Sparsity-Aware Channel Estimation for mmWave Massive MIMO: A Deep CNN-Based Approach

Sicong Liu[*], **Xiao Huang**

Department of Information & Communication Engineering, School of Informatics, Xiamen University, Xiamen 361005, China

[*] The corresponding author, email: liusc@xmu.edu.cn

Abstract: The deep convolutional neural network (CNN) is exploited in this work to conduct the challenging channel estimation for mmWave massive multiple input multiple output (MIMO) systems. The inherent sparse features of the mmWave massive MIMO channels can be extracted and the sparse channel supports can be learnt by the multi-layer CNN-based network through training. Then accurate channel inference can be efficiently implemented using the trained network. The estimation accuracy and spectrum efficiency can be further improved by fully utilizing the spatial correlation among the sparse channel supports of different antennas. It is verified by simulation results that the proposed deep CNN-based scheme significantly outperforms the state-of-the-art benchmarks in both accuracy and spectrum efficiency.

Keywords: deep convolutional neural networks; deep learning; sparse channel estimation; mmWave massive MIMO

I. INTRODUCTION

Massive multiple input multiple output (MIMO) has been adopted as a key alternative in 5G wireless communications since it can significantly increase the data rate and spectral efficiency [1]. However, channel estimation for massive MIMO and millimeter wave (mmWave) massive MIMO systems [2, 3] becomes much more challenging since the overhead of time and frequency resources dramatically increase with

the number of antennas [4, 5]. Exploiting the sparse nature of wireless channels, i.e. the channel impulse response (CIR) is concentrated on only a few dominant taps in the delay domain while most entries are approximately zero [5–7], some compressed sensing (CS)-based methods have been proposed to reduce the overhead and improve the performance for both massive MIMO channels [8] and mmWave massive MIMO channels [9, 10]. Still in severe conditions, e.g. in the low signal-to-noise ratio (SNR) region or in case of insufficient measurement data, the estimation accuracy is limited.

Inspired by the emerging technology of deep learning, several effective deep neural networks have recently been proposed to solve sparse recovery problems. Based on the conventional recursive method, the conventional iterative recovery algorithms with good interpretability, e.g. the iterative shrinkage thresholding algorithm (ISTA) [11], can eventually obtain the globally optimal solution of sparse approximation problems. By unrolling the iterations of ISTA into a multi-layer neural network, the learned ISTA network (LISTA) [12] is proposed to mimic the repeated sparse approximation process of ISTA. An improved method of ISTA-Net casts the ISTA into an enhanced deep neural network architecture, where the convolutional neural network (CNN) is employed to extract a better sparse representation of the signal of interest [13].

Recently, some deep learning-based methods have been incorporated in the estimation of both massive MIMO and mmWave massive MIMO channels. For massive MIMO systems, deep learning is applied to estimate the uplink channels for massive MIMO with

mixed-resolution analog-to-digital converters in [14]. A deep learning-based framework is proposed for direction-of-arrival estimation and massive MIMO channel estimation [15]. A two-stage massive MIMO channel estimation process is proposed based on deep learning, including pilot-aided and data-aided estimation stages [16]. A massive MIMO channel estimator is jointly designed with pilot signals based on deep learning [17]. For mmWave massive MIMO systems, on the other hand, a learned denoising-based approximate message passing (LDAMP) network is proposed to estimate the beamspace mmWave massive MIMO channel [18]. The frequency correlation among channels at adjacent subcarriers is exploited in [19] for mmWave massive MIMO channel estimation using a CNN network.

In fact, the CNN network can be utilized better to extract the sparse feature of the delay-domain wireless channel supports. Besides, due to the spatial correlation of the sparse structure among the channel supports of different antennas, the CIRs of different antennas share a sparse common support [6, 7]. Actually, the spatial correlation might bring considerable benefit for the estimation performance of mmWave massive MIMO channels, which remains to be well exploited. Thus in this work, a novel deep learning and deep CNN-based mmWave massive MIMO channel estimation scheme is proposed. The inherent sparse structure of the wireless channel is effectively extracted by the CNN structure in each layer of the deep network. Moreover, the spatial correlation among a large number of antennas is also fully exploited to improve the estimation accuracy and spectrum efficiency. The main contributions of this work are summarized as follows:

- Accurate and efficient channel inference can be achieved using the well-trained deep CNN-based network, which can effectively extract the inherent sparse features of mmWave massive MIMO channels and reconstruct the channel support.

- The estimation accuracy of the sparse channel support is further refined by fully exploiting the spatial correlation among different MIMO sub-channels, i.e. a refined common channel support is derived by the intersection of all the inferred supports for different antennas.

Simulation results demonstrate that the proposed

scheme can significantly improve the accuracy of mmWave massive MIMO channel estimation, especially in case of low SNR or insufficient measurement data, i.e. frequency-domain pilots, compared with the state-of-the-art benchmark schemes, including the conventional least squares (LS), compressed sensing, and deep learning based methods. Furthermore, since the channel estimation process in practice is conducted by only one feed-forward operation flow exploiting the already well-trained networks, the proposed method is very applicable for practical mmWave communication user equipment.

The structure of this work is as follows. Related works are summarized in Section II. The system model is presented in Section III. The deep CNN-based mmWave massive MIMO channel estimation scheme is proposed in Section IV. Simulation results are given in Section V, followed by the conclusions.

The frequently used abbreviations and terminologies in this paper are listed in Table 1.

