 (a), (c), (d) and (e), the converge time of LISTA, LISTA+DMAL, LISTA+SDL, and LISTA-SMO are almost equal, demonstrating that the increments of training complexity raised by LISTA+DMAL, LISTA+SDL,

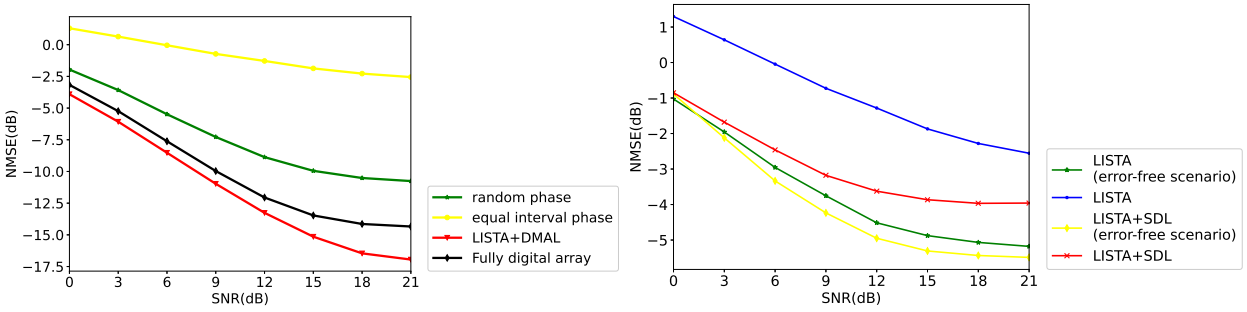
and LISTA-SMO are negligible. Besides, the convergence curves of LISTA-SMO in supervised learning and self-supervised learning are shown in Fig. 4(e) and (f). The comparison of figures shows the self-supervised learning method has similar coverage speed and slight performance loss.



(a) LISTA (b) ISTA-NET (c) LISTA+DMAL

(d) LISTA+SDL (e) LISTA-SMO(supervised) (f) LISTA-SMO(self supervised learning)

Fig. 4: The convergence curve of different algorithms.



(a) Performance comparing of different weighting matrices (b) Performance comparing of different sparsifying dictionaries

Fig. 5: The performance comparison of optimized matrices.

Next, in Fig. 5(a), we show the NMSE of different sensing matrices in different SNRs. To better illustrate the influence of weighting matrices, we compare the NMSE in a scenario in which channel data are generated without sparse representation error, i.e., the angle of channels' path is generated by randomly choosing from the gridding in (28). The black line corresponds to the NMSE that utilizes the LISTA algorithm to estimate the channel of hybrid antenna array with random weights matrix; the yellow, green and red lines correspond to the NMSE that utilizes the LISTA algorithm to estimate the DMA channel with equal-interval-phase weights matrix, random-phase weights matrix, and the weights matrix learned by LISTA+DMAL. Observing Fig. 5(a), we note that the fully-digital array with random weight outperforms 3 dB to DMA with random phase, which justifies Theorem 1 that the DMA random weighting matrix is one of the reasons for low estimation accuracy. Meanwhile, the figure shows that the learned sensing matrix outperforms the sensing matrix with random phase by 2 dB when the SNR is 0 dB and by 4 dB when the SNR is 21 dB. Furthermore, the performance of the learned sensing matrix is also higher than that of the fully-digital array. This comparison confirms the Comment 1 and the effectiveness of our approach that LISTA+DMAL reaches lower NMSE by fitting the data.

Also, we prove Comment 1 by comparing the eigenvalue of the signal subspace of $\mathbf{z}$, which is shown in Table.II. From the table, we have a similar conclusion above that our learned $\mathbf{Q}$ greatly improves the average signal strength.

TABLE II: The eigenvalue of signal subspace in (24)

Eigenvalue	Hybrid Antenna array	DMA with random weighing matrix	DMA with learned weighing matrix
1	3.13	2.35	9.14
2	2.41	1.79	3.45
3	1.60	1.27	2.42

Next, we illustrate the performance influenced by the sparsifying dictionary in Fig. 5(b). To show the performance loss brought by sparse representation, we also present the NMSE of LISTA and LISTA+SDL in the scenario without sparse representation error. The green and yellow lines represent the NMSE of LISTA and LISTA+SDL in the error-free scenario. the blue and red lines represent the NMSE of LISTA and LISTA with spare representation layer that trained on the data set in V-A.

