Deep Learning-Based Near-Field Wideband Channel Estimation: A Joint LISTA-CP Approach

Jiapan Yang, Bo Ai, *Fellow, IEEE*, Wei Chen, *Senior Member, IEEE*,
Songjie Yang, Guowei Shi, Ning Wang, *Member, IEEE*, Chau Yuen, *Fellow, IEEE*

***Abstract*—Extremely large-scale multiple-input-multiple-output (XL-MIMO) is listed as one of the key 6G candidate technologies due to its ability to exponentially increase the communication spectrum efficiency. However, wideband XL-MIMO channel estimation faces a new challenge known as near-field beam split effect. To cope with this challenge, we propose a deep learning based joint learned iterative shrinkage thresholding algorithm with partial weight-coupled (Joint LISTA-CP) to accurately recover the channel. Specifically, by studying the characteristics of the near-field wideband channel, we employ a frequency-dependent polar domain dictionary to ensure that the sparse polar domain channel vectors across different subcarriers share a common sparse support. Then, we transform the near-field wideband channel estimation problem into a multiple measurement vector recovery problem. Utilizing the common sparsity structure of different subcarriers, we design the Joint LISTA-CP network that integrates multiple sub-networks. The proposed network employs a multidimensional shrinkage thresholding operator to guarantee the common sparse property of different subcarrier channels. In addition, we rigorously demonstrate that the Joint LISTA-CP network achieves lower recovery error. Simulation results show that the Joint LISTA-CP method enhances the estimation accuracy of the near-field wideband channel.**

***Index Terms*—Deep learning, XL-MIMO, wideband channel estimation, near-field.**

I. Introduction

With technological progress and rising user demand, the sixth-generation (6G) mobile communication networks will have to accommodate higher data rates and increased communication capacity [1]. Expanding on the massive multiple-input multiple-output (MIMO) technology utilized in the fifth-generation (5G), 6G will progress to extremely large-scale MIMO (XL-MIMO) [2]–[5]. This will involve a substantial increase in the number of antennas, thereby improving the system's spectral efficiency. Additionally, 6G networks will utilize higher frequencies, such as millimeter wave (mmWave) and terahertz (THz) bands, which provides a broader spectrum resource essential for achieving higher data rates. The integration of XL-MIMO with mmWave/THz technology is considered a key advancement for 6G [6]–[8].

In XL-MIMO systems, obtaining highly accurate channel state information (CSI) is crucial for maximizing system performance. However, due to the extremely large number of antennas in XL-MIMO systems, the base station (BS) typically employs a hybrid precoding architecture to lower hardware costs and reduce power consumption. Considering that the number of radio frequency (RF) chains in this architecture is significantly lower than the number of antennas, the channel sampling dimension is reduced, which poses significant challenges to accurate channel estimation [9]–[11]. On the other hand, the transition from massive MIMO to XL-MIMO not only entails a significant increase in the number of antennas but also brings fundamental changes in the electromagnetic (EM) structure [12]. The EM radiation field can be divided into far-field and near-field regions, distinguished by the Rayleigh distance. The Rayleigh distance is directly proportional to the square of the array aperture and inversely proportional to the carrier wavelength. In the far-field region, the EM field can be approximated using a plane wave model. Conversely, in the near-field region, the EM field must be modeled using a spherical wave model. In 5G systems, the Rayleigh distance is relatively small and is usually negligible. However, with a significant increase in the number of antennas in 6G, the near-field region also expands significantly [13]. Consequently, a non-negligible portion of 6G communication will take place in the near-field region.

In current 5G massive MIMO systems, researchers employ a discrete Fourier transform (DFT) dictionary to represent the sparse channel in the angle domain, aiming to minimize pilot overhead for channel estimation. By leveraging angle domain sparsity of massive MIMO channels [14]–[17], compressive sensing (CS) algorithms have been proposed. For example, the orthogonal matching pursuit (OMP) is used to estimate the angle-domain channel of a single subcarrier in [14]. Taking further advantage of the common sparsity structure across

This work was supported by the Fundamental Research Funds for the Central Universities (2022JBQY004), the Natural Science Foundation of China (U2468201, W2421083, 62221001), the State Scholarship Fund, China Scholarship Council (202307090109). The work of Chau Yuen was supported by the Agency for Science, Technology and Research Singapore (M22L1b0110), the Ministry of Education Singapore Tier 2 (T2EP50124-0032), the National Research Foundation (FCP-NTU-RG-2024-025). *(Corresponding authors: Bo Ai; Wei Chen.)*

Jiapan Yang, Bo Ai and Wei Chen are with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China, and also with School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: {jiapanyang,boai,weich}@bjtu.edu.cn). Bo Ai is also with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China.

Songjie Yang is with the National Key Laboratory of Wireless Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: yangsongjie@std.uestc.edu.cn).

Guowei Shi is with the China Academy of Information and Communications Technology, Beijing 100083, China (e-mail: shigw@sohu.com).

Ning Wang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: ienwang@zzu.edu.cn).

Chau Yuen is with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798 (e-mail: chau.yuen@ntu.edu.sg).

multiple subcarrier channels, a simultaneous OMP (SOMP) algorithm is proposed in [15]. Similarly, leveraging the common sparsity among channels, a dynamic sparsity-adjusting approximate message passing (AMP) method is proposed in [16], which iteratively prunes and updates the support set to achieve more accurate sparsity estimation. These channel estimation methods assume that the channel angles perfectly align with the angle grids of the dictionary. However, in practical communication systems, the angles may deviate from the grid, reducing the sparsity in the angle-domain channel and leading to performance degradation in the above methods. To resolve this problem, the method proposed in [17] first uses singular value decomposition (SVD) to obtain a coarse angle estimate, and then iteratively refines the estimated angle towards the actual value using gradient descent, thereby achieving super-resolution estimation. In [18], the atomic norm minimization algorithm is employed to search over an infinitely precise range of angles, thereby improving the estimation performance. In [19], high-accuracy channel estimation is achieved by iteratively refining the angle estimates through minimizing an objective function considering the orthogonality of the subspaces. These schemes are designed based on the sparsity of the channel in the angle domain. Unfortunately, when using a DFT dictionary to transform the near-field channel into the angular domain, the energy of a single near-field path spreads across multiple angles. This results in the near-field channel losing its sparsity in the angular domain.

To achieve the sparsity of the near-field channel, a method leveraging the polar-domain dictionary is proposed in [20]. This polar-domain dictionary considers both angular and distance characteristics, allowing the near-field channel to exhibit sparsity in the polar-domain. A polar-domain based SOMP algorithm (P-SOMP) is proposed for accurate channel estimation in [20]. In [21], the authors address the hybrid-field channel estimation problem, where the number of far-field and near-field paths is known. The channel is estimated based on the sparsity in the far-field angle domain and near-field polar domain, respectively. When the number of far-field and near-field paths is unknown, [22] proposes a channel estimation scheme based on off-grid random gradient tracking to address this issue. In [23], for the channel estimation problem in ultra-large-scale reconfigurable intelligent surface-assisted mmWave communication systems, a three-dimensional multi-measurement vector compressive sensing framework based on polar-domain sparsity is proposed to improve parameter estimation accuracy. In addition to the traditional CS-based algorithms mentioned above, deep learning (DL) methods have been widely applied in the communication field [24]–[26]. Here, we mainly introduce DL-based methods for near-field channel estimation [27]–[31]. Specifically, a multiple residual dense network for near-field channel estimation is proposed in [27] to leverage the polar-domain channel sparsity. A hybrid-field channel estimation method is introduced in [28], where the orthogonal matching pursuit method is first used for initial estimation, and then the convolutional autoencoder (CAE) is used for refined estimation. To address the issue of large storage overhead of polar-domain dictionary, a dictionary learning layer is introduced into the learned iterative shrinkage thresholding algorithm (LISTA) network in [29], leading to a dictionary with fewer atoms.

Near-field wideband systems can significantly enhance data transmission rates and system capacity, making them a focal point of research [32]–[35]. However, in wideband XL-MIMO systems, the large bandwidth and array aperture give rise to beam split effects, resulting in inconsistent sparse supports in the polar-domain channel across different frequencies. To accurately estimate channels in near-field wideband systems, [36] reveals that the sparse supports in both the distance and angle domains grow linearly with the subcarrier frequency. This property is utilized to estimate the angle and distance parameters of near-field paths across all frequencies, improving the accuracy of channel estimation. In [37], a novel frequency-dependent polar dictionary is proposed, which uses the unitary approximate message passing-sparse Bayesian learning (UAMP-SBL) based deep unfolding method and exploits the inherent properties of channel patterns for channel estimation. With the use of frequency-dependent dictionaries, a knowledge-driven wideband mixed-field channel estimation network is proposed in [38]. By using frequency-dependent redundant dictionaries, the sparse channel vectors of different subcarriers exhibit common sparse supports. Utilizing this property, a deep unfolding-based AMP algorithm is designed to enhance channel estimation. While these methods enhance the performance of channel estimation, they suffer from drawbacks such as a large number of network parameters and high computational complexity.

In this paper, we propose a novel deep unfolding-based method for channel estimation in near-field wideband XL-MIMO systems. To address the beam split effect in wideband XL-MIMO systems, we employ a frequency-dependent polar domain dictionary, which ensures that the sparse support set of the corresponding channel remains consistent across different frequencies. To achieve low-complexity channel estimation, we design a channel estimation method that combines multiple subnetworks, namely joint LISTA with partial weight-coupled (Joint LISTA-CP). To reduce the number of learnable parameters in the proposed network, we adopt a LISTA with partial weight-coupled (LISTA-CP) network with fewer parameters in each subnetwork. Additionally, we rigorously prove that the Joint LISTA-CP network achieves a lower recovery error compared to a single LISTA-CP network. The key contributions of this paper are summarized as follows.

- We propose a deep unfolding approach for channel estimation in near-field wideband XL-MIMO systems. To overcome the impact of the near-field beam split, we employ a frequency-dependent polar-domain dictionary, ensuring that the near-field sparse channels across different subcarriers share a common sparse support set. This allows us to transform the near-field wideband channel estimation problem into a multi-measurement sparse signal recovery problem.
- Leveraging the common sparse structure across different subcarriers, we design a Joint LISTA-CP network for channel estimation, where each subnetwork employs a LISTA-CP network with fewer network parameters. Due to the common sparse support, we combine multiple

sparse vectors into a matrix with group-row sparsity and apply a multidimensional shrinkage thresholding operator to ensure group-row sparsity.