Table 1. *List of abbreviations.*

Abbreviation	Description
AWGN	additive white Gaussian noise
CIR	channel impulse response
CNN	convolutional neural networks
CS	compressed sensing
DFT	discrete Fourier transform
ISTA	iterative shrinkage thresholding algorithm
LS	least squares
MIMO	multiple-input multiple-output
NMSE	normalized mean square error
OFDM	orthogonal frequency division multiplexing
SNR	signal-to-noise ratio
SOMP	simultaneous orthogonal matching pursuit

Notation: Matrices and column vectors are denoted by boldface letters; frequency-domain and time-domain vectors are denoted by boldface vectors with tilde $\tilde{\mathbf{v}}$ and without tilde $\mathbf{v}$, respectively; The superscript (t) to a vector $\mathbf{v}^{(t)}$ denotes that the vector is related with the t-th transmit antenna; $\mathbb{C}$ denotes the set of complex number; $(\cdot)^{\dagger}$, $(\cdot)^T$ and $(\cdot)^H$ denote the pseudo-inversion, transpose and conjugate transpose operations, respectively; $\mathbf{A}_{\Pi}$ represents the sub-matrix comprised of the columns of the matrix $\mathbf{A}$ indexed by the set Π.

II. RELATED WORKS

As the number of antennas involved with the channels to be estimated in massive MIMO systems is large, the CS theory has been utilized to decrease the pilot consumption overhead effectively. A spatially non-stationary channel model is adopted to capture the delay-domain sparse characteristics of the massive MIMO channels, and block matching pursuit algorithm is used for massive MIMO channel estimatoin [5]. Qi *et al* provides an analysis of the block coherence of massive MIMO channels and proposes an improved sparse channel estimation scheme in the framework of CS [20]. Based on auxiliary information of the massive MIMO channels, an auxiliary block subspace pursuit algorithm is proposed to reconstruct the massive MIMO channel [21].

To improve the estimation accuracy in massive MIMO systems, some effective sparse channel estimation approaches have been proposed exploiting the spatial correlation, i.e. the sparse common support, among the multiple antennas. The sparse common support property is exploited to estimate the massive MIMO channels in the delay domain in [6]. An adaptive channel estimation scheme based on spatially common sparsity is investigated to achieve an enhanced performance of channel state information acquisition for massive MIMO channels [22]. A doubly sparse approach for mmWave massive MIMO channel estimation is proposed to improve the estimation accuracy greatly with affordable complexity [23].

Recently, the methods based on deep learning have drawn plenty of attention from the research community, and are also being introduced to the area of channel estimation of massive MIMO to achieve a better performance. A deep learning based scheme focused on massive MIMO channel estimation is proposed in [15], where deep neural networks are utilized to learn the statistics of the wireless channel in the angle domain. A channel estimation algorithm for doubly selective fading channels is proposed, which employs deep neural networks to facilitate channel estimation [24]. A deep-learning-based channel estimation scheme called ChanEstNet is designed in [25] for high speed scenarios, where convolutional neural networks are employed to extract the channel response feature vectors, and recurrent neural networks are used to have a better track of the features of the fast varying channels. Besides, since a layer-by-layer training method is employed, only a small proportion of the parameters need to be updated for each new layer of network during training.

III. SYSTEM MODEL

A mmWave massive MIMO transmission system is considered in this work as illustrated in Figure 1. It is equipped with N_t antennas at the transmitter, e.g. a base station, and N_r antennas at the receiver, e.g. a mobile user equipment, where N_r can also be one for scenarios like multi-user MIMO (MU-MIMO) [8]. The length-L CIR $\mathbf{h}^{(t)} \in \mathbb{C}^L$ between the t-th transmit antenna and a certain receive antenna, without loss of generality, can be modeled as

$$\mathbf{h}^{(t)} = \left[h_1^{(t)}, h_2^{(t)}, \cdots, h_L^{(t)} \right]^T. \quad (1)$$

Due to the propagation characteristics of outdoor environment, most of the energy of a wireless channel is concentrated on a few dominant taps, making the channel CIR a sparse or approximately sparse vector [6, 26, 27]. Note that the sparse channel is modeled in the delay-domain, i.e. the CIR vector $\mathbf{h}^{(t)}$ is representing a delay-domain sparse channel. This wireless channel is modeled based on the parametric sparse channel model given in [26]. Specifically, the sparsity of the CIR follows a parametric sparsity-promoting Bernoulli-Gaussian prior distribution. That is, the support of the CIR vector, i.e. the location of nonzero entries, is randomly determined and follows an i.i.d. Bernoulli distribution, with each entry of the CIR vector having an identical probability to be nonzero or zero. If a CIR entry is nonzero subject to the Bernoulli distribution, then it contributes a nonzero channel tap, whose amplitude is determined by a zero-mean Gaussian distribution [26]. In addition, according to the model [26], the parametric sparse channel with uncorrelated-scattering assumption is effective for specular sparse channels, approximately sparse channels and dense channels. Apart from the delay-domain sparse channel model described above, the Saleh-Valenzuela (S-V) model is also used in many cases for mmWave massive MIMO systems [28, 29]. The angle-domain sparsity can be conveniently represented by this model and exploited
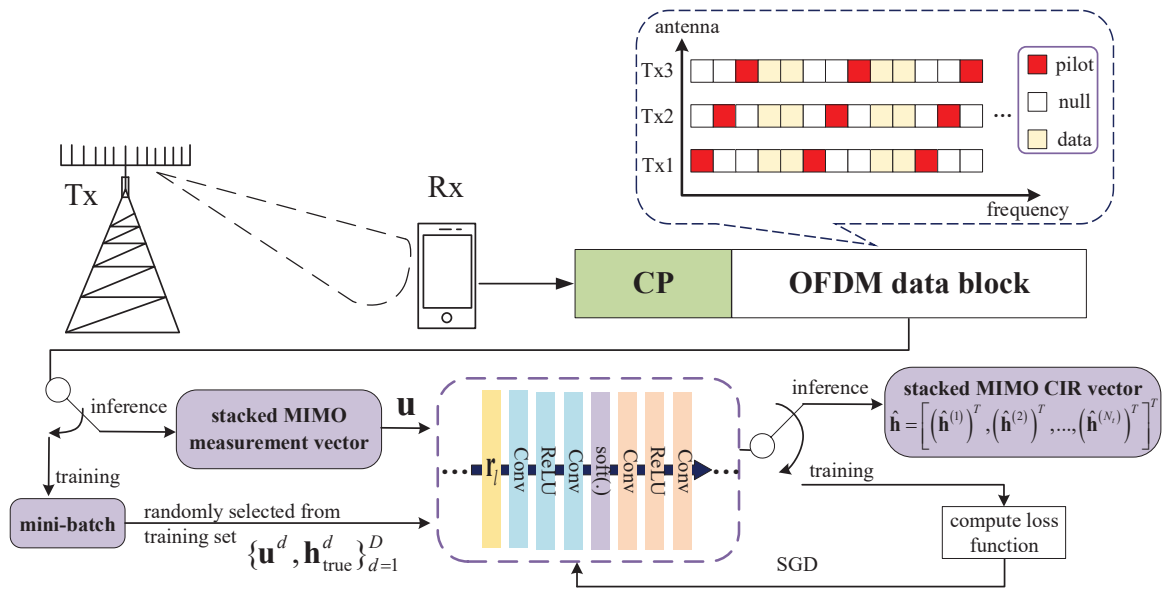
Figure 1. *The signal model and architecture of the proposed deep CNN-based channel estimation scheme for a mmWave massive MIMO system.*

