Deep Learning for Beamspace Channel Estimation in Millimeter-Wave Massive MIMO Systems

Xiuhong Wei, Chen Hu, *Student Member, IEEE*, and Linglong Dai, *Senior Member, IEEE*

***Abstract*— Millimeter-wave massive multiple-input multiple-output (MIMO) can use a lens antenna array to considerably reduce the number of radio frequency (RF) chains, but channel estimation is challenging due to the number of RF chains is much smaller than that of antennas. By exploiting the sparsity of beamspace channels, the beamspace channel estimation can be formulated as a sparse signal recovery problem, which can be solved by the classical iterative algorithm named approximate message passing (AMP), and its corresponding version learned AMP (LAMP) realized by a deep neural network (DNN). However, these existing schemes cannot achieve satisfactory estimation accuracy. To improve the channel estimation performance, we propose a prior-aided Gaussian mixture LAMP (GM-LAMP) based beamspace channel estimation scheme in this paper. Specifically, based on the prior information that beamspace channel elements can be modeled by the Gaussian mixture distribution, we first derive a new shrinkage function to refine the AMP algorithm. Then, by replacing the original shrinkage function in the LAMP network with the derived Gaussian mixture shrinkage function, a prior-aided GM-LAMP network is developed to estimate the beamspace channel more accurately. Simulation results by using both the theoretical channel model and the ray-tracing based channel dataset show that, the proposed GM-LAMP network can achieve better channel estimation accuracy than existing schemes.**

***Index Terms*— Millimeter-wave (mmWave), massive MIMO, beamspace channel estimation, approximate message passing (AMP), deep learning.**

I. Introduction

MILLIMETER-WAVE (mmWave) massive multiple-input multiple-output (MIMO) has been considered as a key technique for 5G and beyond [1]. In order to reduce the hardware cost and power consumption caused by a large number of antennas and the associated radio frequency (RF) chains, the lens antenna array has been recently investigated to provide an energy-efficient realization of hybrid precoding for mmWave massive MIMO [2], [3]. By employing the lens antenna array, which can concentrate signals from different directions on different antennas, the spatial channel can be converted to the beamspace channel [4]. As there are only a few dominant propagation paths with large path gains at mmWave frequencies, the beamspace channel in mmWave massive MIMO systems is sparse in nature [5]. Therefore, by only selecting a small number of dominant beams, the number of RF chains connected to the digital baseband can be considerably reduced. Beam selection requires the accurate channel state information (CSI) in the beamspace [6], which is challenging due to the high channel dimension, especially when the number of RF chains is much smaller than the number of antennas [7]–[9].

A. Prior Works

There are some recently proposed schemes for beamspace channel estimation. Specifically, [10] proposed a two-way channel estimation scheme with low computational complexity, where the antennas corresponding to the dominant beams are firstly determined by beam training between the base station (BS) and users, and then only channel elements corresponding to these selected antennas are estimated. However, the number of pilot symbols required to scan all possible beams is proportional to the number of BS antennas, which is very large (e.g., 256 antennas). Furthermore, by exploiting the sparsity of beamspace channels, some classical compressive sensing (CS) based schemes could estimate the beamspace channel with a reduced pilot overhead [11]–[13], such as the orthogonal matching pursuit (OMP) algorithm used in [11]. Apart from the sparsity, the beamspace channel may exhibit angular spreads. Based on this channel characteristics, [14] proposed a two-stage CS method for channel estimation, which consists of a matrix completion stage and a sparse recovery stage.

Unfortunately, all of these beamspace channel estimation schemes above [11]–[14] cannot achieve satisfactory estimation accuracy in low signal-to-noise ratio (SNR) regions, and they also have high computational complexity especially when the sparsity level of the beamspace channel is high. As a powerful iterative algorithm for sparse signal recovery, the approximate message passing (AMP) algorithm can be used to estimate the beamspace channel with low computational complexity [15], [16]. However, it is difficult to find the optimal shrinkage parameters for the AMP algorithm (the empirical shrinkage parameters are usually used instead), which restricts its channel estimation performance in practice.

Recently, the amazing success of deep learning (DL) in other fields like image recognition [17], [18] and speech

Manuscript received December 30, 2019; revised May 28, 2020 and August 4, 2020; accepted September 15, 2020. Date of publication September 28, 2020; date of current version January 15, 2021. This work was supported by the National Key Research and Development Program of China under Grant 2020YFB1805005 and the National Natural Science Foundation of China for Outstanding Young Scholars under Grant 61722109. The associate editor coordinating the review of this article and approving it for publication was L. Jalloul. *(Corresponding author: Linglong Dai.)*

The authors are with the Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: weixh19@mails.tsinghua.edu.cn; huc16@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.3027027

processing [19] has greatly inspired researchers to use this powerful tool to solve some problems in wireless communications [20]–[23]. With the help of DL, we can extract underlying features of wireless big data and provide some improved solutions to some complicated problems in wireless communications, such as low density parity check (LDPC) decoding [20], sparse code multiple access (SCMA) codebook design [21], end-to-end communications [22], and hybrid precoding for massive MIMO [23].

Inspired by the powerful learning ability of deep neural networks (DNNs), some DL based channel estimation schemes have been proposed. Reference [24] was the first work to exploit the DL tool for channel estimation in the wireless energy transfer system. They developed an autoencoder based channel estimation scheme, where the encoder is used to design pilots and the decoder is used to estimate channels. In order to make the pilot length smaller than the number of the antennas in the massive MIMO system, [25] proposed a joint pilot and data aided channel estimation scheme using DNNs, where the pilot-aided estimation process is realized by a two-layer neural network together with a DNN and the data-aided estimation process is realized by another DNN. Reference [26] proposed a DL based super-resolution channel estimation scheme in the mmWave massive MIMO system, which leverages a DNN for direction-of-arrival (DOA) estimation.