- We conduct rigorous theoretical analysis to unveil why the proposed Joint LISTA-CP network can achieve a faster convergence rate. Theoretical analysis proves that the Joint LISTA-CP network achieves a lower recovery error. Finally, we show that the performance of the Joint LISTA-CP algorithm, which combines multiple sub-networks, is superior to that of the individual LISTA-CP network.

The rest of the paper is arranged as follows. Section II provides a detailed explanation of the system model and problem formulation. Section III presents the ISTA, LISTA, LISTA-CP and the proposed Joint LISTA-CP algorithms for XL-MIMO wideband channel estimation. The convergence analysis of the proposed Joint LISTA-CP algorithm is presented in Section IV. Section V illustrates the simulation results of the Joint LISTA-CP algorithm. Finally, the paper is concluded in Section VI.

Notations: Bold uppercase and lowercase letters represent matrices and vectors, respectively. $(\cdot)^{-1}$ denotes the inverse operation and $(\cdot)^T$ defines the transposition operation. Let $\mathbb{R}$ be the set of real numbers and $\mathbb{C}$ be the set of complex numbers. Given index sets $\mathcal{I} = \{i_1, \ldots, i_n\} \subset N$ and $\mathcal{J} = \{j_1, \ldots, j_m\} \subset M$, for $\mathbf{A} \in \mathbb{R}^{N \times M}$, $\mathbf{A}[\mathcal{I}, \mathcal{J}]$ denotes the submatrix of $\mathbf{A}$ corresponding to the rows indexed by $\mathcal{I}$ and the columns indexed by $\mathcal{J}$. We denote the i-th row and the j-th column of matrix $\mathbf{A}$ by $\mathbf{A}(i,:)$ and $\mathbf{A}(:,j)$, respectively. For a matrix $\mathbf{A}$, $\|\mathbf{A}\|_F$ represents the Frobenius norm, while $\|\mathbf{A}\|_{2,0}$ and $\|\mathbf{A}\|_{2,1}$ denote the mixed ℓ_0/ℓ_2 and ℓ_1/ℓ_2 norms, respectively. $\|\cdot\|_{row,0}$ represents the count of non-zero rows in the input matrix. For a vector $\mathbf{v}$, $\|\mathbf{v}\|_0$ represents the ℓ_0 norm, which is used to count the number of non-zeros elements, and $\|\mathbf{v}\|_1$ means the ℓ_1 norm, which is used to calculate the sum of the absolute values of the vector elements. $\|\mathbf{v}\|_2$ represents the Euclidean norm. The support of $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathbb{R}^N$ is represented as $\mathrm{supp}(\mathbf{v})$.

II. SYSTEM MODEL

In this section, we introduce a wideband near-field communication system based on the THz band. Then, we present the XL-MIMO wideband channel model and a frequency-dependent polar-domain dictionary. Finally, the XL-MIMO channel estimation process is formulated as a classical sparse signal recovery problem.

A. Singal Model

We consider the uplink channel estimation in an XL-MIMO system with orthogonal frequency division multiplexing (OFDM). In this system, the base station (BS) employs a uniform linear array (ULA) with N_{BS} antennas and N_{RF} radio frequency (RF) chains to serve K single-antenna users, as shown in Fig. 1. In this system, we adopt a time-division duplex (TDD) mode, where K users transmit orthogonal pilot signals to the BS. Let $s_{m,p}$ denotes the pilot signal on the m-th subcarrier in the p-th time slot. The length of the pilot sequence is P. The received signal $\tilde{\mathbf{y}}_{m,p} \in \mathbb{C}^{N_{RF}}$ is denoted as

$$\tilde{\mathbf{y}}_{m,p} = \mathbf{W}_p \mathbf{h}_m s_{m,p} + \tilde{\mathbf{n}}_{m,p}, \tag{1}$$

where $\mathbf{W}_p \in \mathbb{C}^{N_{RF} \times N_{BS}}$ represents the combining matrix, and $\mathbf{n}_{m,p} \in \mathbb{C}^{N_{RF}}$ is complex Gaussian noise following $\mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$. Let $s_{m,p} = 1$ for $p = 1, 2, \ldots, P$. Considering the total P time slots for pilot transmission, we can obtain

$$\tilde{\mathbf{y}}_m = \mathbf{W}\mathbf{h}_m + \tilde{\mathbf{n}}_m, \tag{2}$$

where $\tilde{\mathbf{y}}_m = [\tilde{\mathbf{y}}_{m,1}^T, \tilde{\mathbf{y}}_{m,2}^T, \ldots, \tilde{\mathbf{y}}_{m,P}^T]^T \in \mathbb{C}^{PN_{RF}}$ and $\tilde{\mathbf{n}}_m = [\tilde{\mathbf{n}}_{m,1}^T, \tilde{\mathbf{n}}_{m,2}^T, \ldots, \tilde{\mathbf{n}}_{m,P}^T]^T \in \mathbb{C}^{PN_{RF}}$. Besides, $\mathbf{W} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \ldots, \mathbf{W}_P^T]^T \in \mathbb{C}^{PN_{RF} \times N_{BS}}$ is the overall observation matrix, where each element is independently sampled from the binomial distribution $\frac{1}{\sqrt{N_{BS}}}\{-1, 1\}$.

B. Channel Model

Compared to massive MIMO narrowband channel model, the wideband XL-MIMO channel model within the THz band is more complex. It arises from the near-field beam split effect, which influences the array response vector based on factors like angles, distances, and subcarrier frequency. The Saleh-Valenzuela channel model is adopted here. The channel $\mathbf{h}_m \in \mathbb{C}^{N_{BS}}$ at the m-th subcarrier is defined as

$$\mathbf{h}_m = \sqrt{\frac{N_{BS}}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{j=1}^{N_p} \alpha_{i,j} e^{-j2\pi f_m \tau_{i,j}} \mathbf{b}(\phi_{i,j}, r_{i,j}, f_m), \tag{3}$$

where N_c refers to the number of clusters, and N_p represents the number of subpaths in each cluster. The subscripts i and j correspond to the indices of the clusters and subpaths, respectively. τ denotes the delay of the subpath, and α denotes the path gain. In addition, $f_m = f_c + (\frac{2m-M}{2M})B$ represents the frequency of the m-th subcarrier, where B represents the bandwidth and M is the total number of subcarriers. $\phi_{i,j}$ denotes the AoA of the i-th cluster and j-th subpath. $\mathbf{b}(\phi_{i,j}, r_{i,j}, f_m)$ is the array response vector. Based on the geometrical structure of the BS array, $\mathbf{b}(\phi_{i,j}, r_{i,j}, f_m)$ is formulated as

$$\begin{aligned}\mathbf{b}(\phi_{i,j}, r_{i,j}, f_m) = &\sqrt{\frac{1}{N_{BS}}}\Big[e^{-j2\pi \frac{c}{f_m}(r_{i,j}^{(0)} - r_{i,j})}, \ldots, \\ &e^{-j2\pi \frac{c}{f_m}(r_{i,j}^{(N_{BS}-1)} - r_{i,j})}\Big]^T,\end{aligned} \tag{4}$$

where ${r_{i,j}}^{(n)}$ represents the distance between the scatterer and the n-th BS antenna. As described in [20], ${r_{i,j}}^{(n)}$ can be expressed as

$${r_{i,j}}^{(n)} = \sqrt{r_{i,j}^2 + \delta_n^2 d^2 - 2r_{i,j}\delta_n d \sin\phi_{i,j}}, \tag{5}$$

where $\delta_n = n - \frac{N-1}{2}$ with $n \in \{0, 1, \ldots, N_{BS} - 1\}$. Additionally, $d = \frac{\lambda_m}{2}$ represents the antenna spacing, where $\lambda_m = \frac{c}{f_m}$ is the carrier wavelength.

C. Problem Formulation

In communication systems, channel estimation based on transform-domain sparsity is an effective method to reduce pilot overhead and improve estimation performance. The core idea is to transform the channel information from the time or

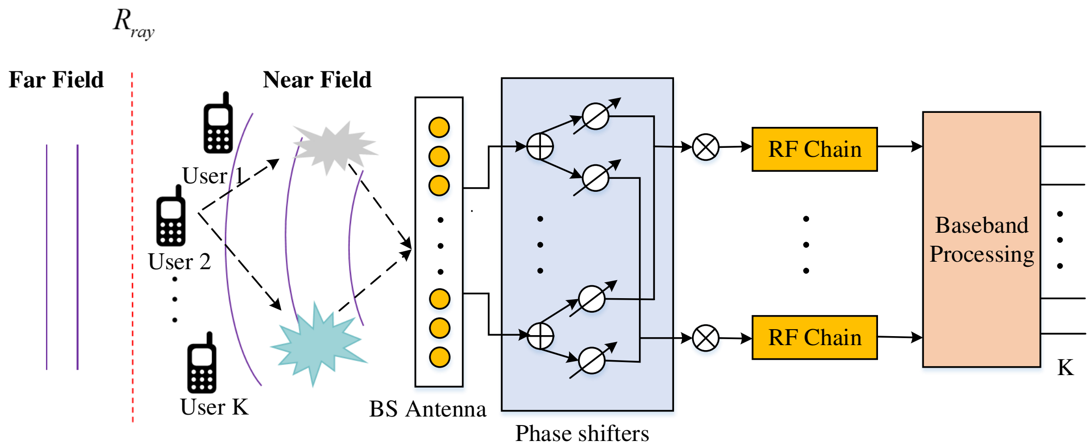

Fig. 1. A Wideband THz XL-MIMO system with the hybrid precoding structure, where R_{ray} represents the Rayleigh distance.