for sparse channel estimation [10, 30]. The proposed method is also applicable to the problem formulated by this model, which can be conducted similarly by measuring and learning the angle-domain sparsity instead of the delay-domain sparsity of the channel.

A certain OFDM signal model sent by the t-th transmit antenna consists of the length-M cyclic prefix (CP) and the length-N OFDM data block $\mathbf{x}^{(t)} \in \mathbb{C}^N$ given by

$$\mathbf{x}^{(t)} = \left[x_1^{(t)}, x_2^{(t)}, ..., x_N^{(t)}\right]^T = \mathbf{F}^H \tilde{\mathbf{x}}^{(t)}, \qquad (2)$$

where $\mathbf{F}$ is the $N \times N$ discrete Fourier transform (DFT) matrix with N being the sub-carrier number. $\tilde{\mathbf{x}}^{(t)}$ denotes the OFDM data block in frequency domain, and N_{p} pilots are randomly distributed over the OFDM sub-carriers. The locations of these pilots are denoted by a set $D^{(t)}$ given by

$$D^{(t)} = \{d_n^{(t)}\}_{n=1}^{N_{\mathrm{p}}}, \qquad (3)$$

where $d_n^{(t)} \in [0, N-1]$ is an index integer denoting a pilot location. The pattern of the pilots is in a random manner, which follows a uniform distribution. The pilots of different transmit antennas are orthogonally distributed as illustrated in the pattern shown in Figure 1.

For a certain receive antenna, i.e. a sub-channel in the MIMO transmission system, the received frequency-domain OFDM data block $\tilde{\mathbf{y}} \in \mathbb{C}^N = [\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_N]^T$ can be represented as

$$\tilde{\mathbf{y}} = \sum_{t=1}^{N_t} \mathrm{diag}(\tilde{\mathbf{x}}^{(t)}) \mathbf{F}_L \mathbf{h}^{(t)} + \tilde{\mathbf{w}}, \qquad (4)$$

where $\mathrm{diag}(\tilde{\mathbf{x}}^{(t)})$ is a diagonal matrix with the diagonal being the vector $\tilde{\mathbf{x}}^{(t)}$, and $\mathbf{F}_L$ is the $N \times L$ partial DFT matrix composed of the first L columns of the $N \times N$ DFT matrix $\mathbf{F}$, with the assumption that the maximum channel length is smaller than the OFDM block length for commonly seen wireless channels, i.e. $L < N$. $\tilde{\mathbf{w}}$ is the frequency-domain additive white Gaussian noise (AWGN) vector. As only the pilots located at the sub-carrier set of $D^{(t)}$ are utilized for the channel estimation of the t-th transmit antenna and the pilot pattern is orthogonal between different transmit antennas, the received normalized pilots at the t-th transmit antenna can be simplified by picking the pilots located at $D^{(t)}$ out of the received signal $\tilde{\mathbf{y}}$ as given by

$$\mathbf{u}^{(t)} = \mathbf{F}_{\mathrm{p}}^{(t)} \mathbf{h}^{(t)} + \tilde{\mathbf{w}}^{(t)}, \quad 1 \le t \le N_t, \qquad (5)$$

where $\mathbf{u}^{(t)} \in \mathbb{C}^{N_{\mathrm{p}}}$ is the channel measurement vector of the t-th transmit antenna, i.e. the received pilots

normalized by the originally transmitted pilot given by

$$\begin{aligned}\mathbf{u}^{(t)} &= \mathbf{S}_{\mathrm{p}}^{(t)}(\mathrm{diag}(\widetilde{\mathbf{x}}^{(t)})^{-1})\widetilde{\mathbf{y}}\\ &= [\tilde{y}_{d_1^{(t)}}/\tilde{x}_{d_1^{(t)}}^{(t)}, \tilde{y}_{d_2^{(t)}}/\tilde{x}_{d_2^{(t)}}^{(t)}, \cdots, \tilde{y}_{d_{N_\mathrm{P}}^{(t)}}/\tilde{x}_{d_{N_\mathrm{P}}^{(t)}}^{(t)}]^T,\end{aligned}$$
(6)

where $\mathbf{S}_{\mathrm{p}}^{(t)} = \mathbf{I}_{N,D^{(t)}}^T$ is an $N_\mathrm{p} \times N$ selection matrix composed of the rows of the identity matrix $\mathbf{I}_N$ corresponding to $D^{(t)}$. The matrix $\mathbf{F}_{\mathrm{p}}^{(t)} = \mathbf{S}_{\mathrm{p}}^{(t)}\mathbf{F}_L$ in (5) is the $N_\mathrm{p} \times L$ partial DFT matrix with its entry in row-n and column-k being $\exp{(-j2\pi d_n^{(t)}(k-1)/N)}/\sqrt{N}$. $\tilde{\mathbf{w}}^{(t)}$ denotes the corresponding background noise vector.