We can observe from Fig. 5(b) that the NMSE in the error-free scenario is much lower than the NMSE on our data set, which testify the performance loss brought by the sparse representation. In the error-free scenario, LISTA and LISTA+SDL achieve similar performance. But a 2 dB performance gap is shown between the NMSE of LISTA+SDL and LISTA in the scenario with errors. This promotion indicates that our learned sparsifying dictionary can achieve better performance. The reason has two aspects: 1) Our learned sparsifying dictionary can represent the channel in sparse form with less sparse representation error, and 2) the LISTA+SDL avoids introducing sparse representation error into the loss function, which is mentioned in Comments 2. However, it can also be observed that the NMSE of LISTA+SDL still has a gap with the NMSE of the LISTA+SDL in the error-free scenario, which indicates that Equation (21) has difficulty in representing the channel in a real scenario.

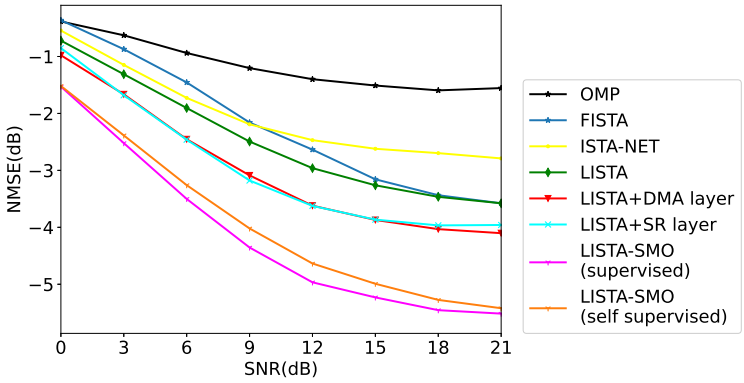
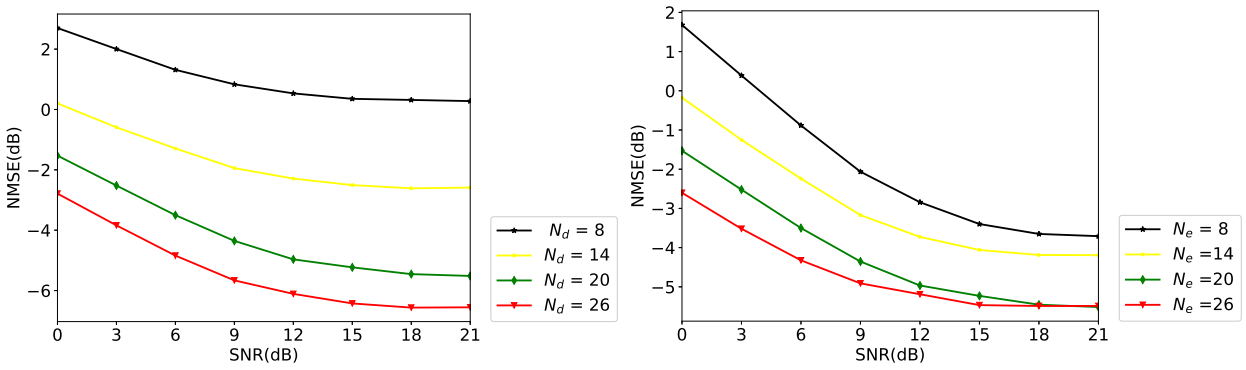


Fig. 6: The performance comparison of different algorithms

Then, in Fig. 6, the performance comparison of the algorithm listed in the subsection V-A is presented. Meanwhile, we compare the performance gap between supervised learning and self-supervised learning. In Fig. 6, LISTA-SMO shows significant superiority in overall algorithms with an almost 3 dB improvement over FISTA, 2 dB improvement over LISTA+DMAL, and 2 dB improvement over LISTA+SDL, indicating that the optimization of the DWM and the SD can collaborate to achieve higher estimation accuracy. We also observe that the NMSE of self-supervised learning is slightly higher than supervised learning, but the difference is small. As expected, LISTA-SMO, without the help of labels, cannot recover some weak channel components.