frequency domain to a sparse representation domain, where the channel exhibits sparsity. This sparsity can be leveraged to substantially reduce the number of pilots for channel estimation, and the CSI can be accurately recovered using algorithms such as OMP, AMP and UAMP-SBL. In far-field channel estimation, researchers commonly utilize angle-domain sparsity of the channel with the DFT dictionary. However, in near-field channel estimation, the angle domain based on the DFT dictionary is no longer sparse. A polar-domain dictionary is proposed in [20] to capture the sparsity of near-field channels. The polar-domain dictionary samples from both angle and distance dimensions, and each column is composed of the near-field steering vector. However, in near-field wideband systems, due to the beam split effect, when all subcarriers use the same polar-domain dictionary, different subcarriers exhibit different sparsity support sets, which poses significant challenges for channel estimation. To ensure that different subcarriers in near-field wideband systems share the same sparsity support set and thereby reduce the complexity of channel estimation, we adopt a frequency-dependent polar-domain dictionary. Specifically, the dictionary for the m-th subcarrier is expressed as

$$\mathbf{B}^m = [\mathbf{B}_0^m, \mathbf{B}_1^m, \dots, \mathbf{B}_{S-1}^m], \tag{6}$$

where S represents the number of distance grids. The s-th sub-dictionary $\mathbf{B}_s^m$ is made up of Q near-field steering vectors

$$\mathbf{B}_s^m = [\mathbf{b}(\phi_0, r_{s,0}, f_m), \dots, \mathbf{b}(\phi_{Q-1}, r_{s,Q-1}, f_m)], \tag{7}$$

where Q represents the number of angle grids. Thus, the total number of grids in the polar domain is given by $G = SQ$. To reduce the maximum coherence between any two atoms, the angles need to be uniformly sampled, while the distances should be sampled non-uniformly, as described

$$\phi_q = \frac{2q - Q + 1}{Q}, \quad q = 0, \cdots, Q-1, \tag{8}$$

$$r_{s,q} = \frac{1}{s} Z_\triangle (1 - \phi_q^2), \quad s = 0, \cdots, S-1, \tag{9}$$

where $Z_\triangle = \frac{N^2 d^2}{2\beta_\triangle^2 c/f_m}$, $\beta_\triangle$ is the threshold used to balance coherence level and grid resolution. S is the smallest integer satisfying $\frac{1}{S} Z_\triangle < \rho_{min}$ with ρ_{min} representing the minimum distance [20]. Utilizing the dictionary in (6), the domain transformation can be readily performed as

$$\mathbf{h}_m = \mathbf{B}^m \bar{\mathbf{x}}_m, \tag{10}$$

where $\bar{\mathbf{x}}_m \in \mathbb{C}^G$ denotes the m-th sparse polar domain sub-channel. Therefore, we can rewrite (2) as

$$\tilde{\mathbf{y}}_m = \mathbf{W}\mathbf{B}^m \bar{\mathbf{x}}_m + \tilde{\mathbf{n}}_m = \bar{\mathbf{D}}_m \bar{\mathbf{x}}_m + \tilde{\mathbf{n}}_m, \tag{11}$$

where $\bar{\mathbf{D}}_m = \mathbf{W}\mathbf{B}^m \in \mathbb{C}^{PN_{RF} \times G}$ represents the sensing matrix for the m-th subcarrier. To facilitate the algorithm design, we restructure (11) by stacking the real and imaginary components together as

$$\begin{aligned}\mathbf{y}_m &= \mathbf{D}_m \mathbf{x}_m + \mathbf{n}_m \\ &= \begin{bmatrix} Re\{\bar{\mathbf{D}}_m\} & -Im\{\bar{\mathbf{D}}_m\} \\ Im\{\bar{\mathbf{D}}_m\} & Re\{\bar{\mathbf{D}}_m\} \end{bmatrix} \begin{bmatrix} Re\{\bar{\mathbf{x}}_m\} \\ Im\{\bar{\mathbf{x}}_m\} \end{bmatrix} + \begin{bmatrix} Re\{\bar{\mathbf{n}}_m\} \\ Im\{\bar{\mathbf{n}}_m\} \end{bmatrix},\end{aligned} \tag{12}$$

where $Re\{\cdot\}$ and $Im\{\cdot\}$ represent the real and imaginary part of the complex matrix, respectively. To estimate the sparse signal $\mathbf{x}_m$, we can address the optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}_m} \ \|\mathbf{x}_m\|_0 \\ &\text{s.t.} \quad \|\mathbf{y}_m - \mathbf{D}_m \mathbf{x}_m\|_2 \le \delta, \end{aligned} \tag{13}$$

where δ denotes the error tolerance parameter. By substituting the ℓ_0-norm with the ℓ_1-norm, various CS algorithms can be employed to address this problem.

III. PROPOSED DL-BASED CHANNEL ESTIMATION

In this section, we begin with a classic sparse signal reconstruction algorithm, i.e., the iterative shrinkage thresholding algorithm (ISTA), to solve the problem mentioned above. The main idea of ISTA is to iteratively approximate the solution to the problem, enforcing sparsity through the use of a soft-thresholding function. Although ISTA is capable of finding the optimal solution to the problem, its convergence rate is relatively slow. LISTA is an algorithm that accelerates the convergence of ISTA using deep learning techniques. The LISTA model with weight coupling algorithm (LISTA-CP) is a variant of LISTA, which introduces partial weight

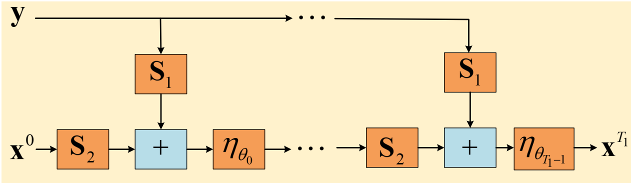

Fig. 2. The structure of the LISTA network.

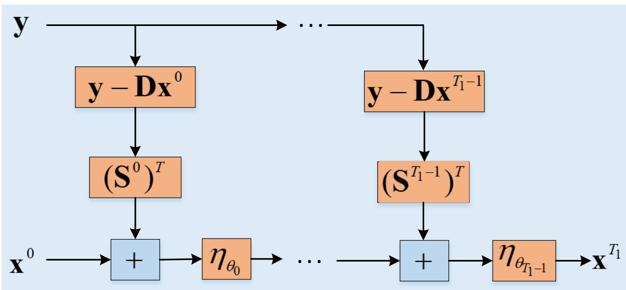

Fig. 3. The structure of the LISTA-CP network.

coupling mechanisms to reduce the parameters that need to be learned, making the training process more efficient. To fully leverage the common sparsity structure provided by the frequency-dependent dictionary, the Joint LISTA-CP algorithm is proposed based on the LISTA-CP network. Joint LISTA-CP utilizes the parallel processing of multiple LISTA-CP sub-networks to improve channel estimation accuracy.

A. ISTA and LISTA Based Channel Estimation

The ISTA method is a popular algorithm to address the problem of sparse channel recovery. The core principle of the algorithm is to iteratively approximate the sparse solution through continuous refinements. It has the following iterative steps:

$$\mathbf{x}_m^{t+1} = \eta_{\theta_t}(\mathbf{x}_m^t + \frac{1}{\lambda}\mathbf{D}_m^T(\mathbf{y}_m - \mathbf{D}_m\mathbf{x}_m^t)), \tag{14}$$

where λ denotes the largest eigenvalue of $\mathbf{D}_m^T\mathbf{D}_m$. η_θ represents the soft-thresholding function. The effect of this function on each component can be expressed as:

$$\eta_\theta(\mathbf{x}_m) = \text{sign}(\mathbf{x}_m)\text{max}(0, |\mathbf{x}_m| - \theta), \tag{15}$$

where $\theta = \frac{L}{\lambda}$ denotes the threshold parameter and L is a hyperparameter. In general, for any given dictionary $\mathbf{D}_m$, ISTA recovers a sparse signal $\mathbf{x}_m$ with sublinear convergence in terms of the objective function. LISTA, as presented in [39], is an accelerated version of ISTA, aiming to reduce the number of iterations and improve solution efficiency through machine learning methods. The LISTA algorithm improves the performance of the ISTA algorithm by using learned weights and threshold. The recursive expression for each layer of LISTA is:

$$\mathbf{x}_m^{t+1} = \eta_{\theta_t}(\mathbf{S}_1^t\mathbf{y} + \mathbf{S}_2^t\mathbf{x}_m^t), \tag{16}$$

where $\mathbf{S}_1 = \frac{1}{\lambda}\mathbf{D}_m^T$, $\mathbf{S}_2 = \mathbf{I} - \frac{1}{\lambda}\mathbf{D}_m^T\mathbf{D}_m$. Unlike ISTA, which has no learnable parameters except for the parameter λ, LISTA is regarded as a neural network with a specific structure. It is trained by applying stochastic gradient descent on the given training dataset $\{\mathbf{y}_m, \mathbf{x}_m^*\}$. The parameters $\mathbf{\Theta} = \{(\mathbf{S}_1^t, \mathbf{S}_2^t, \theta^t)\}_{t=0}^{T_1-1}$ are subject to learning. The training is formulated as

$$\min_{\mathbf{\Theta}} \ \mathbb{E}\{\left\|\mathbf{x}_m^{T_1}(\mathbf{y}_m, \mathbf{x}_m^0, \mathbf{\Theta}) - \mathbf{x}_m^*\right\|_2^2\}, \tag{17}$$

where $\mathbf{x}_m^{T_1}(\cdot)$ denotes the output of the LISTA network after T_1 unfolding layers. Some empirical results indicate that during the training of a LISTA network, the number of layers is typically set between 10 and 20. Fig. 2 illustrates the structure of the LISTA network.

B. LISTA-CP Based Channel Estimation

Although LISTA achieves faster convergence and improved performance compared to ISTA, it involves a large number of learnable parameters. In particular, it requires the learning of two weight matrices with high dimensionality, which increases the difficulty of network training. To solve this problem, [39] proposed LISTA-CP, which proved a necessary condition for LISTA convergence, i.e., the partial weight coupling relationship is satisfied between the parameter matrices $\mathbf{S}_1, \mathbf{S}_2$:

$$\mathbf{S}_2^t = \mathbf{I} - \mathbf{S}_1^t\mathbf{D}_m. \tag{18}$$

Therefore, using the above form of partial weight coupling for the network and setting $\mathbf{S}^t = (\mathbf{S}_1^t)^T$, the simplified deep unfolding network structure for the T-th layer is obtained as

$$\mathbf{x}_m^{t+1} = \eta_{\theta_t}(\mathbf{x}_m^t + (\mathbf{S}^t)^T(\mathbf{y}_m - \mathbf{D}_m\mathbf{x}_m^t)), \tag{19}$$

where $\{\mathbf{S}, \theta\}_{t=0}^{T_1}$ are the learnable parameters. The architecture of the LISTA-CP network is illustrated in Fig. 3.

The LISTA-CP algorithm preserves the fast convergence property of LISTA, while significantly reducing the number of learnable parameters through parameter coupling.