By stacking the CIR vectors of all the transmit antennas $\{\mathbf{h}^{(t)}\}_{t=1}^{N_t}$ into a single long vector, the deep CNN network can be trained to exploit the inherent spatial correlation among the mmWave massive MIMO channels. This is because one of the most important advantages of CNN is that a convolutional layer uses the same convolution kernel, i.e. parameter sharing, for the data at different spatial locations to extract a higher-level common feature, such as in image processing and pattern recognition applications [31]. Then the stacked MIMO CIR vector is given by

$$\mathbf{h} = [(\mathbf{h}^{(1)})^T, (\mathbf{h}^{(2)})^T, ...(\mathbf{h}^{(N_t)})^T]^T \in \mathbb{C}^{N_t L}, \quad (7)$$

Consequently, the problem of sparse MIMO channel recovery can be formulated as

$$\mathbf{u} = \Phi\mathbf{h} + \tilde{\mathbf{w}}, \tag{8}$$

where $\mathbf{u} = [(\mathbf{u}^{(1)})^T, (\mathbf{u}^{(2)})^T, ...(\mathbf{u}^{(N_t)})^T]^T \in \mathbb{C}^{N_t N_\mathrm{p}}$ is the stacked measurement vector formulated by stacking all the received normalized pilots $\{\mathbf{u}^{(t)}\}_{t=1}^{N_t}$ into a single vector. The observation matrix $\Phi \in \mathbb{C}^{N_t N_\mathrm{p} \times N_t L}$ is a block-diagonal matrix with the t-th diagonal block sub-matrix being $\mathbf{F}_{\mathrm{p}}^{(t)}$, and $\tilde{\mathbf{w}}$ denotes the stacked noise vector of the pilots at all the transmit antennas.

Due to the spatial correlation of the mmWave massive MIMO system, i.e. the spatial distance between different transmit antennas is much smaller than the propagation distance between the transmitter and the receiver, the CIR vectors of different transmit-receive antenna pairs share a sparse common support [7]. More explicitly, the spatial correlation can be explained by the following formula as,

$$\Pi_0^{(1)} = \Pi_0^{(2)} = \cdots = \Pi_0^{(N_t)}, \tag{9}$$

where $\Pi_0^{(t)}$ denotes the support of the CIR vector $\mathbf{h}^{(t)}$ for the t-th transmit antenna. Note that the support of a sparse vector is defined as the locations of the nonzero entries of this sparse vector. Utilizing the spatial correlation of the MIMO sub-channels, the accuracy of the deep CNN-based channel inference can be further improved, which is discussed in the following section.

Although the proposed method is presented using a fully digital MIMO structure in this paper, it is also suitable for hybrid precoding structures [32–34], where the measurement model of the sparse MIMO channels can be built up in a way similar to the modeling given by (5) through (8), and the spatial correlation can be exploited to refine the support estimation, too.

IV. PROPOSED DEEP CNN-BASED MMWAVE MASSIVE MIMO CHANNEL ESTIMATION SCHEME

In this section, the novel deep CNN-based mmWave massive MIMO channel estimation scheme is proposed. The CNN-based scheme is devised based on the classical iterative sparse recovery algorithm of ISTA. Using the ISTA algorithm, the original unknown signal $\mathbf{x}$ can be recovered from the measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ through the iterations given by

$$\mathbf{r}_k = \mathbf{x}_{k-1} - \beta\mathbf{A}^T(\mathbf{A}\mathbf{x}_{k-1} - \mathbf{y}), \tag{10}$$

$$\mathbf{x}_k = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{r}_k\|_2^2 + \lambda\|\Psi\mathbf{x}\|_1, \tag{11}$$

where k is the iteration index and β is the stepsize. $\mathbf{r}_k$ denotes an intermediate proxy of $\mathbf{x}_k$, which is updated using the gradient descent of the squared estimation error $\|\mathbf{A}\mathbf{x}_{k-1} - \mathbf{y}\|_2^2$. The sparse solution of $\mathbf{x}$ can be obtained by minimizing the ℓ_1-norm of its sparse representation $\Psi\mathbf{x}$ using the sparse basis Ψ, and λ is a weighting parameter indicating the weight of sparsity on the minimization problem in (11).

In this work, a deep CNN-based sparse recovery network is devised to infer the mmWave massive MIMO channels, which only requires one forward propagation in the practical channel inference phase. The
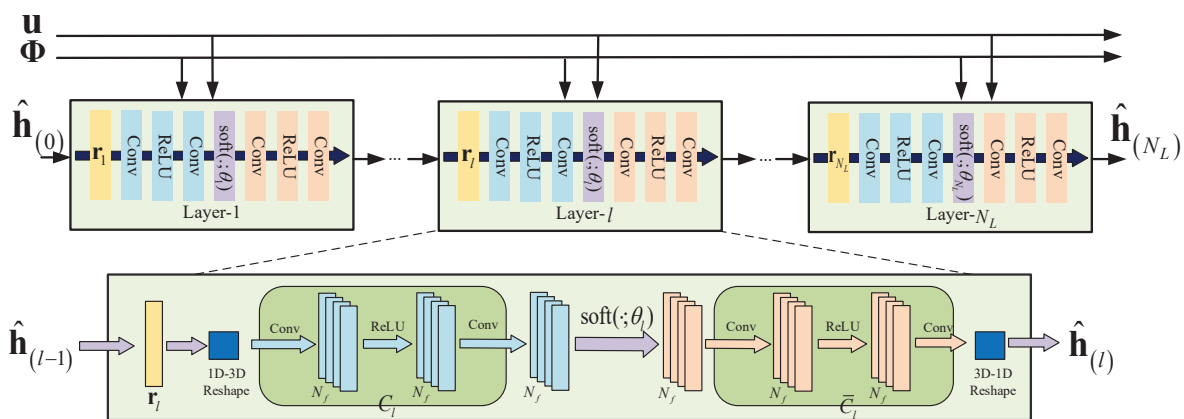
Figure 2. *The structure of the l-th layer in the proposed CNN-based network for mmWave massive MIMO channel estimation: $C_l(\cdot)$ denotes a learnable operator for sparse feature compression using CNNs, and $\bar{C}_l(\cdot)$ denotes the corresponding inverse operator for decompression.*

deep CNN network has a large fitting capacity and the sparse features can be effectively extracted to further improve the channel estimation accuracy.