Finally, we examine the performance of LISTA-SMO with the impact of array size. In Fig. 7(a), the NMSE for the DMA with 8, 14, 20, and 26 microstrips and 20 elements is present. From

Fig. 7(a), the algorithm can hardly solve this problem when the number of microstrips is 14 because the information gathered by microstrips is not enough to recover the sparse signal. TConsequently, there is a significant performance gap between the Nd=14 and Nd=20 lines. Additionally, the figure demonstrates that increasing the number of microstrips leads to better performance. This is because a higher compression ratio can be achieved, which allows for more information to be retained. In Fig. 7(b), we show the performance of the DMA with the number of elements in each microstrip being 8, 14, 20, 26, and the microstrip of DMA is fixed in 20. As the number of elements increases, performance gains are primarily brought about by the higher degree of freedom of the dictionary because we set the dictionary dimension as the same with N. The DMA array with more elements can support more orthogonal vectors for dictionaries. But the performance gap between N_e = 20 and N_e = 26 is smaller than the gap between N_e = 14 and N_e = 20 due to the higher compressed ratio.



(a) Different Microstrip

(b) Different Element

Fig. 7: The SNR-NMSE curve in different DMA settings

VI. Conclusion

In this paper, we studied the channel estimation problem in DMA-assisted MIMO systems. We first formulated it as a CS problem by sparsifying the channel with the space-gridding sparsifying dictionary (SD). We solved the CS problem with an efficient model-based algorithm, LISTA, under the random DMA weighting matrix (DWM). However, the DWM and SD could result in decreased channel estimation accuracy, so we proposed a model-based learning algorithm, LISTA-SMO, to optimize the DWM and SD. LISTA-SMO is the LISTA neural network appended

with the DMA layer and the sparse representation layer, which are built by unfolding the DMA signal process and the sparsifying dictionary into neural network layers. The DMA layer and the sparse representation layer contain the DWM and SD as learnable parameters, and thereby, the optimized matrices can be optimized while LISTA training. To reduce the requirement of the data set, we proposed a self-supervised learning method to train LISTA-SMO. Simulation results demonstrated that the model-based algorithm, LISTA, outperforms the other channel estimation algorithms, and LISTA-SMO significantly overperformed LISTA. Our findings highlight the importance of optimizing the DWM and SD and the effectiveness of our algorithm.

APPENDIX A
PROOF OF THEOREM 1

Proof: Theorem 1 is mainly for demonstrating the influence of the DWM. We first introduce some definitions of CS that we will utilize in the proof.

Definition 1: (RIP) A matrix $\mathbf{\Psi}$ satisfied the Restricted Isometry Property (RIP) of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\left\|\boldsymbol{\alpha}\right\|_2^2 \leq \left\|\mathbf{\Psi}x\right\|_2^2 \leq (1+\delta_k)\left\|\boldsymbol{\alpha}\right\|_2^2, \tag{29}$$

holds for all $\|\boldsymbol{\alpha}\| = k$. δ_k is the k ord Restricted Isometry Constant (RIC).

Theorem 3: (Theorem 1.9 of [14]) When the sensing matrix $\mathbf{\Psi}$ satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2}-1$, then $\boldsymbol{\alpha}$ can be recovery from $\mathbf{z}$.

According to Definition A1, we need to investigate the $2L_p$ order restricted isometry constant of $\mathbf{Q}$. We denote $\mathbf{Q}_{2L_p}$ as the submatrix that randomly chooses the $2L_p$ column from $\mathbf{Q}$ with the same probability. To satisfy the RIP, $\mathbf{Q}_{2L_p}$ must be a full column rank matrix, which means it must choose $N_i = 2L_p$ irrelevant column of $\mathbf{Q}$. Following this, the probability of full column rank means choosing $2L_q$ elements from DMA with any N_i elements belonging to different microstrip.

Lemma 1: The probability that $\mathbf{Q}_{2L_p}$ is a full column rank matrix is $p = \frac{\mathcal{X}_{N_i}}{C_N^{2L_p}}$, where $\mathcal{X}_{N_i}$ is the combination of $2L_q$ elements comes from the N_i microstrips.