C. Joint LISTA-CP Approach Based Channel Estimation

The aforementioned methods are all channel estimation techniques based on the sparsity of channel vectors in the polar domain. Considering that the observations across the M subcarriers share a common sparse structure, we can exploit this property to further improve the accuracy of channel estimation. Based on (13), we can reformulate the optimization problem into

$$\begin{aligned} &\min_{\mathbf{X}} \sum_{g=1}^{2G} \|\mathbf{X}[g,:]\|_2 \\ &\text{s.t.} \ \sum_{m=1}^{M} \|\mathbf{y}_m - \mathbf{D}_m\mathbf{x}_m\|_2 \leq \varepsilon, \end{aligned} \tag{20}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M] \in \mathbb{R}^{2G\times M}$ is the matrix formed by channel vectors from M subcarriers, and $\mathbf{X}[g,:]$ represents the g-th row of the matrix $\mathbf{X}$. $\varepsilon > 0$ is the allowable error threshold. As we employ frequency-dependent dictionaries to mitigate the beam split effects, this results in distinct sensing matrices for each subcarrier channel vector. Therefore, existing sparse channel recovery methods based on the multiple measurement vector (MMV) problem with a shared sensing matrix cannot be directly applied to solve (20).

Here, we modify the ISTA to jointly estimate sparse channels of different subcarriers. During each iteration, the modified ISTA jointly estimates M column vectors $\mathbf{x}_m$ using a multidimensional shrinkage thresholding function to enforce the common sparse structure. Each iteration of the modified ISTA is given by

$$\begin{aligned}\mathbf{X}^{t+1} =&[\mathbf{x}_1^{t+1}, \cdots, \mathbf{x}_m^{t+1}]\\ =&\widetilde{\eta}_{\theta_t}\big([\mathbf{x}_1^t - (\mathbf{D}_1)^T(\mathbf{y}_1 - \mathbf{D}_1\mathbf{x}_1^t),\\ &\dots, \mathbf{x}_M^t - (\mathbf{D}_M)^T(\mathbf{y}_M - \mathbf{D}_M\mathbf{x}_M^t)]\big),\end{aligned} \tag{21}$$

where θ_t represents the thresholding parameter at the t-th layer,and $\widetilde{\eta}_\theta$ denotes the multidimensional shrinkage thresholding function, defined as

$$\widetilde{\eta}_\theta(\mathbf{X}(g,:)) = \max\left\{0, \frac{\|\mathbf{X}(g,:)\|_2 - \theta}{\|\mathbf{X}(g,:)\|_2}\right\}\mathbf{X}(g,:), \tag{22}$$

where $\mathbf{X}(g,:)$ denotes the g-th row of $\mathbf{X}$. The main function of the multidimensional shrinkage thresholding function is to shrink the signal's multidimensional structure, removing the unimportant parts and achieving sparsification of the signal. To further improve the performance of channel estimation, we extend the modified ISTA method to the learning-based Joint LISTA. The recursive expression for each iteration of Joint LISTA is given by:

$$\begin{aligned}\mathbf{X}^{t+1} =&[\mathbf{x}_1^{t+1}, \cdots, \mathbf{x}_M^{t+1}]\\ =&\widetilde{\eta}_{\theta_t}\big([\mathbf{S}_{11}^t\mathbf{y}_1 - \mathbf{S}_{21}^t\mathbf{x}_1^t, \dots, \mathbf{S}_{1M}^t\mathbf{y}_M - \mathbf{S}_{2M}^t\mathbf{x}_M^t]\big),\end{aligned} \tag{23}$$

where the learnig parameters of Joint LISTA include $\mathbf{\Theta} = \{\mathbf{S}_{11}^t, \mathbf{S}_{21}^t, \dots, \mathbf{S}_{1M}^t, \mathbf{S}_{2M}^t, \theta^t\}_{t=0}^{T_1-1}$.The dimension of $\mathbf{S}_{1M}$ is $2G \times 2PN_{RF}$ and the dimension of $\mathbf{S}_{2M}$ is $2G \times 2G$. These large-dimensional learning parameters total $2M$, resulting in significant storage and computational overhead during network training. To address this challenge, we propose a Joint LISTA-CP algorithm with fewer learnable parameters. The Joint LISTA-CP network enhances the accuracy of sparse channel estimation by simultaneously estimating multiple vectors sharing a common sparse structure. The Joint LISTA-CP algorithm is described as follows.

*Definition 1 (**Joint LISTA-CP**):* The t-th layer of the Joint LISTA-CP network is characterized by the following iterative process:

$$\begin{aligned}\mathbf{X}^{t+1} =&[\mathbf{x}_1^{t+1}, \cdots, \mathbf{x}_M^{t+1}]\\ =&\widetilde{\eta}_{\theta_t}\big([\mathbf{x}_1^t - (\mathbf{S}_1^t)^T(\mathbf{y}_1 - \mathbf{D}_1\mathbf{x}_1^t),\\ &\dots, \mathbf{x}_M^t - (\mathbf{S}_M^t)^T(\mathbf{y}_M - \mathbf{D}_M\mathbf{x}_M^t)]\big),\end{aligned} \tag{24}$$

where the learning parameters of Joint LISTA-CP include $\mathbf{\Theta} = \{\mathbf{S}_1^t, \dots, \mathbf{S}_M^t, \theta^t\}_{t=0}^{T_1-1}$.

As illustrated in Fig. 4, the proposed Joint LISTA-CP consists of M subnetworks. Each subnetwork employs a LISTA-CP network with fewer parameters. Joint LISTA-CP network exploits the property that the sparse channels of different subcarriers have common sparse support and applies a multidimensional shrinkage thresholding operator to guarantee the group-row sparsity property.

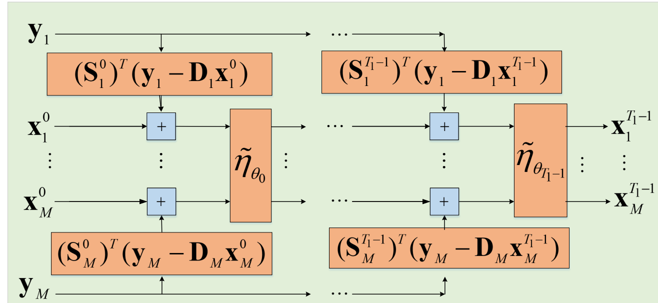

Fig. 4. The structure of the proposed Joint LISTA-CP network.

The Joint LISTA-CP is trained using the Adam optimizer to reduce the MSE loss function, which is defined as

$$\begin{aligned}\mathcal{L}(\mathbf{\Theta}) &= \sum_{u=1}^{U}(\|\hat{\mathbf{X}}^u - \mathbf{X}^{*u}\|_F^2)\\ &= \sum_{u=1}^{U}\sum_{m=1}^{M}(\|\hat{\mathbf{x}}_m^u - \mathbf{x}_m^{*u}\|_2^2)\\ &= \sum_{u=1}^{U}\sum_{m=1}^{M}(\|r(\mathbf{y}_m^u, \mathbf{\Theta}_m) - \mathbf{x}_m^{*u}\|_2^2),\end{aligned} \tag{25}$$

where U denotes the number of training data in each batch set. $\mathbf{\Theta}_m$ represents the trainable parameter of the m-th subnetwork. $r(\cdot)$ denotes the neural network (NN).

IV. THEORETICAL ANALYSIS

In the preceding section, we introduced a deep unfolding-based framework for estimating near-field wideband channels. This approach reduces the number of parameters that the network needs to learn, thereby enhancing the effectiveness of sparse channel estimation. Unlike conventional NNs, which are often viewed as inscrutable black boxes, this section examines the convergence properties of the proposed Joint LISTA-CP. With the proposed network adopting the ISTA structure, its interpretability can be analyzed from an optimization standpoint.

For theoretical analysis, we assume the matrix $\mathbf{X}^*$ and noise $\mathbf{Z}$ belong to the set:

$$\begin{aligned}\mathcal{X}(\beta, s, \sigma) =&\{(\mathbf{X}^*, \mathbf{Z})|\|\mathbf{X}^*[i,:]\|_2 \le \beta, \forall i,\\ &\|\mathbf{X}^*\|_{row,0} \le s, \|\mathbf{Z}\|_F \le \sigma\}.\end{aligned} \tag{26}$$

In other words, $\mathbf{X}^*$ and $\mathbf{Z}$ are bounded, and $\mathbf{X}^*$ contains no more than s non-zero rows.

Theorem 1: **[Convergence Rate of Joint LISTA-CP]** Given $\{\mathbf{S}_m\}_{m=1}^M$ and $\{\theta_t\}_{t=1}^{T_1}$, let $\{\mathbf{X}_t\}_{t=1}^{T_1}$ be generated by Joint LISTA-CP with layers in (24). The initial point is set as $\mathbf{X}_0 = 0$. If s is sufficiently small, then $(\mathbf{X}^*, \mathbf{Z}) \in \mathcal{X}(\beta, s, \sigma)$, and the error bound is given by:

$$\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_F \le s\beta\exp(-at) + A\sigma, \tag{27}$$

where a and A are positive constants, and their specific expressions will be provided in the subsequent proof. For the noiseless case, i.e., $\sigma = 0$, it can be observed that as the number of layers approaches infinity, the error bound decreases to 0 at a linear rate.

The steps of proving Theorem 1 include defining "good" learned parameters, establishing an error bound for one sample, and extending this bound to the entire dataset. The detailed proof process is presented as follows.

A. "Good" Learned Parameters

In this subsection, we define the "good" learned parameters in the Joint LISTA-CP algorithm that ensure a linear convergence rate.

To begin, it is essential to present some basic definitions and a lemma.

Definition 2: The mutual coherence μ_m of $\mathbf{D}_m \in \mathbb{R}^{2PN_{RF}\times 2G}$ can be described as

$$\mu_m(\mathbf{D}_m) = \max_{\substack{i\neq j\\ 1\leq i,j\leq 2G}} |\mathbf{D}_m[:,i]^T\mathbf{D}_m[:,j]|, \tag{28}$$

where $\mathbf{D}_m$ denotes the sensing matrix corresponding to the m-th subnetwork, and each column of $\mathbf{D}_m$ is normalized.

The generalized mutual coherence of $\mathbf{D}_m$, with its columns normalized, is defined as

$$\begin{aligned}&\widetilde{\mu}_m(\mathbf{D}_m)\\ &= \inf_{\substack{\mathbf{W}_m\in\mathbb{R}^{2PN_{RF}\times 2G}\\ (\mathbf{W}_m[:,i])^T\mathbf{D}_m[:,j]=1}} \left\{ \max_{\substack{i\neq j\\ 1\leq i,j\leq 2G}} |\mathbf{W}_m[:,i]^T\mathbf{D}_m[:,j]| \right\}.\end{aligned} \tag{29}$$

Lemma 1: There exists a matrix $\widetilde{\mathbf{W}}_m \in \mathbb{R}^{2PN_{RF}\times 2G}$ satisfying the infimum given in (29) and

$$\begin{aligned}&(\widetilde{\mathbf{W}}_m[:,i])^T\mathbf{D}_m[:,i] = 1,\\ &\max_{\substack{i\neq j\\ 1\leq i,j\leq 2G}} |\widetilde{\mathbf{W}}_m[:,i]^T\mathbf{D}_m[:,j]| = \tilde{\mu}_m.\end{aligned} \tag{30}$$

The proof of this lemma follows the process outlined in [39].