The proposed network structure is illustrated in Figure 2, which consists of N_L identical cascaded layers. A learnable operator denoted by $C(\cdot)$, which is composed of two convolutional (Conv) layers separated by a rectified linear unit (ReLU), is adopted for sparse representation and sparse feature compression, playing the role of the conventional sparse basis Ψ. Both the two Conv layers in $C(\cdot)$ have N_f feature maps, respectively. Thus, the ability to extract sparse features is greatly improved and the capacity of the network to fit different mappings is substantially extended through using CNNs.

Utilizing the CNN-based operator $C(\cdot)$ for sparse regularization instead of the ℓ_1-norm $\|\Psi\mathbf{x}\|_1$, the classical optimization problem in (11) can be transferred into a deep network with much better sparse feature extraction capability. Specifically, for the l-th layer of the network, given the stacked received measurement vector $\mathbf{u}$ and the observation matrix Φ, the stacked MIMO CIR vector estimated by the l-th layer denoted by $\hat{\mathbf{h}}_{(l)}$ can be estimated by

$$\mathbf{r}_l = \hat{\mathbf{h}}_{(l-1)} - \beta_l \Phi^H (\Phi \hat{\mathbf{h}}_{(l-1)} - \mathbf{u}), \quad (12)$$

$$\hat{\mathbf{h}}_{(l)} = \bar{C}_l(\text{soft}(C_l(\mathbf{r}_l), \theta_l)), \quad (13)$$

where $\text{soft}(v; \theta)$ is the *soft threshold* function, with $\text{soft}(v; \theta) = 0, |v| \le \theta$, and $\text{soft}(v; \theta) = \text{sign}(v) \cdot (|v| - \theta), |v| > \theta$. $\bar{C}_l(\cdot)$ is the corresponding inverse

operator of $C_l(\cdot)$. Note that the learnable parameters up to layer-l in the network in Figure 2 can be aggregated into a set $\Theta_l = \left\{ \beta_i, \theta_i, C_i(\cdot), \bar{C}_i(\cdot) \right\}_{i=1}^{l}$, which contains the network weights and parameters associated with all the l layers. It can be learnt and optimized using back-propagation and gradient descent during training.

The training process is described in **Algorithm 1**, where the network is trained layer-by-layer in a sequential manner. The whole training process is composed of some number of stages indexed by l, and l is increasing sequentially during training. Specifically, in a certain stage l, the current network contains l layers and a corresponding current loss function $\mathcal{L}_l(\Theta_l)$ is formulated for this l-layers network with the parameter set Θ_l. The number of layers in the network is indicated by l, which should be increased as long as the current loss function $\mathcal{L}_l(\Theta_l)$ is still decreasing with the increase of l. When the loss function does not decrease with l any more, the final number of layers N_L can be determined, i.e. $N_L = l - 1$. The reason is that at this time the halting condition is met, i.e. $\mathcal{L}_l(\Theta_l) \ge \mathcal{L}_{l-1}(\Theta_{l-1})$, so the network with $l - 1$ layers in the previous stage is the one with the minimum training loss.

The current loss function $\mathcal{L}_l(\Theta_l)$ for the l-layers network is composed of two parts: the normalized mean square error contributed by the channel estimation output $\hat{\mathbf{h}}_{(l)}$ of the l layers, and the regularization term to impose symmetry constraints on the operator pairs $C_i(\cdot)$ and $\bar{C}_i(\cdot)$, $i = 1, 2, ..., l$. Hence, the loss func-

tion of layer-l $\mathcal{L}_l(\Theta_l)$ is given by

$$
\mathcal{L}_l(\Theta_l) = \underbrace{\frac{1}{D}\sum_{d=1}^{D}\frac{\left\|\mathbf{h}_{\text{true}}^d - \hat{\mathbf{h}}_{(l)}\right\|_2^2}{\left\|\mathbf{h}_{\text{true}}^d\right\|_2^2}}_{\text{channel estimation error}}
$$
$$
+ \delta \underbrace{\frac{1}{D}\sum_{d=1}^{D}\sum_{i=1}^{l}\frac{\left\|\mathbf{h}_{\text{true}}^d - \bar{C}_i(C_i(\mathbf{h}_{\text{true}}^d))\right\|_2^2}{\left\|\mathbf{h}_{\text{true}}^d\right\|_2^2}}_{\text{symmetry constraint}},
$$
$$
\tag{14}
$$

where δ is the penalty parameter for the symmetric constraints regularization. The parameter set $\Theta_l = \left\{\beta_i, \theta_i, C_i(\cdot), \bar{C}_i(\cdot)\right\}_{i=1}^{l}$ for the l-layers network is learnt through minimizing the loss function $\mathcal{L}_l(\Theta_l)$ over the training set $\{\mathbf{u}^d, \mathbf{h}_{\text{true}}^d\}_{d=1}^{D}$. Like the procedures in [19], the complex numbers are divided into real and imaginary parts, and trained separately in two networks with the same structure as shown in Figure 2. Then the output of the two networks are combined to form the estimated channel vector in complex numbers.