All of these works above [24]–[26] adopt classical DNNs (e.g., multilayer perceptron) to solve the channel estimation problems and achieve better performance in different scenarios. However, their neural networks are usually regarded as the black-boxes, which may lack stable performance guarantees [27]. By contrast, there are some other DL based channel estimation schemes by integrating the conventional algorithms that have certain performance guarantees with the DL tool. Reference [28] has presented a learned denoising-based approximate message passing (LDAMP) network for channel estimation, where a denoising convolutional neural network (DnCNN) for image recovery is incorporated into the AMP algorithm to replace the original shrinkage function. In order to improve the performance of the sparse signal recovery, [29] proposed a learned AMP (LAMP) network by directly unfolding the iterations of the AMP algorithm into the corresponding layer-wise network structure, where its linear transform coefficients and nonlinear shrinkage parameters are jointly optimized by the DNN. However, when used to solve the beamspace channel estimation problem, the existing LAMP network cannot achieve the satisfactory estimation accuracy.

B. Our Contributions

In this article, in order to improve the estimation performance, we propose a complex-valued Gaussian mixture LAMP (GM-LAMP) based beamspace channel estimation scheme by fully utilizing the prior information of the beamspace channel.[1]

[1]Simulation codes are provided to reproduce the results presented in this article: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html.

Specifically, we first exploit the prior information that beamspace channel elements follow the Gaussian mixture distribution to derive a new shrinkage function. Then, by replacing the original shrinkage function in the LAMP network with the derived Gaussian mixture shrinkage function, a prior-aided GM-LAMP network is developed to estimate the beamspace channel. Finally, we verify our work by using the widely used channel model for theoretical analysis and the publicly-available channel dataset based on ray-tracing, respectively. Simulation results show that, compared with conventional algorithms, the proposed GM-LAMP network can achieve better estimation accuracy in the above two channels.

C. Organization and Notation

The rest of the paper is organized as follows. In section II, the beamspace channel estimation problem in mmWave massive MIMO systems is formulated as a sparse signal recovery problem, and the conventional AMP algorithm and LAMP network for solving this problem are briefly reviewed. In section III, we derive the new shrinkage function based on the Gaussian mixture distribution, and propose the GM-LAMP network for improved beamspace channel estimation. The computational complexity of the proposed scheme is also analyzed in Section III. Simulation results are provided to show the performance of the proposed GM-LAMP network in Section IV. Finally, conclusions are given in Section V.

Notation: Lower-case and upper-case boldface letters $\mathbf{a}$ and $\mathbf{A}$ denote a vector and a matrix, respectively; $\mathbf{a}^*$ denotes the conjugate of vector $\mathbf{a}$; $\mathbf{A}^H$ and $\mathbf{A}^T$ denote the conjugate transpose and transpose of matrix $\mathbf{A}$, respectively; $\|\mathbf{a}\|_2$ denotes the l_2-norm of vector $\mathbf{a}$; $|a|$ denotes the amplitude of scalar a; a^* denotes the conjugate of scalar a; $\mathbf{A}\otimes\mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$; $\mathcal{U}(-a,a)$ denotes the probability density function of uniform distribution on $(-a,a)$; $\delta(x)$ denotes the Dirac delta function; $\mathrm{sinc}(\mathrm{x}) \triangleq \frac{\sin(\mathrm{N}\pi\mathrm{x})}{\mathrm{N}\pi\mathrm{x}}$ denotes the Dirichlet sinc function. Finally, $\mathbf{I}_K$ is the $K \times K$ identity matrix.

II. System Model

In this section, we first introduce the beamspace channel model, and then formulate the beamspace channel estimation problem as a sparse signal recovery problem. Finally, the conventional AMP algorithm [15] and its corresponding LAMP network proposed in [29] to solve this problem are reviewed.

A. Beamspace Channel

We consider a time division duplex (TDD) based mmWave massive MIMO system, as shown in Fig. 1 [30], where the BS employs a lens antenna array with N antennas and N_{RF} RF chains to simultaneously serve K single-antenna users.

In order to formulate the beamspace channel estimation problem, we start with the conventional mmWave massive MIMO channel in the spatial domain. According to the widely used Saleh-Valenzuela channel model [11], the channel vector $\mathbf{h}_k$ of size $N \times 1$ between the kth $(k = 1, 2, \cdots, K)$ user and the N-antenna BS can be presented by

$$\mathbf{h}_k = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \beta_{k,l} \mathbf{a}\left(\theta_{k,l}^{\mathrm{azi}}, \theta_{k,l}^{\mathrm{ele}}\right) = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \mathbf{c}_{k,l}, \quad (1)$$

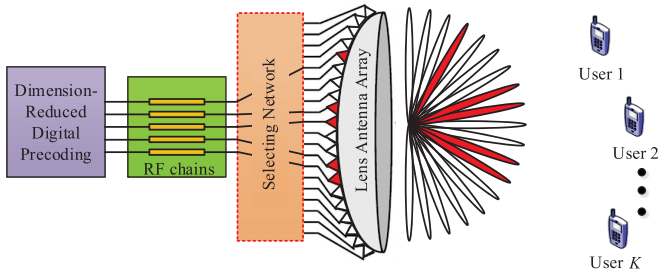

Fig. 1. mmWave massive MIMO with lens antenna array [30].

where L_k is the number of resolvable paths, and $\mathbf{c}_{k,l} = \beta_{k,l}\mathbf{a}\left(\theta_{k,l}^{\text{azi}}, \theta_{k,l}^{\text{ele}}\right)$ is the lth path component. $\beta_{k,l}$, $\theta_{k,l}^{\text{azi}}$ and $\theta_{k,l}^{\text{ele}}$ are the complex gain, azimuth angle and elevation angle of the lth path, respectively. $\mathbf{a}\left(\theta_{k,l}^{\text{azi}}, \theta_{k,l}^{\text{ele}}\right)$ is the $N \times 1$ array steering vector, which depends on the array geometry. Ignoring the subscripts without loss of generality, for the simpler uniform linear arrays (ULAs), the array steering vector can be determined by one angle, which can be presented by [11]