Definition 3: For the m-th subnetwork, given $\mathbf{D}_m \in \mathbb{R}^{2PN_{RF}\times 2G}$, the "good" weight matrix satisfies the following conditions:

$$\begin{aligned}\mathcal{W}_m(\mathbf{D}_m) = \mathop{\arg\min}_{\mathbf{W}_m\in\mathbb{R}^{2PN_{RF}\times 2G}} \Big\{ &\max_{1\leq i,j\leq 2G} |\mathbf{W}_m[i,j]| :\\ &\max_{\substack{i\neq j\\ 1\leq i,j\leq 2G}} |\mathbf{W}_m[:,i]^T\mathbf{D}_m[:,j]| = \tilde{\mu}_m\\ &(\mathbf{W}_m[:,i])^T\mathbf{D}_m[:,i] = 1\Big\}.\end{aligned} \tag{31}$$

With definitions in (29) and (31), the "good" parameters of the Joint LISTA-CP satisfy

$$\begin{aligned}&\mathbf{W}_m^t \in \mathcal{W}_m(\mathbf{D}_m),\\ &\theta^t = \widetilde{\mu}_{min} \sup_{(\mathbf{X}^*,\mathbf{Z})} \|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} + \sigma C_W,\end{aligned} \tag{32}$$

where $\widetilde{\mu}_{min}$ denotes the minimum of all the M generalized mutual coherence $\widetilde{\mu}_m$, the best weight matrix of the m-th subnetwork corresponding to $\widetilde{\mu}_{min}$ is $\mathbf{W}$, and $C_W = \max_{t\geq 0}\|\mathbf{W}\|_{2,1}$. In the next section, we use these good parameters to prove the convergence rate of the Joint LISTA-CP.

B. Error Bound for One Sample

To obtain information about group-row sparsity, we first introduce a new function $\gamma(\mathbf{X})$:

$$\gamma(\mathbf{X}) = \left[\|\mathbf{X}[1,:]\|_2, \|\mathbf{X}[2,:]\|_2, \cdots, \|\mathbf{X}[2G,:]\|_2\right]^T. \tag{33}$$

For a vector $\mathbf{a} = [a_1, a_2, \ldots, a_{2G}]^T \in \mathbb{R}^{2G}$, we have $\gamma(\mathbf{a}[i]) = |a_i|$, $i \in [2G]$. Using this function, $\text{supp}(\gamma(\mathbf{X}))$ can reveal the sparsity of a matrix $\mathbf{X}$. To analyze the error of Joint LISTA-CP, we first demonstrate that none of the rows in $\mathbf{X}^t$ for any t false positives.

Lemma 2: With $(\mathbf{X}^*, \mathbf{Z}) \in \mathcal{X}(\beta, s, \sigma)$, as long as good parameters are used, we have

$$\text{supp}(\gamma(\mathbf{X}^t)) \subset \text{supp}(\gamma(\mathbf{X}^*)),\ \forall t. \tag{34}$$

The proof can be found in Appendix A. The no false positives indicates this property in the Joint LISTA-CP:

$$\mathbf{X}^t[i,:] = 0,\ \forall i \notin S,\ \forall t, \tag{35}$$

where $S = \text{supp}(\gamma(\mathbf{X}^*))$. This indicates that for elements not included in the set S, the recovery process is perfect with no error.

Next, we focus on the components in S. For each $i \in S$, the Joint LISTA-CP in (24) gives

$$\begin{aligned}\mathbf{X}^{t+1}[i,:] &= \widetilde{\eta}_{\theta_t}(\mathbf{X}^t[i,:] - \mathbf{a}_1 + \mathbf{a}_2)\\ &\in \mathbf{X}^t[i,:] - \mathbf{a}_1 + \mathbf{a}_2 - \theta_t\partial\|\mathbf{X}^{t+1}[i,:]\|_2,\end{aligned} \tag{36}$$

where $\mathbf{a}_1 = [(\mathbf{W}_1^t[:,i])^T\mathbf{D}_1[:,S](\mathbf{x}_1^t[S] - \mathbf{x}_1^*[S]), \ldots, (\mathbf{W}_M^t[:,i])^T\mathbf{D}_M[:,S](\mathbf{x}_M^t[S] - \mathbf{x}_M^*[S])]$, and $\mathbf{a}_2 = [(\mathbf{W}_1^t[:,i])^T\mathbf{z}_1, \ldots, (\mathbf{W}_M^t[:,i])^T\mathbf{z}_M]$. $\partial\|\cdot\|_2$ is the sub-gradient of $\mathbf{X}[i,:]$:

$$\partial\|\mathbf{X}[i,:]\|_2 = \begin{cases} \{\frac{\mathbf{X}[i,:]}{\|\mathbf{X}[i,:]\|_2}\}, & if\ \mathbf{X}[i,:] \neq \mathbf{0}\\ \{\mathbf{h} \in \mathbb{R}^M \mid \|\mathbf{h}\|_2 \leq 1\}, & if\ \mathbf{X}[i,:] = \mathbf{0}\end{cases} \tag{37}$$

Since $(\mathbf{W}^k[:,i])^T\mathbf{D}[:,i] = 1$, (36) can be written as

$$\mathbf{X}^{t+1}[i,:] - \mathbf{X}^*[i,:] \in -\widetilde{\mathbf{a}}_1 + \mathbf{a}_2 - \theta_t\partial\|\mathbf{X}^{t+1}[i,:]\|_2, \tag{38}$$

where $\widetilde{\mathbf{a}}_1 = [\sum_{j\in S, j\neq i}(\mathbf{W}_1^t[:,i])^T\mathbf{D}_1[:,j](\mathbf{x}_1^t[j] - \mathbf{x}_1^*[j]), \cdots, \sum_{j\in S, j\neq i}(\mathbf{W}_M^t[:,i])^T\mathbf{D}_M[:,j](\mathbf{x}_M^t[j] - \mathbf{x}_M^*[j])]$.

According to (37), we can get $\|\partial(\|\mathbf{X}^{t+1}[i,:]\|_2)\|_2 \leq 1$. By taking the norm of both sides of (38), we can obtain

$$\begin{aligned}&\|\mathbf{X}^{t+1}[i,:] - \mathbf{X}^*[i,:]\|_2\\ &\overset{(a)}{\leq} [\sum_{j\in S, j\neq i} |(\mathbf{W}_1^t[:,i])^T\mathbf{D}_1[:,j](\mathbf{x}_1^t[j] - \mathbf{x}_1^*[j])|, \cdots,\\ &\quad \sum_{j\in S, j\neq i} |(\mathbf{W}_M^t[:,i])^T\mathbf{D}_M[:,j](\mathbf{x}_M^t[j] - \mathbf{x}_M^*[j])|]\\ &\quad + \theta_t + \|\mathbf{a}_2\|_2\\ &\overset{(b)}{\leq} \mu_{min} \sum_{j\in S, j\neq i} \|\mathbf{X}^t[j,:] - \mathbf{X}^*[j,:]\|_2 + \theta_t + \|\mathbf{W}^t[:,i]\|_2\|\mathbf{Z}\|_F,\end{aligned} \tag{39}$$

where (a) is derived from the triangle inequality. The first term in (b) uses the smallest generalized mutual coherence μ_{min} and the optimal parameters given in (32). The third term in

(b) is obtained through $\|(\mathbf{W}^t[:,i])^T\mathbf{Z}\|_2 \leq \|\mathbf{W}^t[:,i]\|_2\|\mathbf{Z}\|_F$. Since there are no false positives, relationship $\|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} = \|\mathbf{X}^t[S,:] - \mathbf{X}^*[S,:]\|_{2,1}$ holds. Based on (39), we can get

$$\begin{aligned}
&\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_{2,1} \\
=&\sum_{i\in S}\|\mathbf{X}^{t+1}[i,:] - \mathbf{X}^*[i,:]\|_2 \\
=&\mu_{min}(|S|-1)\sum_{i\in S}\|\mathbf{X}^{t}[i,:] - \mathbf{X}^*[i,:]\|_2 + |S|\theta^t + \sigma C_W \\
\leq&\mu_{min}(|S|-1)\|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} + |S|\theta^t + \sigma C_W.
\end{aligned} \tag{40}$$

Now, we derive the recovery error of $l_{2,1}$-norm for one sample data.

C. Error Bound for Complete Support Set

To obtain the error bound for the entire support set of the sparse signal, we first establish a bound on the number of nonzero support indices, denoted as $|S| = \text{supp}(\gamma(\mathbf{X}^*)) \leq s$. Taking the upper bound on both sides of (40) gives:

$$\begin{aligned}
&\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_{2,1} \\
\leq&\mu_{min}(s-1)\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} + s\theta^t + \sigma C_W
\end{aligned} \tag{41}$$

According to (32), the "good" parameters satisfy $\theta^t = \widetilde{\mu}_{min}\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} + \sigma C_W$, then we have

$$\begin{aligned}
&\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_{2,1} \\
\leq&(2\mu_{min}s - \mu_{min})^{t+1}s\beta + \frac{(s+1)C_W}{1+\mu_{min} - 2\mu_{min}s}\sigma
\end{aligned} \tag{42}$$

By letting $a = -\log(2\mu_{min}s - \mu_{min})$, $A = \frac{(s+1)C_W}{1+\mu_{min}-2\mu_{min}s}$, we can obtain

$$\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_{2,1} \leq s\beta\exp(-a(t+1)) + A\sigma. \tag{43}$$

Since $\|\mathbf{X}\|_F \leq \|\mathbf{X}\|_{2,1}$ for arbitrary $\mathbf{X}$, we have the upper bound in the Frobenius norm as

$$\begin{aligned}
\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_F \leq& \sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^{t+1} - \mathbf{X}^*\|_{2,1} \\
\leq& s\beta\exp(-at) + A\sigma.
\end{aligned} \tag{44}$$

As long as $s < (1+1/\mu_{min})/2$, $a = -log(2\mu_{min}s - \mu_{min}) > 0$, the error bound holds for all $\mathbf{X}^*$. This completes the proof of the error bound in (27). The procedure for proving the error bound for a single LISTA-CP network is provided in Appendix B.