After the network has been well trained, the mmWave massive MIMO channels can be estimated accurately in the inference phase. First, the estimated stacked MIMO CIR vector $\hat{\mathbf{h}} = [(\hat{\mathbf{h}}^{(1)})^T, (\hat{\mathbf{h}}^{(2)})^T, ..., (\hat{\mathbf{h}}^{(N_t)})^T]^T \in \mathbb{C}^{N_t L}$ can be obtained through the trained network with the learned parameters Θ_{N_L}. The estimated CIR vector $\hat{\mathbf{h}}^{(t)}$ for the t-th transmit antenna can be extracted from the corresponding t-th length-L sub-vector of the estimated stacked MIMO CIR vector $\hat{\mathbf{h}}$ as given by

$$
\hat{\mathbf{h}}^{(t)} = [\hat{h}_{(t-1)L+1}, \cdots, \hat{h}_{tL}]^T, \quad t = 1, 2, ..., N_t,
\tag{15}
$$

Note that, $\hat{\mathbf{h}}^{(t)}$ is a sparse CIR vector, and therefore its support, i.e. the locations of the nonzero entries of $\hat{\mathbf{h}}^{(t)}$, can be obtained simply by aggregating the indices of the nonzero entries of $\hat{\mathbf{h}}^{(t)}$. Then, the spatial correlation between the MIMO sub-channels as described in Section III can be fully utilized to derive a refined common channel support, i.e. by the intersection of all the inferred supports for different antennas obtained from the output of the CNN-based network. Namely, the common support of the mmWave massive MIMO

Algorithm 1. *Deep CNN-Based Sparse Channel Estimation for mmWave Massive MIMO Systems.*

Input:

Training set with size D: $\{\mathbf{u}^d, \mathbf{h}_{\text{true}}^d\}_{d=1}^{D}$

Initialization:

1: $\hat{\mathbf{h}}_{(0)} \leftarrow \mathbf{0}, l \leftarrow 1$

Iterations:

2: **repeat**

3: Compute intermediate proxy $\mathbf{r}_l$ based on (12)

4: Generate the sparse stacked MIMO CIR vector $\hat{\mathbf{h}}_{(l)}$ based on (13)

5: Calculate current loss function $\mathcal{L}_l(\Theta_l)$ based on (14) and update current network parameters $\Theta_l = \left\{\beta_i, \theta_i, C_i(\cdot), \bar{C}_i(\cdot)\right\}_{i=1}^{l}$ through gradient descent and back-propagation over the training set $\{\mathbf{u}^d, \mathbf{h}_{\text{true}}^d\}_{d=1}^{D}$.

6: Increase the number of network layers $l \leftarrow l+1$

7: **until** loss function does NOT decrease, i.e. $\mathcal{L}_l(\Theta_l) \geq \mathcal{L}_{l-1}(\Theta_{l-1})$

8: Set the final number of network layers as $N_\text{L} \leftarrow l - 1$

Output:

Learned parameters $\Theta_{N_\text{L}} = \left\{\beta_i, \theta_i, C_i(\cdot), \bar{C}_i(\cdot)\right\}_{i=1}^{N_\text{L}}$

---

system can be refined as given by

$$
\bar{\Pi} = \bigcap_{t=1}^{N_t} \Pi^{(t)},
\tag{16}
$$

where $\Pi^{(t)}$ is the support of $\hat{\mathbf{h}}^{(t)}$ for the t-th transmit antenna. Finally, the corresponding refined CIR vector $\bar{\mathbf{h}}^{(t)}$ for the t-th transmit antenna based on the spatial correlation can be obtained by solving the least squares (LS) problem derived from (5) as given by

$$
\bar{\mathbf{h}}^{(t)}\Big|_{\bar{\Pi}} = \underset{\mathbf{h} \in \mathbb{C}^K}{\text{argmin}} \left\|\mathbf{u}^{(t)} - \mathbf{B}_{\bar{\Pi}}\mathbf{h}\right\|_2,
\tag{17}
$$

where the matrix $\mathbf{B} = \mathbf{F}_\text{p}^{(t)}$ and $\mathbf{B}_{\bar{\Pi}} \in \mathbb{C}^{N_\text{p} \times K}$ denotes the projecting matrix composed of the columns of $\mathbf{B} = \mathbf{F}_\text{p}^{(t)}$ corresponding to the support $\bar{\Pi}$. Solving the LS problem (17) yields $\bar{\mathbf{h}}^{(t)}|_{\bar{\Pi}} = (\mathbf{B}_{\bar{\Pi}}^H \mathbf{B}_{\bar{\Pi}})^{-1} \mathbf{B}_{\bar{\Pi}}^H \mathbf{u}^{(t)}$.

V. SIMULATION RESULTS

The performance of the proposed deep CNN-based sparse channel estimation scheme is evaluated by simulations. The conventional LS-based channel estimation method with linear interpolation [35] and the CS-based algorithm of simultaneous orthogonal matching pursuit (SOMP) [8], are evaluated for comparison. As a deep learning based benchmark, the state-of-the-art method of ISTA-Net [13] is also evaluated. The simulation parameters are configured in a mmWave massive MIMO system with the carrier frequency $f_c = 30$ GHz and $N_t = N_r = 64$ antennas. The length of the OFDM data block length and the CP is $N = 4096$ and $M = 256$, respectively. The maximum channel length is set as $L = 256$ [26, 36], which is the same as the CP length. It is a conservative maximum range of the possible CIR delay spreads generated according to the statistical channel distribution [26, 36]. Then, the CP length is sufficiently large to cover most of the CIR delay spreads to avoid inter-block interference. The actual sparsity level of the CIR vector is conservatively assumed to be $K = 12$, which denotes the number of dominant CIR taps with much greater amplitude than the other small-scale nonzero taps in the parametric sparse channel model [26]. The number of pilots adopted for channel measurements is set as $N_p = 25$. The SNR denoted by γ used in the simulations is calculated by $\gamma = P_r/\sigma^2$, where P_r denotes the power of the received signal, and σ^2 denotes the power of the AWGN $\tilde{w}$ as in equation (4).