Proof: This is a typical combination problem. First, the total combination of selecting N_i column from $\mathbf{Q}$ is $C_N^{2L_q}$. Then, we need the combination of choosing $2L_q$ columns with N_i

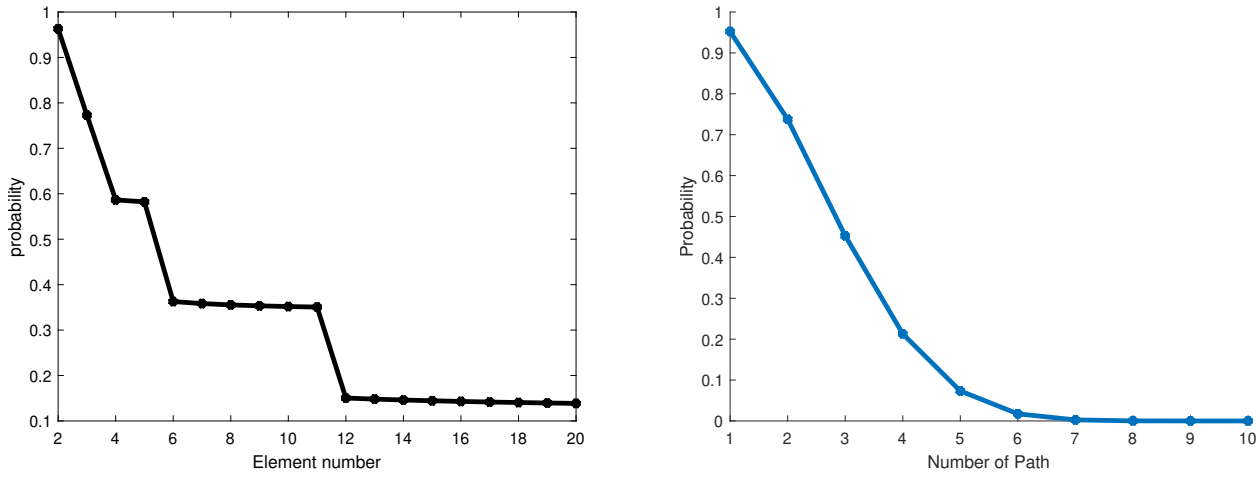
irrelevant columns. We can represent the combination of $2L_q$ elements that comes from the N_i microstrips is

$$\mathcal{X}_{N_i} = C_{N_d}^{N_i}(C_{(N_i)N_e}^{2L_p} - \sum_{i=1}^{N_i-1} C_{N_i}^{i}\mathcal{X}_i), \tag{30}$$

where $\mathcal{X}_1 = C_{N_e}^{2L_p}$. Following this, the probability is $p = \frac{\mathcal{X}_{N_i}}{C_N^{2L_p}}$.

■

From Lemma A1, we can get the probability that the $\mathbf{Q}_{2L_p}$ is a full column rank matrix is low. Also, we can see the probability of full rank is higher when there are fewer elements in each microstrip. Fig. 8 shows the probability of full column rank $\mathbf{Q}_{2L_p}$ with respect to the element number in each microstrip and the number of a path.



(a) the element number in each microstrip

(b) The number of Path

Fig. 8: Probability that $\mathbf{Q}$ is a full column rank matrix

Then, we investigate the RIC of $\mathbf{\Psi}_{2L_p}$ when it is a full column rank matrix. The RIC depends on the singular value of the sensing matrix, which is $1-\delta \leq s_{\min}(\mathbf{\Psi}) \leq s_{\max}(\mathbf{\Psi}) < 1+\delta$. For easy illustration, we assume the sparsifying direction is the orthogonal matrix. Then, the singular value of the sensing matrix only depends on the sensing matrix $\mathbf{Q}$. Because the DWM has only one nonzero value in each column, the singular value of $\mathbf{Q}$ can be written as

$$\mathbf{s}(\mathbf{Q}) = \left[\|\mathbf{q}_1\|_2, \cdots, \|\mathbf{q}_{N_d}\|_2\right], \tag{31}$$

where $\mathbf{q}_k$ is kth row of $\mathbf{Q}$.

The singular value of $\mathbf{Q}_{2L_p}$ follows the same form with (31). From (4) and (2), we have $\|q_{n,l}h_{n,l}\|_2 \leq \|h_{n,l}\|_2 \simeq 1$. We assume the phase of element's weighting value $\varphi_{n,l}$ follows $U[0, 2\pi]$ and $q_{n,l}h_{n,l}$ follows uniform distribution $U[0, 1]$. The k-th elements s_k can be approximated as

- **Case 1:** $s_k \sim U[0, 1]$, when $\|\mathbf{q}_k\|_0 = 1$;
- **Case 2:** $s_k \sim \mathcal{N}(\|\mathbf{q}_k\|_0/2, \|\mathbf{q}_k\|_0/12)$, when $\|\mathbf{q}_k\|_0 > 1$ (according to the central limit theorem).