Remark: The analysis shows that the error bound of Joint LISTA-CP algorithm is smaller than that of the single LISTA-CP network, indicating that the Joint LISTA-CP achieves improved recovery performance by leveraging the group-row sparsity property.

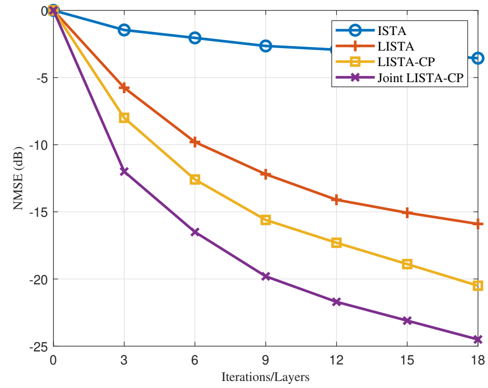

Fig. 5. The NMSE performance of different methods in noiseless cases.

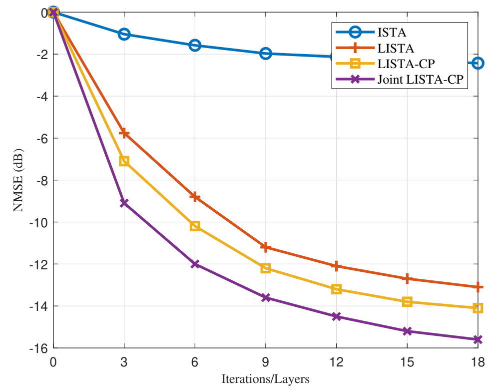

Fig. 6. The NMSE performance of different methods under SNR=15dB.

V. Numerical Results

In this section, we design a series of experiments to illustrate the efficiency of the proposed Joint LISTA-CP method in general scenarios and XL-MIMO near-field wideband systems. In general cases, we use the randomly generated Gaussian matrix as the sensing matrix. We demonstrate the advantages of Joint LISTA-CP method in terms of convergence rate and accuracy when solving sparse reconstruction problems in both noiseless and noisy conditions. For XL-MIMO wideband system channel estimation, we demonstrate the performance of the Joint LISTA-CP method under varying conditions of signal-to-noise ratio (SNR), system bandwidth, and number of antennas and pilots. These experiments are designed to demonstrate the performance enhancements achieved through joint estimation across multiple LISTA-CP networks. Additionally, in this paper, the training and testing datasets used for network training are obtained through simulation.

A. Simulation Experiments

In this experiment, we set that $\mathbf{D}_m$ follows the Gaussian distribution, and its column vectors have unit norms. The nonzeros indexes in the sparse vector $\mathbf{x}$ are distributed according to a Bernoulli distribution, while the values follow a Gaussian distribution. We set the number of subnetworks to 16, and the dimension of the sparse signal $\mathbf{X}$ to 1000×16. For training the network, we use online training with the batch size of 128. The learning rate is adjusted in stages, starting with an initial rate of 0.001. If the loss function fails to decrease after 200 iterations, the learning rate is reduced by a specific percentage. This adjustment is made three times, with the learning rate being scaled by 0.5, 0.1, and 0.01, respectively. We utilize the NMSE as the evaluation metric to measure the performance of different methods, and it is described as follows

$$\mathrm{NMSE}(\hat{\mathbf{X}}, \mathbf{X})(\mathrm{dB}) = 10 \log_{10} \left(\mathbb{E} \left(\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{F}}^2}{\|\mathbf{X}\|_{\mathrm{F}}^2} \right) \right). \quad (45)$$

To demonstrate the performance advantages of the Joint LISTA-CP algorithm in terms of reconstruction accuracy and computation complexity, we compare the NMSE of the reconstructed sparse signal $\mathbf{X}$ obtained by various methods with varying numbers of iterations. As each layer of the Joint LISTA-CP network corresponds to one iteration of the traditional algorithm, the change of NMSE with network depth can be used to demonstrate the convergence property [40]. As shown in Fig. 5 and Fig. 6, the Joint LISTA-CP demonstrates faster convergence and higher accuracy compared to the other methods in both noiseless and noisy scenarios. Since the proposed network has the same iteration complexity as the original algorithm ISTA but requires significantly fewer iterations to attain the same accuracy, its overall complexity is lower. Compared to the LISTA-CP network, the Joint LISTA-CP method achieves performance improvements through joint estimation across multiple LISTA-CP networks. The experimental results using random matrix demonstrate the potential of the proposed Joint LISTA-CP in other applications.

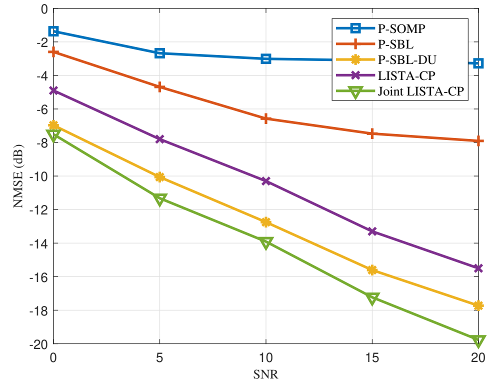

Fig. 7. The NMSE performance versus different SNRs with $P = 64$ and $N_{BS} = 256$.

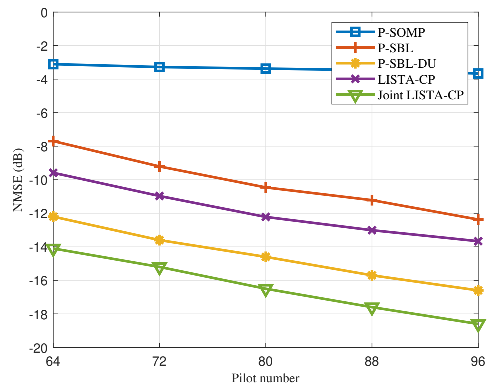

Fig. 8. The NMSE performance against different numbers of pilots with SNR $= 10dB$ and $N_{BS} = 256$.

B. Wideband Channel Estimation for Near-field

In the experiment, the number of BS antennas is set to 256 and the number of RF chains is set to 8. The carrier frequency is set to 100GHz, with a bandwidth of 10GHz. The number of clusters is 3, and each cluster contains 10 subpaths. Assume that the distances and angles of subpaths within a cluster follow Laplacian distributions, with their means following uniform distributions within specified ranges. Specifically, the mean of the angles ranges from $0°$ to $360°$, while the mean of the distances ranges from 5 meters to 30 meters, with standard deviations of $4°$ and 1 meter, respectively. The parameter $\beta_\triangle$ of the polar dictionary is 0.8. The number of angle-domain samples is 512 and distance-domain samples is 4. Thus, the polar-domain dictionary $\mathbf{A}$ is constructed with 2048 sampled grids. In addition, the environment for the experiment is Python 3.6 and Tensorflow 1.12. The experiments are conducted using the GTX1050 GPU.

The Adam optimizer is used to train the Joint LISTA-CP network to minimize the loss. The batch size that we employ for network training is 128. In addition, the learning rate is adjusted at different stages of the model training process to better fit the data and improve the learning effect of the model. The learning rate is adjusted in the same way as explained in the previous section. To evaluate the performance, we compare the proposed Joint LISTA-CP method with existing traditional CS-based methods and DL-based methods.

- P-SOMP [9]: The polar domain-based SOMP (P-SOMP) method is used to solve the sparse channel recovery problem for near-field multi-measurement vectors. In each iteration, the channel estimate is refined by updating the residual.
- P-SBL [41]: It is a sparse signal recovery method based on sparse Bayesian learning (SBL) in the polar domain (P-SBL). The core of the P-SBL method is to model the sparsity of each signal element, assuming that each element follows some Gaussian distribution. Through the process of Bayesian inference, the sparsity of the signal can be modeled and estimated.
- P-SBL-DU [37]: It is a deep unfolding version of the P-SBL (P-SBL-DU) method. The key idea of P-SBL-DU is to unfold the expectation-maximization (EM) iterative process, mapping each iteration to an SBL layer. In each SBL layer, the expectation step remains unchanged, while the original maximization step is implemented using a deep neural network (DNN).
- LISTA-CP [42]: It is a weight-coupled version based on LISTA, which has been described in detail earlier.
- Joint LISTA-CP: The method proposed in this paper primarily involves jointly using multiple LISTA-CP subnetworks to achieve near-field wideband channel estimation.

The computational complexity of the above algorithms is analyzed as follows. The computational complexity of Joint LISTA-CP mainly arises from the matrix multiplication operations, with a complexity of $O(MGPN_{RF})$ per iteration. The computational complexity of P-SBL and LISTA-CP algorithms per iteration is the same as that of Joint LISTA-CP. The com-

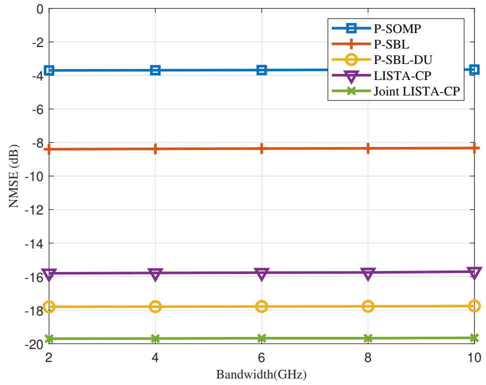

Fig. 9. The NMSE performance against different system bandwidths with SNR $= 20dB$ and $N_{BS} = 256$.

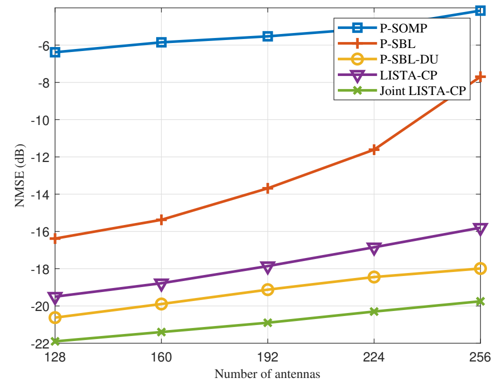

Fig. 10. The NMSE performance against different numbers of antennas with SNR $= 20dB$ and $P = 64$.

putational complexity of P-SOMP is $O(N_c N_p MGPN_{RF})$, where $N_c N_p$ is the total number of paths. Additionally, the computational complexity of P-SBL-DU is $O(MGPN_{RF} + 800MG)$. Although P-SOMP, LISTA-CP, and Joint LISTA-CP have the same computational complexity, Joint LISTA-CP exhibits faster convergence, thus effectively reducing the computational cost. The simulation results of Joint LISTA-CP and the comparison algorithms are analyzed below.