The training set $\{u^d, h_{true}^d\}_{d=1}^D$, with $D = 1000$, is generated randomly according to the distribution of the parametric sparse channel [26], with each training sample d containing a stacked measurement vector u^d and the corresponding ground-truth stacked MIMO CIR vector h_{true}^d. The support of the delay taps of the MIMO channel is randomly distributed in all the L possible locations determined by a Gaussian-Bernoulli distribution [26] as stated in Section III. The testing set is generated in the same way. The stochastic gradient descent method and the Adam optimizer are utilized to train the network using mini-batch size of 64 with the learning rate of 10^{-4}, and the penalty parameter in the loss function (14) is set as $\delta = 0.05$. Each Conv layer has $N_f = 16$ feature maps of size 32×32, and the convolution kernel is of size 5×5 with stride 1.

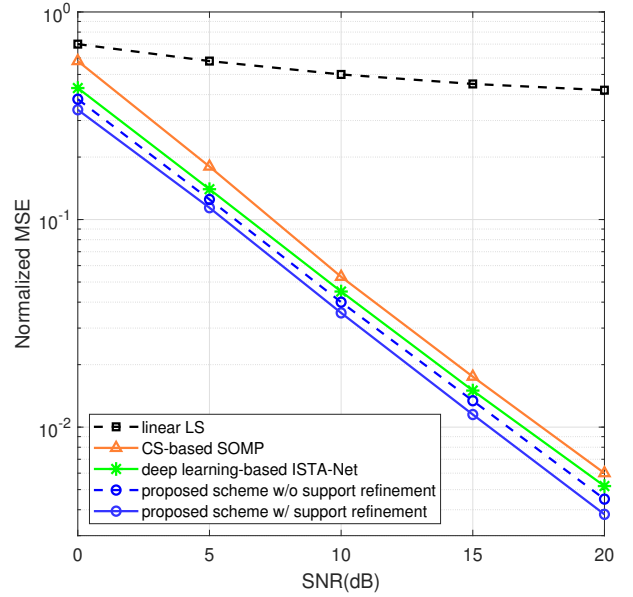Figure 3 presents the NMSE performance of chan-

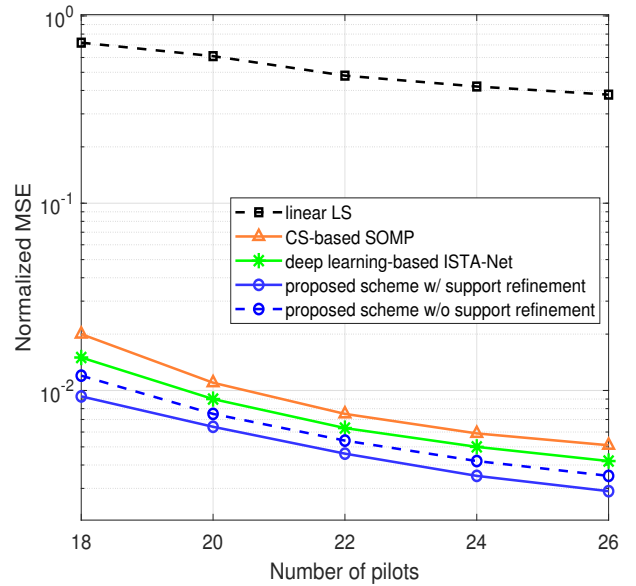**Figure 3.** *NMSE performance of mmWave massive MIMO channel estimation with respect to SNR.*

Figure 4. *NMSE performance of mmWave massive MIMO channel estimation with respect to the number of pilots N_p.*

nel estimation in a mmWave massive MIMO system with multipath fading. It can be noted from Figure 3 that, the proposed deep CNN-based scheme outperforms the LS-based method significantly, since the number of available pilots is much smaller than the channel length, making it an underdetermined problem that LS cannot tackle but the proposed scheme

can solve well. An NMSE gain of 2.0 dB over the CS-based method at the NMSE of 10^{-2} is achieved by the proposed deep CNN-based scheme, validating the high accuracy of the proposed scheme. The proposed scheme outperforms the deep learning-based algorithm of ISTA-Net by 1.3 dB, which indicates that the proposed network has been optimized for mmWave massive MIMO channel estimation. Moreover, the gap between the proposed scheme with and without support refinement is about 0.8 dB, which verifies that the spatial correlation is fully exploited by the proposed support refinement method given in (16).

The channel estimation performance with respect to the number of available pilots N_p adopted for channel measurements is reported in Figure 4 at the SNR of 20 dB. It can be observed from Figure 4 that, the proposed deep CNN-based scheme can achieve better accuracy compared with the existing CS and deep learning-based algorithms using the same number of pilots. Meanwhile, the proposed scheme requires much less pilots for channel estimation than the benchmarks to reach the same target accuracy, which saves the sub-carrier resource and improves the spectrum efficiency greatly. Besides, the proposed method with support refinement requires much less pilots than that without support refinement, which validates the benefits of utilizing the spatial correlation.

VI. CONCLUSION

In this work, a novel deep CNN-based mmWave massive MIMO channel estimation scheme has been proposed. By exploiting the CNN-based sparse recovery network, the mmWave massive MIMO channel is estimated with higher accuracy and spectrum efficiency than state-of-the-art methods. Simulation results verify that the proposed scheme can significantly improve the accuracy of mmWave massive MIMO channel estimation, especially in case of low SNR or insufficient measurement data, i.e. frequency-domain pilots, compared with the conventional LS-based method and the state-of-the-art CS-based and deep learning-based methods. The proposed deep CNN-based sparse recovery framework is promising to be widely applied in various mmWave massive MIMO scenarios, including indoor mmWave systems and outdoor mobility scenarios.