We treat $s_{\min} = \min(s(\mathbf{Q}))$ and $s_{\max} = \max(s(\mathbf{Q}))$ are two different random variables. The cumulative distribution function for $s_{\min}, s_{\max}$ is

$$\begin{aligned} F(s_{\min}) &= 1 - (1 - F(s))^{2L_p} \\ F(s_{\max}) &= (F(s))^{2L_p} \end{aligned} . \tag{32}$$

where is the Then, the probability of $s_{\min} > (2 - \sqrt{2})/2$ and the probability of $s_{\max} < \sqrt{2}/2$ are both $(\frac{\sqrt{2}}{2})^{2L_p}$.

To sum up, when the phase of q follows the phase of element's weighting value $\varphi_{n,l}$ follows $U[0, 2\pi]$, the algorithm can recover the sparse signal $\boldsymbol{\alpha}$ from (10) from the signal with the probability of $\frac{\mathcal{X}_{2L_p}}{C_N^{2L_p}}(\frac{\sqrt{2}}{2})^{2L_p}$. Theorem 1 is proofed.

■

APPENDIX B
PROOF OF THEOREM 2

Proof: We assume the actual channel is incident from θ_t, ϕ_t, and the incident angles of the corresponding sparse channel are θ_d, ϕ_d. Then, the loss of estimation accuracy of a single path can be represented as $\varrho = a_{l_p}\|\mathbf{a}(\theta_t, \phi_t) - \mathbf{a}(\theta_d, \phi_d)\|$, which depends on $\|\theta_t - \theta_d\|, \|\phi_t - \phi_d\|$. As the θ_t, ϕ_t follows the uniform distribution, $\mathbb{E}\{\|\theta_t - \theta_d\|\} = \Delta\theta/4$, $\mathbb{E}\{\|\phi_t - \phi_d\|\} = \Delta\phi/4$.

Hence, we have

$$\begin{aligned}\varrho_{\text{mean}} &= \|\mathbf{a}\left(\theta_t,\phi_t\right) - \mathbf{a}\left(\theta_t+\frac{\Delta\theta}{4},\phi_t+\frac{\Delta\phi}{4}\right)\| = 2-2\mathbf{a}\left(\theta_t,\phi_t\right)\mathbf{a}\left(\theta_t+\frac{\Delta\theta}{4},\phi_t+\frac{\Delta\phi}{4}\right)^{\mathrm{H}} \\ &= 2-2\sum_{n_e=1}^{N_e}\sum_{n_d=1}^{N_d}\exp\left(j2\pi n_e\frac{\Delta d_x}{\lambda}(\cos\theta_t\sin\phi_t-\cos(\theta_t+\frac{\Delta\theta}{4})\sin(\phi+\frac{\Delta\phi}{4}))\right) \\ &\quad \exp\left(j2\pi n_d\frac{\Delta d_y}{\lambda}(\sin\theta_t\sin\phi_t-\sin(\theta_t+\frac{\Delta\theta}{4})\sin(\phi+\frac{\Delta\phi}{4}))\right) \\ &\simeq 2-2\exp\left(1-\cos\frac{\Delta\theta}{4}\cos\frac{\Delta\phi}{4}\right)\end{aligned} \tag{33}$$

in which we utilize the approximation $\cos(\theta_t+\frac{\Delta\theta}{4})\sin(\phi+\frac{\Delta\phi}{4}) \simeq \cos\theta_t\sin\phi_t\cos(\frac{\Delta\theta}{4})\cos(\frac{\Delta\phi}{4})$ and $\sin(\theta_t+\frac{\Delta\theta}{4})\sin(\phi+\frac{\Delta\phi}{4}) = \sin(\theta_t)\sin(\phi)\cos(\frac{\Delta\theta}{4})\cos(\frac{\Delta\phi}{4})$, as $\sin\frac{\Delta\theta}{4}\simeq 0$ and $\sin\frac{\Delta\phi}{4}\simeq 0$. Hence, the proof is completed. ■

REFERENCES

[1] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[2] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, 2020.