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

where $\mathbf{n}=[0, 1, \cdots, N-1]^T$. For the widely considered uniform planar arrays (UPAs) with $N_1 \times N_2$ ($N = N_1 \times N_2$) antennas, we have [31]

$$\mathbf{a}_{\text{UPA}}\left(\theta^{\text{azi}}, \theta^{\text{ele}}\right) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi d\sin\left(\theta^{\text{azi}}\right)\sin\left(\theta^{\text{ele}}\right)\mathbf{n}_1/\lambda}\right] \otimes \left[e^{-j2\pi d\cos\left(\theta^{\text{ele}}\right)\mathbf{n}_2/\lambda}\right], \tag{3}$$

where $\mathbf{n}_1 = [0, 1, \cdots, N_1-1]^T$ and $\mathbf{n}_2 = [0, 1, \cdots, N_2-1]^T$. In (2) and (3), λ is the wavelength of carrier, and d is the antenna spacing usually satisfying $d = \lambda/2$ in mmWave communications [32]. Then, we can respectively define $\psi \triangleq d\sin(\theta)/\lambda$ as the spatial angle for ULAs, and $\psi^{\text{azi}} \triangleq d\sin\left(\theta^{\text{azi}}\right)\sin\left(\theta^{\text{ele}}\right)/\lambda$ and $\psi^{\text{ele}} \triangleq d\cos\left(\theta^{\text{ele}}\right)/\lambda$ as the spatial angles for UPAs.

The spatial domain channel can be directly transformed to the beamspace channel by using a lens antenna array. As a matter of fact, the lens antenna array plays the role of a spatial discrete fourier transform (DFT) matrix $\mathbf{U}$ of size $N \times N$ [30]. For ULAs, the matrix $\mathbf{U}$ can be expressed as

$$\mathbf{U} = \left[\bar{\mathbf{a}}_{\text{ULA}}\left(\bar{\psi}_1\right), \bar{\mathbf{a}}_{\text{ULA}}\left(\bar{\psi}_2\right), \cdots, \bar{\mathbf{a}}_{\text{ULA}}\left(\bar{\psi}_N\right)\right]^H, \tag{4}$$

where $\bar{\psi}_n = \frac{1}{N}\left(n - \frac{N+1}{2}\right)$ for $n = 1, 2, \cdots, N$ are the spatial directions predefined by the lens antenna array. Similar to (2), $\bar{\mathbf{a}}_{\text{ULA}}(\psi)$ can be presented by

$$\bar{\mathbf{a}}_{\text{ULA}}(\psi) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\psi\mathbf{n}}\right]. \tag{5}$$

For UPAs, $\mathbf{U}$ can be expressed as

$$\mathbf{U} = \left[\bar{\mathbf{a}}_{\text{UPA}}\left(\bar{\psi}_1^{\text{azi}}, \bar{\psi}_1^{\text{ele}}\right), \cdots, \bar{\mathbf{a}}_{\text{UPA}}\left(\bar{\psi}_1^{\text{azi}}, \bar{\psi}_{N_2}^{\text{ele}}\right), \cdots, \bar{\mathbf{a}}_{\text{UPA}}\left(\bar{\psi}_{N_1}^{\text{azi}}, \bar{\psi}_1^{\text{ele}}\right), \cdots, \bar{\mathbf{a}}_{\text{UPA}}\left(\bar{\psi}_{N_1}^{\text{azi}}, \bar{\psi}_{N_2}^{\text{ele}}\right)\right]^H, \tag{6}$$

where $\bar{\psi}_n^{\text{azi}} = \frac{1}{N_1}\left(n - \frac{N_1+1}{2}\right)$ for $n = 1, 2, \cdots, N_1$ and $\bar{\psi}_n^{\text{ele}} = \frac{1}{N_2}\left(n - \frac{N_2+1}{2}\right)$ for $n = 1, 2, \cdots, N_2$ are respectively predefined spatial angles of azimuth and elevation by the lens antenna array. Similar to (3), $\bar{\mathbf{a}}_{\text{UPA}}\left(\psi^{\text{azi}}, \psi^{\text{ele}}\right)$ can be presented by

$$\bar{\mathbf{a}}_{\text{UPA}}\left(\psi^{\text{azi}}, \psi^{\text{ele}}\right) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\psi^{\text{azi}}\mathbf{n}_1}\right] \otimes \left[e^{-j2\pi\psi^{\text{ele}}\mathbf{n}_2}\right]. \tag{7}$$

Finally, the beamspace channel vector $\tilde{\mathbf{h}}_k$ of size $N \times 1$ between the kth user and the N-antenna BS can be presented by

$$\tilde{\mathbf{h}}_k = \mathbf{U}\mathbf{h}_k = \sqrt{\frac{N}{L_k}}\sum_{l=1}^{L_k}\tilde{\mathbf{c}}_{k,l}, \tag{8}$$

where $\tilde{\mathbf{c}}_{k,l} = \mathbf{U}\mathbf{c}_{k,l}$ is the lth channel component of the beamspace channel.

B. Problem Formulation

In order to acquire the CSI, all users should transmit the known pilot symbols to the BS over Q instants. Due to the reciprocity of the TDD channel, we can only consider the uplink to formulate the channel estimation problem. Then, the downlink channel can be directly obtained according to the estimated uplink channel. In this article, we adopt the widely used orthogonal pilot transmission strategy [30], where the uplink channel estimation for each user is independent due to the pilot orthogonality, and thus we can estimate the beamspace channel vectors between all K users and the BS one by one. Without loss of generality, we take the beamspace channel vector $\tilde{\mathbf{h}}_k$ between the kth user and the BS as an example to formulate the channel estimation problem.