References

[1] S. Chen, S. Sun, *et al.*, "Beam-space multiplexing: Practice, theory, and trends-from 4g td-lte, 5g, to 6g and beyond," *IEEE Wireless Communications*, vol. 27, no. 2, 2020.

[2] S. Chen, Q. Gao, *et al.*, "A CSI acquisition approach for mmWave massive MIMO," *China Communications*, vol. 16, no. 9, 2019, pp. 1–14.

[3] Y. Zhao, W. Zhao, *et al.*, "AoA-based channel estimation for massive MIMO OFDM communication systems on high speed rails," *China Communications*, vol. 17, no. 3, 2020, pp. 90–100.

[4] Y. Li, M. Zhang, *et al.*, "Performance evaluation for medium voltage mimo-ofdm power line communication system," *China Communications*, vol. 17, no. 1, 2020, pp. 151–162.

[5] Y. Zhang, H. Cao, *et al.*, "Non-orthogonal multiple access in cell-free massive mimo networks," *China Communications*, vol. 17, no. 8, 2020, pp. 81–94.

[6] M. Masood, L. H. Afify, *et al.*, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, 2014, pp. 104–118.

[7] Y. Barbotin, A. Hormati, *et al.*, "Estimation of sparse MIMO channels with common support," *IEEE Transactions on Communications*, vol. 60, no. 12, 2012, pp. 3705–3716.

[8] M. Ke, Z. Gao, *et al.*, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Transactions on Signal Processing*, vol. 68, 2020, pp. 764–779.

[9] Z. Wan, Z. Gao, *et al.*, "Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, 2019, pp. 2337–2342.

[10] A. Liao, Z. Gao, *et al.*, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Transactions on Communications*, vol. 67, no. 12, 2019, pp. 8329–8345.

[11] U. S. Kamilov and H. Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Processing Letters*, vol. 23, no. 5, 2016, pp. 747–751.

[12] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on international conference on machine learning*, 2010, pp. 399–406.

[13] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1828–1837.

[14] S. Gao, P. Dong, *et al.*, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs,"

IEEE Communications Letters, vol. 23, no. 11, 2019, pp. 1989–1993.

[15] H. Huang, J. Yang, *et al.*, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, 2018, pp. 8549–8560.

[16] C.-J. Chun, J.-M. Kang, *et al.*, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, 2019, pp. 1228–1231.

[17] X. Ma and Z. Gao, "Data-driven deep learning to design pilot and channel estimator for massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, 2020, pp. 5677–5682.

[18] H. He, C.-K. Wen, *et al.*, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, 2018, pp. 852–855.

[19] P. Dong, H. Zhang, *et al.*, "Deep CNN-based channel estimation for mmWave massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, 2019, pp. 989–1000.

[20] C. Qi, Y. Huang, *et al.*, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 4558–4563.

[21] Y. Nan, L. Zhang, *et al.*, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for TDD massive MU-MIMO systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, 2015, pp. 345–348.

[22] Z. Gao, L. Dai, *et al.*, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, 2015, pp. 6169–6183.

[23] S. Gao, X. Cheng, *et al.*, "Making wideband channel estimation feasible for mmwave massive MIMO: A doubly sparse approach," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[24] Y. Yang, F. Gao, *et al.*, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, 2019, pp. 36 579–36 589.

[25] Y. Liao, Y. Hua, *et al.*, "ChanEstNet: A deep learning based channel estimation for high-speed scenarios," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[26] T. L. Hansen, P. B. Jørgensen, *et al.*, "An iterative receiver for OFDM with sparsity-based parametric channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, 2018, pp. 5454–5469.

[27] X. He, R. Song, *et al.*, "Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, 2015, pp. 2990–3004.

[28] Y. Chen, D. Chen, *et al.*, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems," *IEEE Transactions on Communications*, vol. 67, no. 12, 2019, pp. 8315–8328.

[29] Y. Chen, D. Chen, *et al.*, "Spatial lobes division-based low complexity hybrid precoding and diversity combining for mmWave IoT systems," *IEEE Internet of Things Journal*, vol. 6, no. 2, 2018, pp. 3228–3239.

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

[31] I. Goodfellow, Y. Bengio, *et al.*, *Deep learning*. MIT press Cambridge, 2016.

[32] O. El Ayach, S. Rajagopal, *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, 2014, pp. 1499–1513.

[33] Y. Chen, D. Chen, *et al.*, "Non-uniform quantization codebook-based hybrid precoding to reduce feedback overhead in millimeter wave MIMO systems," *IEEE Transactions on Communications*, vol. 67, no. 4, 2018, pp. 2779–2791.

[34] Y. Chen, D. Chen, *et al.*, "Millimeter-wave massive MIMO systems relying on generalized sub-array-connected hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, 2019, pp. 8940–8950.

[35] Y. Takano and H.-J. Su, "A low-complexity LS turbo channel estimation technique for MU-MIMO systems," *IEEE Signal Processing Letters*, vol. 25, no. 5, 2018, pp. 710–714.

[36] M. Series, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," *Report ITU*, vol. 638, 2009, pp. 1–72.

Biographies

Sicong Liu received his B.S.E. and PhD degree (with highest honor) both in electronic engineering from Tsinghua University, Beijing, China in 2012 and 2017. He was a visiting scholar in City University of Hong Kong, China. He served as a senior engineer in Huawei Technologies. Currently, he is a specially-appointed associate researcher and an assistant professor in Xiamen University, China. His research interests include sparse signal processing, machine learning, and wireless communications.

Xiao Huang received the B.S. degree in communication engineering from FuZhou University, China, in 2018. He is currently pursuing the M.S. degree with Department of Information and Communication Engineering, School of Informatics, Xiamen University, P. R. China. His current research interests include sparse signal processing, wireless communications, and machine learning.