[3] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6g extreme massive MIMO communications," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 106–113, Apr. 2021.

[4] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[5] N. Shlezinger, O. Dicker, Y. C. Eldar, I. Yoo, M. F. Imani, and D. R. Smith, "Dynamic metasurface antennas for uplink massive MIMO systems," *IEEE transactions on communications*, vol. 67, no. 10, pp. 6829–6843, 2019.

[6] H. Wang, N. Shlezinger, Y. C. Eldar, S. Jin, M. F. Imani, I. Yoo, and D. R. Smith, "Dynamic Metasurface Antennas for MIMO-OFDM Receivers With Bit-Limited ADCs," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2643–2659, Apr. 2021.

[7] I. Yoo, M. F. Imani, T. Sleasman, H. D. Pfister, and D. R. Smith, "Enhancing capacity of spatial multiplexing systems using reconfigurable cavity-backed metasurface antennas in clustered mimo channels," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1070–1084, 2018.

[8] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Beam focusing for near-field multi-user MIMO communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[9] H. Jiang, L. You, J. Wang, W. Wang, and X. Gao, "Hybrid RIS and DMA assisted multiuser MIMO uplink transmission with electromagnetic exposure constraints," *IEEE Journal of Selected Topics in Signal Processing*, 2022.

[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

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

[12] A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Sparse Bayesian learning-based channel estimation in millimeter wave hybrid MIMO systems," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2017, pp. 1–5.

[13] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2748–2759, 2017.

[14] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[15] W. Jin, H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Adaptive channel estimation based on model-driven deep learning for wideband mmwave systems," *arXiv:2104.13656*, Sep. 2021.

[16] L. Wan, K. Liu, and W. Zhang, "Deep learning-aided off-grid channel estimation for millimeter wave cellular systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3333–3348, 2022.

[17] Z. Chen, J. Tang, X. Zhang, Q. Wu, Y. Wang, D. K. C. So, S. Jin, and K.-K. Wong, "Offset learning based channel estimation for intelligent reflecting surface-assisted indoor communication," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2021.

[18] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *arXiv preprint arXiv:2012.08405*, 2020.

[19] J. Sun, H. Li, Z. Xu *et al.*, "Deep admm-net for compressive sensing mri," *Advances in neural information processing systems*, vol. 29, 2016.

[20] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, Jun. 2018, pp. 1828–1837.

[21] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *2010 18th European Signal Processing Conference*, 2010, pp. 427–431.

[22] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[23] N. González-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wideband channel tracking and hybrid precoding for mmwave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2161–2174, 2021.

[24] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmwave massive MIMO with hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8954–8958, 2018.

[25] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[26] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[27] B. Dumitrescu and P. Irofti, *Dictionary learning algorithms and applications*. Springer, 2018.

[28] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.

[29] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for fdd massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5437–5451, 2018.

[30] W.-C. Chang and Y. T. Su, "Sparse bayesian learning based tensor dictionary learning and signal recovery with application to mimo channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 847–859, 2021.

[31] H. Xie and N. González-Prelcic, "Dictionary learning for channel estimation in hybrid frequency-selective mmwave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7407–7422, 2020.

[32] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-fed metasurface antenna," *Phys.rev.applied*, vol. 8, no. 5, p. 054048, 2017.

[33] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 4558–4563.

[34] J. Yang, C.-K. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmwave systems via cosparse image reconstruction technique," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4767–4782, Oct. 2018.

[35] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in Millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

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

[37] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[38] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 399–406.

[39] J. Zhang, C. Pan, F. Pei, G. Liu, and X. Cheng, "Three-dimensional fading channel models: A survey of elevation angle research," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 218–226, 2014.

[40] J. Bian, C.-X. Wang, X. Gao, X. You, and M. Zhang, "A General 3D Non-Stationary Wireless Channel Model for 5G and Beyond," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 3211–3224, May 2021.

[41] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds," *arXiv:1808.10038*, Nov. 2018.

[42] X. C. Jialin Liu, "Alista: analytic weights are as good as learned weights in lista," *International Conference on Learning Representations (ICLR)*.

[43] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

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

[45] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 693–696.

[46] L. Yang, Y. Zeng, and R. Zhang, "Efficient channel estimation for millimeter wave MIMO with limited rf chains," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.