In the qth instant for pilot transmission, the $N_{\text{RF}} \times 1$ measurement signal $\mathbf{y}_{k,q}$ in the baseband at the BS after beam selection can be presented as [30]

$$\mathbf{y}_{k,q} = \mathbf{A}_{k,q}\tilde{\mathbf{h}}_k s_{k,q} + \bar{\mathbf{n}}_{k,q}, \quad q = 1, 2, \cdots, Q, \tag{9}$$

where $\mathbf{A}_{k,q}$ is the $N_{\text{RF}} \times N$ beam selection network, $s_{k,q}$ is the transmitted pilot symbol, $\bar{\mathbf{n}}_{k,q} = \mathbf{A}_{k,q}\mathbf{n}_{k,q}$ is the effective noise vector, where $\mathbf{n}_{k,q} \sim \mathcal{CN}\left(0, \sigma_n^2\mathbf{I}_{\mathbf{N}}\right)$ is the $N \times 1$ noise vector with σ_n^2 representing the noise power.

After Q instants of pilot transmission, we can obtain the $M \times 1$ ($M = QN_{\text{RF}}$) overall measurement signal $\mathbf{y}_k$ by assuming $s_{k,q} = 1$ for $q = 1, 2, \cdots, Q$ as

$$\mathbf{y}_k = \begin{bmatrix}\mathbf{y}_{k,1}\\ \mathbf{y}_{k,2}\\ \vdots\\ \mathbf{y}_{k,Q}\end{bmatrix} = \mathbf{A}_k\tilde{\mathbf{h}}_k + \mathbf{n}_k, \tag{10}$$

where $\mathbf{A}_k = [\mathbf{A}_{k,1}^T, \mathbf{A}_{k,2}^T, \cdots, \mathbf{A}_{k,Q}^T]^T$ is the $M \times N$ selection matrix with the entry being $\pm\frac{1}{\sqrt{M}}$ [30], and $\mathbf{n}_k = [\bar{\mathbf{n}}_{k,1}^T, \bar{\mathbf{n}}_{k,2}^T, \cdots, \bar{\mathbf{n}}_{k,Q}^T]^T$ is the $M \times 1$ effective noise vector for Q instants.

Since the channel estimation method is the same for all K users due to the pilot orthogonality, the subscript k in the problem (10) can be omitted, then (10) can be expressed as

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

Note that only the elements of the beamspace channel $\tilde{\mathbf{h}}$ that are close to the practical spatial angles of the channel paths have large values. As there are only a few propagation paths due to limited scattering at mmWave frequencies, the beamspace channel $\tilde{\mathbf{h}}$ is approximately sparse [5]. Consequently, we can apply the sparse signal recovery algorithms in CS to estimate the beamspace channel with a low pilot overhead, where the matrix $\mathbf{A}$ in (11) can be regarded as the sensing matrix in CS. That is to say, the beamspace channel estimation problem in (11) can be formulated as a sparse signal recovery problem

$$\min \left\|\tilde{\mathbf{h}}\right\|_0, \quad \text{s.t.} \left\|\mathbf{y}-\mathbf{A}\tilde{\mathbf{h}}\right\|_2 \leq \varepsilon, \tag{12}$$

where $\left\|\tilde{\mathbf{h}}\right\|_0$ is the number of non-zero elements of $\tilde{\mathbf{h}}$, ε is the error tolerance parameter.

Due to the non-convexity of the l_0-norm, the problem in (12) is NP-hard [33]. Therefore, this problem is usually converted to a convex optimization problem by replacing the l_0-norm with the l_1-norm [34]–[36]. There have been some conventional greedy algorithms to solve it, such as OMP [11] and compressive sampling matching pursuit (CoSaMP) [37]. However, these greedy algorithms cannot achieve satisfactory estimation accuracy. Especially, with the increase of the sparsity level, the computational complexity of the greedy algorithms will be also increased.

C. AMP Algorithm and LAMP Network

Since the number of antennas in mmWave massive MIMO systems is usually large, the dimension of the sparse signal in (12) is high. Thanks to faster convergence, the iterative AMP algorithm can be used to recover the sparse signal with low computational complexity, especially for the high-dimensional sparse signal [15]. In this subsection, we introduce how the complex-valued AMP algorithm estimates the beamspace channel, as shown in **Algorithm 1**.

Algorithm 1 Approximate Message Passing (AMP)

Input: The measurement vector $\mathbf{y}$, the sensing matrix $\mathbf{A}$, the number of iterations T.

Initialization: $\mathbf{v}_{-1}=\mathbf{0}, b_0=0, c_0=0, \hat{\tilde{\mathbf{h}}}_0=\mathbf{0}$.

for $t=0,1,\cdots,T-1$ **do**

1. $\mathbf{v}_t=\mathbf{y}-\mathbf{A}\hat{\tilde{\mathbf{h}}}_t+b_t\mathbf{v}_{t-1}+c_t\mathbf{v}_{t-1}^*$
2. $\sigma_t^2=\frac{1}{M}\left\|\mathbf{v}_t\right\|_2^2$
3. $\mathbf{r}_t=\hat{\tilde{\mathbf{h}}}_t+\mathbf{A}^T\mathbf{v}_t$
4. $\hat{\tilde{\mathbf{h}}}_{t+1}=\boldsymbol{\eta}_{\mathrm{st}}(\mathbf{r}_t;\lambda_t,\sigma_t^2)$
5. $b_{t+1}=\frac{1}{M}\sum_{i=1}^{N}\frac{\partial\eta_{\mathrm{st}}(r_{t,i};\lambda_t,\sigma_t^2)}{\partial r_{t,i}}$
6. $c_{t+1}=\frac{1}{M}\sum_{i=1}^{N}\frac{\partial\eta_{\mathrm{st}}(r_{t,i};\lambda_t,\sigma_t^2)}{\partial r_{t,i}^*}$

end for

Output: Sparse signal recovery results: $\hat{\tilde{\mathbf{h}}}=\hat{\tilde{\mathbf{h}}}_T$.