In Fig. 7, we compare the performance of near-field channel estimation under different SNR, with the number of pilots is 64. As observed in Fig. 7, the NMSE values of all methods decrease as the SNR increases. However, the rate of decrease and the final NMSE values achieved by different algorithms vary. First, the NMSE of P-SOMP decreases slightly as the SNR increases. By comparing the P-SBL and P-SBL-DU algorithms, it can be observed that P-SBL-DU achieves better performance. This performance improvement comes from the effective utilization of the inherent characteristics of the channel by deep neural networks. The Joint LISTA-CP algorithm shows the best performance at all SNRs. Specifically, at SNR = 20 dB, the performance of Joint LISTA-CP is 4 dB better than that of LISTA-CP. This performance gain comes from the joint learning of multiple networks.

Fig. 8 shows the NMSE performance of various algorithms with different numbers of pilots. As the number of pilots increases, the NMSE values of all algorithms show a downward trend. This trend indicates that when the system receives more pilot information, the error in channel estimation decreases, which is the expected result. The NMSE of P-SOMP decreases at a slower rate as the number of pilots increases, and even when the number of pilots reaches 96, its performance is still significantly lower than the other algorithms. This indicates that P-SOMP is not sensitive to the increase in the number of pilots. Across all the numbers of pilots, P-SBL-DU outperforms P-SBL. This indicates that P-SBL-DU is more effective in improving channel estimation accuracy when handling more pilot data. Joint LISTA-CP achieves the best NMSE performance under different pilot lengths. For example, to reach NMSE = -14dB, P-SBL-DU requires 76 pilots, while Joint LISTA-CP only needs 64 pilots. This indicates that the Joint LISTA-CP algorithm can effectively save pilot resources.

In Fig. 9, the NMSE performance of the comparison methods under different system bandwidths is presented. Notably, the NMSE of all algorithms in the figure remains nearly unchanged across different bandwidth conditions. This is due to the use of a frequency-dependent polar-domain dictionary, which allows the near-field channel to maintain good channel sparsity and sparse structure in the polar domain, effectively overcoming the beam split effect. Additionally, the Joint LISTA-CP method exhibits better NMSE performance under all considered system bandwidth conditions, indicating its robustness and accuracy in near-field wideband channel estimation.

Fig. 10 illustrates the NMSE performance of different methods under different numbers of antennas. As the number of antennas increases, the NMSE values for all algorithms also increase. This is because channel estimation becomes more challenging as the number of antennas increases, resulting in a decrease in performance of all methods. The figure shows that the P-SBL algorithm performs effectively in channel estimation when the number of antennas is limited. However, as the number of antennas increases, the channel estimation performance decreases dramatically. The reason is that increasing the number of antennas leads to greater correlation among the sparse dictionary atoms, which negatively impacts the channel estimation result of the P-SBL method. However, the P-SBL-DU algorithm improves estimation performance by leveraging the inherent characteristics of the near-field channel through a DNN network. The Joint LISTA-CP shows the best performance at all antenna numbers and has the least increase in NMSE as the number of antennas increases, which indicates its good robustness and stability.

C. Other Simulation Results

To further verify the accuracy and adaptability of the proposed Joint LISTA-CP method, we refer to the channel model used in [29] and conduct corresponding simulations using this model.

First, we introduce the channel model used in [29]. The narrowband channel considered in [29] is extended to a wideband scenario, and the corresponding channel model is as follows:

$$\mathbf{h}_m^{'} = \sqrt{\frac{N_{BS}}{L}} \sum_{l=1}^{L} \frac{\lambda_m \sqrt{\rho}}{(4\pi)^{3/2} r_l} e^{-j2\pi f_m \tau_l} \mathbf{b}(\phi_l, r_l, f_m), \quad (46)$$

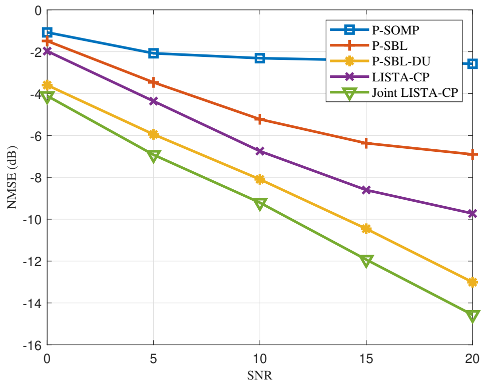

Fig. 11. The NMSE performance under different SNRs for channel model (46), with SNR $= 20dB$ and $P = 64$.

where λ_m represents the wavelength of the m-th subcarrier. ρ represents the path loss coefficient, L denotes the number of paths. r_l represents the distance from the scatterer to the BS antenna for the l-th path. f_m represents the frequency of the m-th subcarrier. τ_l denotes the delay of the l-th subpath. $\mathbf{b}(\phi_l, r_l, f_m)$ is the array response vector. As introduced earlier, it is not be repeated here.

In the simulation experiment, the BS is equipped with 256 antennas and 8 RF chains. The system bandwidth is 10 GHz, and the carrier frequency is set to 100 GHz. The number of paths is assumed to be 6, with 16 subcarriers. The polar domain dictionary has 512 samples in the angle domain and 4 samples in the distance domain. The training setup for the Joint LISTA-CP network remains consistent with the previous section. The following are the simulation results based on the channel model (46).

Fig. 11 presents the NMSE performance of different methods under varying SNR conditions based on channel model (46). It can be observed that the proposed Joint LISTA-CP method still achieves favorable near-field wideband channel estimation performance under channel model (46). At high SNRs, the Joint LISTA-CP achieves lower NMSE compared to other methods. This indicates that the proposed method exhibits good robustness across different channel models and achieves superior channel estimation performance.

VI. Conclusions

In this paper, we propose a deep learning-based algorithm for wideband near-field channel estimation in XL-MIMO systems. To address the beam split effect in wideband near-field scenarios, we employ a frequency-dependent polar-domain dictionary, ensuring that the corresponding channel sparse support set remains consistent across different subcarriers. Then, we propose a channel estimation method that joint multiple LISTA-CP networks, namely Joint LISTA-CP. The proposed method uses a multidimensional shrinkage thresholding operator to ensure the common sparsity of different subcarrier channels. Additionally, we rigorously prove that the Joint LISTA-CP network achieves a lower recovery error. Simulation results demonstrate that the Joint LISTA-CP method achieves high channel estimation accuracy in XL-MIMO wideband systems. Generalization is a common issue in DL methods, while the main focus of this paper is the performance of the proposed Joint LISTA-CP method in near-field wideband channel estimation. Generalization is an additional technical aspect that can be further explored and improved in future work.

Appendix A
The Proof of Lemma 2

We demonstrate that the no-false-positives property holds for the Joint LISTA-CP. Specifically, we prove that if (32) is satisfied, then $\mathbf{X}^t[i,:] = 0, \forall i \notin S, \forall t$.

Proof: Let $(\mathbf{X}^*, \mathbf{Z}) \in \mathcal{X}(\beta, s, \sigma)$ and $S = \text{supp}(\gamma(\mathbf{X}^*))$.

(i) When $t = 0$, $\mathbf{X}^0 = 0$.

(ii) Assume $\mathbf{X}^t[i,:] = 0$ for all $i \notin S$. Then by (24) it follows that

$$\begin{aligned}\mathbf{X}^{t+1}[i,:] =& [\mathbf{x}_1^{t+1}[i], \cdots, \mathbf{x}_M^{t+1}[i]]\\ =&\widetilde{\eta}_\theta([\mathbf{x}_1^t[i] - (\mathbf{S}_1[:,i])^T(\mathbf{y}_1 - \mathbf{D}_1\mathbf{x}_1^t)), \cdots,\\ &\mathbf{x}_M^t[i] - (\mathbf{S}_M[:,i])^T(\mathbf{y}_M - \mathbf{D}_M\mathbf{x}_M^t)])\\ =&\widetilde{\eta}_\theta([\mathbf{x}_1^t[i] - (\mathbf{S}_1[:,i])^T\mathbf{D}_1(\mathbf{x}_1^t - \mathbf{x}_1^*)\\ &+ (\mathbf{S}_1[:,i])^T\mathbf{z}_1, \cdots, \mathbf{x}_M^t[i] -\\ &(\mathbf{S}_M[:,i])^T\mathbf{D}_M(\mathbf{x}_M^t - \mathbf{x}_M^*)\\ &+ (\mathbf{S}_M[:,i])^T\mathbf{z}_M]),\end{aligned} \tag{47}$$

for all $i \notin S$. For any given vector $\mathbf{x}_m$, it follows from the triangle inequality that

$$\begin{aligned}&\|-(\mathbf{S}_m[:,i])^T\mathbf{D}_m(\mathbf{x}_m^t - \mathbf{x}_m^*)\|_2\\ \leq&\sum_{j\in S}|\mathbf{S}_m[:,i]^T\mathbf{D}_m[:,j]|\|\mathbf{x}_m^t - \mathbf{x}_m^*\|_2 + \|\mathbf{S}_m[:,i]^T\mathbf{z}_m\|_2\\ \leq&\widetilde{\mu}_{min}\|\mathbf{x}_m^t - \mathbf{x}_m^*\|_2 + C_W\|\mathbf{z}_m\|_2.\end{aligned} \tag{48}$$

Combine M vectors,

$$\begin{aligned}&\mathbf{X}^{t+1}[i,:]\\ =&[\mathbf{x}_1^{t+1}[i], \cdots, \mathbf{x}_M^{t+1}[i]]\\ \leq&\widetilde{\mu}_{min}\|\mathbf{X}_m^t - \mathbf{X}_m^*\|_{2,1} + C_W\|\mathbf{Z}\|_F \leq \theta^t.\end{aligned} \tag{49}$$

The last inequality holds because $\theta^t = \widetilde{\mu}_{min}\sup_{(\mathbf{X}^*,\mathbf{Z})}\|\mathbf{X}^t - \mathbf{X}^*\|_{2,1} + \sigma C_W$. This implies $\mathbf{X}^{t+1}[i,:] = 0, \forall i \notin S$ by the definition of thresholding function η_θ. Using induction, we obtain that $\|\mathbf{X}^t[i,:]\|_2 = 0, \forall i \notin S, \forall t$.