In **Algorithm 1**, the term $b_t\mathbf{v}_{t-1}$ and term $c_t\mathbf{v}_{t-1}^*$ in Step 1 are called Onsager Correction [15], which are introduced into the AMP algorithm to accelerate the convergence.

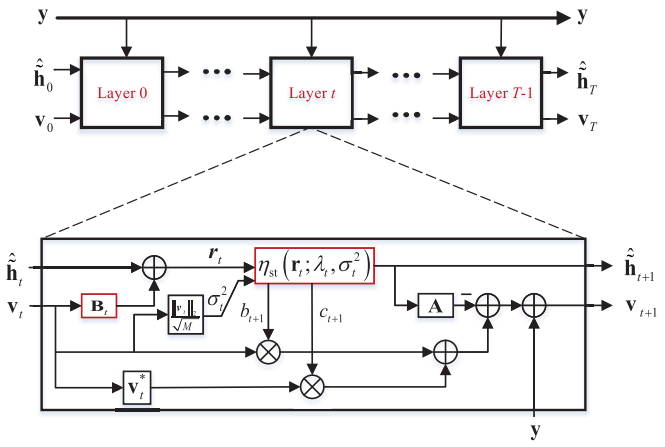

Fig. 2. LAMP network structure (the tth layer is explained in detail) [29].

The critical step of the AMP algorithm is Step 4, in which the estimate $\hat{\tilde{\mathbf{h}}}_{t+1}$ in the tth iteration is obtained through the soft threshold shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$: $\mathbb{C}^N\rightarrow\mathbb{C}^N$. The shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ is nonlinear element-wise operation, which takes the sparsity of the vector $\tilde{\mathbf{h}}$ into consideration, and makes the estimate $\hat{\tilde{\mathbf{h}}}_{t+1}$ sparser. For the ith element $r_{t,i}=|r_{t,i}|\,e^{j\omega_{t,i}}$ $(i=1,2,\cdots,N)$ of input vector $\mathbf{r}_t$, we have

$$\begin{aligned}\left[\boldsymbol{\eta}_{\mathrm{st}}\left(\mathbf{r}_t;\lambda_t,\sigma_t^2\right)\right]_i &= \eta_{\mathrm{st}}\left(|r_{t,i}|\,e^{j\omega_{t,i}};\lambda_t,\sigma_t^2\right)\\ &= \max\left(|r_{t,i}|-\lambda_t\sigma_t,0\right)e^{j\omega_{t,i}},\end{aligned} \tag{13}$$

where $\omega_{t,i}$ is the phase of complex-valued element $r_{t,i}$, λ_t is the predefined and fixed parameter in the tth iteration, and σ_t^2 is updated via estimating the noise variance in Step 2. From (13), we can find that the soft threshold shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ can shrink the amplitude of complex-valued input with low power to zero. In Step 5 and Step 6, the element-wise derivatives of the shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ at the input vector $\mathbf{r}$ and its conjure vector $\mathbf{r}^*$ are respectively calculated to obtain b_{t+1} and c_{t+1}.

Although the AMP algorithm is good at dealing with the large-scale sparse signal recovery problem, there are still two problems when it is used for the sparse beamspace channel estimation. First, the shrinkage parameter λ_t in (13) usually takes the same empirical value for all iterations. Second, the general AMP algorithm cannot fully exploit the prior distribution of the beamspace channel. These two problems limit the performance of the AMP algorithm.

To solve the first problem, the LAMP network based on the classical AMP algorithm has been recently proposed to optimize the nonlinear shrinkage parameter λ_t in each iteration [29]. As shown in Fig. 2, each iteration of the classical AMP algorithm is mapped to each layer of the LAMP network. To be specific, the inputs of the tth layer are $\mathbf{y}\in\mathbb{C}^M$, $\hat{\tilde{\mathbf{h}}}_t\in\mathbb{C}^N$ and $\mathbf{v}_t\in\mathbb{C}^M$, where $\mathbf{y}$ is the measurement vector in (11), $\hat{\tilde{\mathbf{h}}}_t$ and $\mathbf{v}_t$ are the outputs of the $(t-1)$th layer. Following the principle of the AMP algorithm, each layer of the LAMP network processes the signal as follows, which is

similar to **Algorithm 1**:

$$\hat{\tilde{\mathbf{h}}}_{t+1} = \boldsymbol{\eta}_{\mathrm{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2), \tag{14}$$

$$\mathbf{v}_{t+1} = \mathbf{y} - \mathbf{A}\hat{\tilde{\mathbf{h}}}_t + b_{t+1}\mathbf{v}_t + c_{t+1}\mathbf{v}_t^*, \tag{15}$$

where

$$\mathbf{r}_t = \hat{\tilde{\mathbf{h}}}_t + \mathbf{B}_t\mathbf{v}_t, \tag{16}$$

$$\sigma_t^2 = \frac{1}{M}\left\|\mathbf{v}_t\right\|_2^2, \tag{17}$$

$$b_{t+1} = \frac{1}{M}\sum_{i=1}^{N}\frac{\partial \eta_{\mathrm{st}}\left(r_{t,i}; \lambda_t, \sigma_t^2\right)}{\partial r_{t,i}}, \tag{18}$$

$$c_{t+1} = \frac{1}{M}\sum_{i=1}^{N}\frac{\partial \eta_{\mathrm{st}}\left(r_{t,i}; \lambda_t, \sigma_t^2\right)}{\partial r_{t,i}^*}, \tag{19}$$

where the shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ of the AMP algorithm plays a role of the nonlinear activation function in the conventional DNN [29]. What's more, from (16), we can find that different from the Step 3 in **Algorithm 1**, the LAMP network can choose the different linear coefficients $\mathbf{B}_t$ for each layer t, which can replace $\mathbf{A}^T$ as the linear transform from the measurement signal space to the original sparse signal space. It is worth noting that $\mathbf{A}^T$ is selected only for the convenience of derivation in the AMP algorithm. In the training stage of the LAMP network, the linear transform coefficients $\mathbf{B}_t$ of size $N \times M$ in (16) and the nonlinear shrinkage parameters λ_t in (14), (18) and (19) can be optimized. Therefore, given enough training data, the LAMP network can find better shrinkage parameters by leveraging the powerful learning ability of the DNN.

However, the second problem of the AMP algorithm for beamspace channel estimation has not been solved. The conventional AMP algorithm and its corresponding LAMP network only consider the sparsity of signals to be recovered, which are general for any sparse signal recovery problem. In particular, compared with the activation function without an explicit physical meaning in the conventional DNN, the shrinkage function of the LAMP network is not specifically designed for the beamspace channel estimation problem under investigation. The LAMP based beamspace channel estimation schemes still cannot achieve satisfactory estimation accuracy. In order to improve the estimation accuracy, we will utilize the prior distribution of the sparse beamspace channel to propose a more suitable network for the beamspace channel estimation problem in mmWave massive MIMO systems in the next section.

III. Proposed GM-LAMP Network for Beamspace Channel Estimation

In this section, we first derive a new shrinkage function according to the Gaussian mixture distribution of beamspace channel elements. Then, based on the derived shrinkage function, the GM-LAMP based beamspace channel estimation scheme is proposed. After that, we also discuss how to extend the idea of the GM-LAMP network for other sparse signal recovery problems. Finally, the computational complexity analysis between the proposed algorithm and the existing algorithms is provided.

A. Gaussian Mixture Distribution and Its Corresponding Shrinkage Function

As we all know, we are likely to get a more accurate estimate with more prior information of the channel. Next, we will utilize more specific prior distribution (besides sparsity) of the beamspace channel to refine the LAMP network.

There have been some previous works to consider the Gaussian mixture distribution to model the prior distribution of beamspace channel elements for ULAs [13] and for UPAs [38] and verify its validity. Specifically, the probability density function of the element $\tilde{h}$ of the beamspace channel $\tilde{\mathbf{h}}$ can be expressed as:

$$p\left(\tilde{h}; \boldsymbol{\theta}\right) = \sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(\tilde{h}; \mu_k, \sigma_k^2\right), \tag{20}$$

where $\boldsymbol{\theta} = \left\{p_0, \cdots, p_{N_c-1}, \mu_0, \cdots, \mu_{N_c-1}, \sigma_0^2, \cdots, \sigma_{N_c-1}^2\right\}$ is the set of all distribution parameters. N_c is the number of Gaussian components in the Gaussian mixture distribution, p_k is the probability of kth Gaussian component, μ_k and σ_k^2 denote the mean and variance of the kth Gaussian component, respectively. $\mathcal{CN}\left(\tilde{h}; \mu_k, \sigma_k^2\right) = \frac{1}{\pi\sigma_k^2}\mathrm{e}^{-\frac{(\tilde{h}-\mu_k)^*(\tilde{h}-\mu_k)}{\sigma_k^2}}$ denotes the probability density function of the kth Gaussian component. Take the ULA as an example, the rationality of the Gaussian mixture distribution can be explained based on the following two observations.

From (1), (4) and (8), the nth element $\tilde{h}_n$ of the beamspace channel $\tilde{\mathbf{h}}$ can be expressed by

$$\tilde{h}_n = \sqrt{\frac{N}{L}}\sum_{l=1}^{L}\beta_l \mathrm{sinc}\left(\Delta\psi_n\right), \tag{21}$$

where $\Delta\psi_n = \bar{\psi}_n - \psi_l$. Firstly, it is noted that the complex gain β_l follows the complex Gaussian distribution. Secondly, when the practical spatial direction ψ_l for the lth path is close to the predefined spatial direction $\bar{\psi}_n$, $\mathrm{sinc}\left(\Delta\psi_n\right)$ has a large value, which brings the large power for $\tilde{h}_n$. Similarly, when the practical spatial direction ψ_l for the lth path is far away from the predefined spatial direction $\bar{\psi}_n$, $\mathrm{sinc}\left(\Delta\psi_n\right)$ has a small value, which brings the small power for $\tilde{h}_n$. It is due to the random of the practical spatial direction ψ_l that the different $\tilde{h}_n$ can be regarded as the different Gaussian component. So, the Gaussian mixture distribution is expected to model the distribution of the beamspace channel elements.

It is worth noting that when the mean and variance of a Gaussian component are both zero, the probability density function of Gaussian distribution will be changed to

$$\mathcal{CN}\left(\tilde{h}; 0, 0\right) = \delta\left(\tilde{h}\right), \tag{22}$$

where the $\delta\left(\tilde{h}\right)$ is the Dirac delta function, which means the random variable $\tilde{h}$ will be exact zero. Thus, the Gaussian mixture distribution can also describe the sparsity of the beamspace channel as a special case.