Appendix B
Convergence Rate of LISTA-CP

According to the bounds and sparsities of $\mathbf{X}$, for any $\mathbf{x}_m$, the corresponding bounds and sparsities are denoted by

$$\begin{aligned}\mathcal{X}_1(\frac{\beta}{M}, s, \frac{\sigma}{\sqrt{M}}) =& \{(\mathbf{x}^*, \mathbf{z})| \|\mathbf{x}_i^*\|_2 \leq \frac{\beta}{M}, \forall i,\\ &\|\mathbf{x}^*\|_0 \leq s, \|\mathbf{z}\|_2 \leq \frac{\sigma}{\sqrt{M}}\}.\end{aligned} \tag{50}$$

The "good" weight matrix satisfies (31), the "good" parameter of the LISTA-CP satisfy

$$\begin{aligned}&\mathbf{W}_m^t \in \mathcal{W}_m(\mathbf{D}_m),\\ &(\theta^t)' = \widetilde{\mu}_m \sup_{(\mathbf{x}^*,\mathbf{z})}\|\mathbf{x}^t - \mathbf{x}^*\|_1 + \frac{\sigma}{\sqrt{M}}C_W\end{aligned} \tag{51}$$

For the convergence proof procedure of LISTA-CP sparse vector signal refer to [39]. Here we give the convergence bound for LISTA-CP:

$$\begin{aligned}&\sup_{(\mathbf{x}^*,\mathbf{z})} \|\mathbf{x}^{t+1}-\mathbf{x}^*\|_2\\ \le&(2\tilde{\mu}_m s-\tilde{\mu}_m)^t s\frac{\beta}{M}+\frac{2sC_W\sigma}{(1+\tilde{\mu}_m-2\tilde{\mu}_m s)\sqrt{M}}.\end{aligned} \tag{52}$$

For the M vector signal $\mathbf{x}_m$, the convergence bound for LISTA-CP:

$$\begin{aligned}&\sup_{(\mathbf{X}^*,\mathbf{Z})} \|\mathbf{X}^{t+1}-\mathbf{X}^*\|_{2,1}\\ \le&(2\tilde{\mu}_m s-\tilde{\mu}_m)^t s\beta+\frac{2sC_W\sigma\sqrt{M}}{1+\tilde{\mu}_m-2\tilde{\mu}_m s}\\ =&s\beta exp(-ct)+C\sigma,\end{aligned} \tag{53}$$

where $c=-log(2\tilde{\mu}_m s-\tilde{\mu}_m)$, $C=\frac{2sC_W\sqrt{M}}{1+\tilde{\mu}_m-2\tilde{\mu}_m s}$. Provided that $s<(1+1/\tilde{\mu}_m)/2$, $c=-log(2\tilde{\mu}_m s-\tilde{\mu}_m)>0$, the error bound in (53) holds for all $(\mathbf{X}^*,\mathbf{Z})$.

References

[1] Z. Wang, J. Zhang, H. Du, D. Niyato, S. Cui, B. Ai, M. Debbah, K. B. Letaief, and H. V. Poor, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Communications Surveys & Tutorials*, 2024.

[2] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Communications Magazine*, vol. 61, no. 1, pp. 40–46, 2022.

[3] Y. Lu and L. Dai, "Near-field channel estimation in mixed LoS/NLoS environments for extremely large-scale MIMO systems," *IEEE Transactions on Communications*, vol. 71, no. 6, pp. 3694–3707, 2023.

[4] C. Ruan, Z. Zhang, H. Jiang, J. Dang, L. Wu, and H. Zhang, "Wideband near-field channel covariance estimation for XL-MIMO systems in the face of beam split," *IEEE Transactions on Vehicular Technology*, 2024.

[5] K. Zhi, C. Pan, H. Ren, K. K. Chai, C.-X. Wang, R. Schober, and X. You, "Performance analysis and low-complexity design for XL-MIMO with near-field spatial non-stationarities," *IEEE Journal on Selected Areas in Communications*, 2024.

[6] Y. Chen and L. Dai, "Non-stationary channel estimation for extremely large-scale MIMO," *IEEE Transactions on Wireless Communications*, 2023.

[7] H. Yan, H. Chen, W. Liu, S. Yang, G. Wang, and C. Yuen, "RIS-enabled joint near-field 3D localization and synchronization in siso multipath environments," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2024.

[8] S. Yang, X. Chen, Y. Xiu, W. Lyu, Z. Zhang, and C. Yuen, "Performance bounds for near-field localization with widely-spaced multi-subarray mmWave/THz MIMO," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2024.

[9] Y. Lu and L. Dai, "Near-field channel estimation in mixed LoS/NLoS environments for extremely large-scale MIMO systems," *IEEE Transactions on Communications*, vol. 71, no. 6, pp. 3694–3707, 2023.

[10] M. Cui and L. Dai, "Near-field channel estimation for extremely large-scale MIMO with hybrid precoding," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[11] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 633–637, 2020.

[12] X. Zhang, H. Zhang, and Y. C. Eldar, "Near-field sparse channel representation and estimation in 6G wireless communications," *IEEE Transactions on Communications*, 2023.

[13] H. Lu, Y. Zeng, C. You, Y. Han, J. Zhang, Z. Wang, Z. Dong, S. Jin, C. Wang, T. Jiang *et al.*, "A tutorial on near-field XL-MIMO communications towards 6G," *IEEE Communications Surveys & Tutorials*, 2024.

[14] 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.

[15] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, 2018.

[16] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE transactions on signal processing*, vol. 63, no. 23, pp. 6169–6183, 2015.

[17] 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.

[18] H. Chu, L. Zheng, and X. Wang, "Super-resolution mmwave channel estimation for generalized spatial modulation systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1336–1347, 2019.

[19] B. Qi, W. Wang, and B. Wang, "Off-grid compressive channel estimation for mm-wave massive MIMO with hybrid precoding," *IEEE Communications Letters*, vol. 23, no. 1, pp. 108–111, 2018.

[20] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663–2677, 2022.

[21] X. Wei and L. Dai, "Channel estimation for extremely large-scale massive MIMO: Far-field, near-field, or hybrid-field?" *IEEE Communications Letters*, vol. 26, no. 1, pp. 177–181, 2021.

[22] H. Lei, J. Zhang, Z. Wang, B. Ai, and D. W. K. Ng, "Hybrid-field channel estimation for XL-MIMO systems with stochastic gradient pursuit algorithm," *IEEE Transactions on Signal Processing*, 2024.

[23] S. Yang, W. Lyu, Z. Hu, Z. Zhang, and C. Yuen, "Channel estimation for near-field XL-RIS-Aided mmwave hybrid beamforming architectures," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 8, pp. 11 029–11 034, 2023.

[24] J. Zhang, M. Liu, J. Tang, N. Zhao, D. Niyato, and X. Wang, "Joint design for RIS-aided ISAC via deep unfolding learning," *IEEE Transactions on Cognitive Communications and Networking*, 2024.

[25] Y. Wang, Z. Gao, S. Chen, C. Hu, and D. Zheng, "Deep learning–based channel extrapolation and multiuser beamforming for RIS-aided terahertz massive MIMO systems over hybrid-field channels," *Intelligent Computing*, vol. 3, p. 0065, 2024.

[26] J. Zhang, S. Gong, W. Lu, C. Xing, N. Zhao, D. W. K. Ng, and D. Niyato, "Joint design for STAR-RIS aided ISAC: Decoupling or learning," *IEEE Transactions on Wireless Communications*, 2024.

[27] Y. Jin, J. Zhang, C. Huang, L. Yang, H. Xiao, B. Ai, and Z. Wang, "Multiple residual dense networks for reconfigurable intelligent surfaces cascaded channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2134–2139, 2021.

[28] H. Nayir, E. Karakoca, A. Görçin, and K. Qaraqe, "Hybrid-field channel estimation for massive MIMO systems based on OMP cascaded convolutional autoencoder," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. IEEE, 2022, pp. 1–6.

[29] X. Zhang, Z. Wang, H. Zhang, and L. Yang, "Near-field channel estimation for extremely large-scale array communications: A model-based deep learning approach," *IEEE Communications Letters*, vol. 27, no. 4, pp. 1155–1159, 2023.

[30] J. Xiao, J. Wang, Z. Chen, and G. Huang, "U-MLP-based hybrid-field channel estimation for XL-RIS assisted millimeter-wave MIMO systems," *IEEE Wireless Communications Letters*, vol. 12, no. 6, pp. 1042–1046, 2023.

[31] H. Nayir, E. Karakoca, A. Görçin, and K. Qaraqe, "Channel estimation using RIDNet assisted OMP for hybrid-field THz massive MIMO systems," in *ICC 2023-IEEE International Conference on Communications*. IEEE, 2023, pp. 2625–2630.

[32] S. Yang, C. Xie, W. Lyu, B. Ning, Z. Zhang, and C. Yuen, "Near-field channel estimation for extremely large-scale reconfigurable intelligent surface (XL-RIS)-aided wideband mmwave systems," *IEEE Journal on Selected Areas in Communications*, 2024.

[33] Y. Sun, C. Yang, and M. Peng, "Subarray-based hybrid-field channel estimation for terahertz wideband UM-MIMO systems without prior location knowledge," *IEEE Transactions on Vehicular Technology*, 2023.

[34] Z. Wang, X. Mu, and Y. Liu, "Beamfocusing optimization for near-field wideband multi-user communications," *IEEE Transactions on Communications*, 2024.

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

[36] M. Cui and L. Dai, "Near-field wideband channel estimation for extremely large-scale MIMO," *Science China Information Sciences*, vol. 66, no. 7, p. 172303, 2023.

[37] J. Gao, X. Chen, and G. Y. Li, "Deep unfolding based channel estimation for wideband terahertz near-field massive MIMO systems," *Frontiers of Information Technology & Electronic Engineering*, pp. 1–11, 2024.

[38] K. Wang, Z. Gao, S. Chen, B. Ning, G. Chen, Y. Su, Z. Wang, and H. V. Poor, "Knowledge and data dual-driven channel estimation and feedback for ultra-massive MIMO systems under hybrid field beam squint effect," *IEEE Transactions on Wireless Communications*, 2024.

[39] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[40] J. Liu and X. Chen, "ALISTA: Analytic weights are as good as learned weights in LISTA," in *International Conference on Learning Representations (ICLR)*, 2019.

[41] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse bayesian learning (SBL) based approaches," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1251–1266, 2018.

[42] Y. Shi, H. Choi, Y. Shi, and Y. Zhou, "Algorithm unrolling for massive access via deep neural networks with theoretical guarantee," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 945–959, 2021.