Then, we can derive the scalar version η_{gm}: $\mathbb{C} \to \mathbb{C}$ of element-wise Gaussian mixture shrinkage function based on

the the Bayesian minimize mean square error (MMSE) estimation principle [15] as follows:

$$\eta_{\mathrm{gm}} = \mathbb{E}\left\{\tilde{h}\left|r;\boldsymbol{\theta},\sigma^2\right.\right\} = \frac{\int \tilde{h} p\left(r\left|\tilde{h}\right.;\sigma^2\right) p\left(\tilde{h};\boldsymbol{\theta}\right) d\tilde{h}}{\int p\left(r\left|\tilde{h}\right.;\sigma^2\right) p\left(\tilde{h};\boldsymbol{\theta}\right) d\tilde{h}}, \tag{23}$$

where the input element r of the shrinkage function is modeled by [29]

$$r = \tilde{h} + n, \tag{24}$$

where n is the additive Gaussian noise following $\mathcal{CN}\left(0,\sigma^2\right)$. Thus, we have

$$p\left(r\left|\tilde{h}\right.;\sigma^2\right) = \mathcal{CN}\left(r;\tilde{h},\sigma^2\right). \tag{25}$$

Given $p\left(\tilde{h};\boldsymbol{\theta}\right)$ by (20), we have

$$\begin{aligned} &p\left(r\left|\tilde{h}\right.;\sigma^2\right) p\left(\tilde{h};\boldsymbol{\theta}\right) \\ &\quad= \mathcal{CN}\left(r;\tilde{h},\sigma^2\right) \sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(\tilde{h};\mu_k,\sigma_k^2\right) \\ &\quad= \sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(r;\tilde{h},\sigma^2\right) \mathcal{CN}\left(\tilde{h};\mu_k,\sigma_k^2\right) \\ &\quad= \sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(r;\mu_k,\sigma^2+\sigma_k^2\right) \mathcal{CN}\left(\tilde{h};\tilde{\mu}_k\left(r\right),\tilde{\sigma}_k^2\right), \end{aligned} \tag{26}$$

where $\tilde{\mu}_k\left(r\right) = \frac{\sigma^2\mu_k+\sigma_k^2 r}{\sigma^2+\sigma_k^2}$ and $\tilde{\sigma}_k^2\left(r\right) = \frac{\sigma^2\sigma_k^2}{\sigma^2+\sigma_k^2}$.

Finally, by substituting (26) in (23), we can derive a new shrinkage function based on the Gaussian mixture distribution as:

$$\eta_{\mathrm{gm}}\left(r;\boldsymbol{\theta},\sigma^2\right) = \frac{\sum_{k=0}^{N_c-1} p_k \tilde{\mu}_k\left(r\right) \mathcal{CN}\left(r;\mu_k,\sigma^2+\sigma_k^2\right)}{\sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(r;\mu_k,\sigma^2+\sigma_k^2\right)}, \tag{27}$$

where a set of all distribution parameters $\boldsymbol{\theta}$ can also be called as the shrinkage parameters. Compared with the general soft threshold shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ in the existing LAMP network, the Gaussian mixture shrinkage function $\boldsymbol{\eta}_{\mathrm{gm}}$ considering the prior distribution of the beamspace channel is designed for the specific beamspace channel estimation problem.

Now we have derived the Gaussian mixture shrinkage function, based on which we will propose the GM-LAMP network for the beamspace channel estimation in the next subsection.

B. Proposed GM-LAMP Network

In order to estimate the beamspace channel more accurately, we integrate the LAMP network and the new shrinkage function derived from the Gaussian mixture distribution to propose a prior-aided GM-LAMP network.

Specifically, we replace the original soft threshold shrinkage function in the existing LAMP network by the Gaussian mixture shrinkage function. Therefore, the proposed GM-LAMP network is still constructed on the AMP algorithm. Similar to Fig. 2, the GM-LAMP network also have T homogeneous layers, where the inputs and outputs of each layer are the same as those of the LAMP network. The inputs of the tth layer are represented by $\mathbf{y} \in \mathbb{C}^M$, $\hat{\tilde{\mathbf{h}}}_t \in \mathbb{C}^N$ and $\mathbf{v}_t \in \mathbb{C}^M$, where $\mathbf{y}$ is the measurement vector, $\hat{\tilde{\mathbf{h}}}_t$ and $\mathbf{v}_t$ are the outputs of the $(t-1)$th layer. The outputs of the tth layer can be represented by $\hat{\tilde{\mathbf{h}}}_{t+1}$ and $\mathbf{v}_{t+1}$, representing the estimate vector and residual vector of the tth layer, respectively. The difference is that the soft threshold shrinkage function $\boldsymbol{\eta}_{\mathrm{st}}$ of each layer is replaced by the Gaussian mixture shrinkage function $\boldsymbol{\eta}_{\mathrm{gm}}$. To do this, the channel estimate $\hat{\tilde{\mathbf{h}}}_{t+1}$ of the tth layer in the GM-LAMP network can be obtained by

$$\mathbf{r}_t = \hat{\tilde{\mathbf{h}}}_t + \mathbf{B}_t \mathbf{v}_t, \tag{28}$$

$$\hat{\tilde{\mathbf{h}}}_{t+1} = \boldsymbol{\eta}_{\mathrm{gm}}\left(\mathbf{r}_t;\boldsymbol{\theta}_t,\sigma^2\right), \tag{29}$$

where $\sigma^2 = \frac{\|\mathbf{v}_t\|_2^2}{M}$ is obtained in the same way as the AMP algorithm and the LAMP network [29], the linear transform coefficients $\mathbf{B}_t$ and nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ are trainable variables to be optimized in the training phase.

Next, we discuss how the GM-LAMP network works for the beamspace channel estimation problem in mmWave massive MIMO systems. Like most existing DNNs [24]–[26], the GM-LAMP network mainly works in two phases: offline training phase and online estimation phase. In the offline training phase, given a large number of known training data, the GM-LAMP network aims to optimize overall trainable variables $\boldsymbol{\Omega}_{T-1} = \{\mathbf{B}_t,\boldsymbol{\theta}_t\}_{t=0}^{T-1}$ by minimizing the loss function. In the online estimation phase, by inputting the new measurements $\mathbf{y}$, the trained GM-LAMP network can output the estimated beamspace channel $\hat{\tilde{\mathbf{h}}}$. Next, we introduce these two phases in detail.

1) Offline Training Phase: In this article, we adopt the supervised learning to train the GM-LAMP network. The training dataset can be represented $\{\mathbf{y}^d,\tilde{\mathbf{h}}^d\}_{d=1}^D$, where $\mathbf{y}^d$ is the input of the GM-LAMP network, $\tilde{\mathbf{h}}^d$ is the corresponding label, and D represents the number of the training data. In order to avoid overfitting, the layer-by-layer training method adopted by [29] is used to train the GM-LAMP network. Generally speaking, the layer-by-layer training method can be explained from three steps.

Firstly, the whole training procedure can be divided into T sequential training sub-procedures [29]. For the tth training sub-procedure, we aim to refine the trainable variables $\boldsymbol{\Omega}_t = \{\mathbf{B}_i,\boldsymbol{\theta}_i\}_{i=0}^t$ of the $i = 0,\cdots,i = t$th layer. Each layer of the GM-LAMP network has its own loss functions.

Secondly, we define two types of loss functions as follows, which are related to the linear transform operation and the nonlinear shrinkage operation:

$$L_t^{\mathrm{linear}}\left(\boldsymbol{\Omega}_t\right) = \frac{1}{D}\sum_{d=1}^{D}\left\|\mathbf{r}_t^d(\mathbf{y}^d,\boldsymbol{\Omega}_t) - \tilde{\mathbf{h}}^d\right\|_2^2, \tag{30}$$

$$L_t^{\text{nonlinear}}(\boldsymbol{\Omega}_t) = \frac{1}{D}\sum_{d=1}^{D}\left\|\hat{\tilde{\mathbf{h}}}_{t+1}^{d}(\mathbf{y}^d, \boldsymbol{\Omega}_t) - \tilde{\mathbf{h}}^d\right\|_2^2, \quad (31)$$

where $\mathbf{r}_t^d$ is the output of the linear transform operation in (28), and $\hat{\tilde{\mathbf{h}}}_{t+1}^{d}$ is the output of the nonlinear shrinkage operation in (29) (i.e., the estimated channel of the tth layer). Based on these two loss functions, the training sub-procedure for the tth layer are further divided into the two parts: the linear training for aiming to minimizing L_t^{linear} and the nonlinear training for aiming to minimizing $L_t^{\text{nonlinear}}$.

Thirdly, the hybrid method of "individual" and "joint" optimization is further adopted in linear training and the nonlinear training [29]. Specifically, in the linear training of the tth training sub-procedure, only the linear transform coefficients $\mathbf{B}_t$ are first optimized individually, and all trainable variables $\boldsymbol{\Omega}_{t-1}$ of the previous $i = 0, \cdots, i = (t-1)$th layer together with $\mathbf{B}_t$ are optimized jointly. Similarly, in the nonlinear training of the tth training sub-procedure, the nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ are first optimized individually, and then all trainable variables $\boldsymbol{\Omega}_{t-1}$ of the previous $i = 0, \cdots, i = (t-1)$th layer together with $\mathbf{B}_t$ and $\boldsymbol{\theta}_t$ are optimized jointly. Based on the above three steps, the trained GM-LAMP network can be efficiently fine-tuned in each layer, and therefore avoid bad local optimum caused by overfitting [29].

Algorithm 2 Layer-by-Layer Training Method

Initialization: $\mathbf{B}_0 = \boldsymbol{A}^T, \boldsymbol{\theta}_0 = \boldsymbol{\theta}^0$.
1. Learn $\mathbf{B}_0$ to minimize L_0^{linear}
2. Learn $\boldsymbol{\theta}_0$ with fixed $\mathbf{B}_0$ to minimize $L_0^{\text{nonlinear}}$
3. Re-learn $\boldsymbol{\Omega}_0 = \{\mathbf{B}_0, \boldsymbol{\theta}_0\}$ to minimize $L_0^{\text{nonlinear}}$

for $t = 1, \cdots, T-1$ **do**
4. Initialization: $\mathbf{B}_t = \mathbf{B}_{t-1}, \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$
5. Learn $\mathbf{B}_t$ with fixed $\boldsymbol{\Omega}_{t-1}$ to minimize L_t^{linear}
6. Re-learn $\{\boldsymbol{\Omega}_{t-1}, \mathbf{B}_t\}$ to minimize L_t^{linear}
7. Learn $\boldsymbol{\theta}_t$ with fixed $\{\boldsymbol{\Omega}_{t-1}, \mathbf{B}_t\}$ to minimize $L_t^{\text{nonlinear}}$
8. Re-learn $\boldsymbol{\Omega}_t = \{\boldsymbol{\Omega}_{t-1}, \mathbf{B}_t, \boldsymbol{\theta}_t\}$ to minimize $L_t^{\text{nonlinear}}$

end for

Output: $\boldsymbol{\Omega}_{T-1}$.

Algorithm 2 shows the specific layer-by-layer training method. Steps 1-3 represent the training sub-procedure for the initial layer (i.e., $t = 0$), where $\mathbf{B}_0$ and $\boldsymbol{\theta}_0$ are first optimized individually and then optimized jointly. Then, the training sub-procedure for the $t = 1$, $t = 2$, $\cdots$, $t = (T-1)$th layer are performed sequentially. Before training, trainable variables of the tth layer are initialized as the values for those of the $(t-1)$th layer, as shown in Step 4. Steps 5-6 and Steps 7-8 represent the linear training and the nonlinear training of the training sub-procedure for the tth layer, respectively. Step 5 represents the individual optimization of linear transform coefficients $\mathbf{B}_t$, while Step 6 represents the joint optimization of $\boldsymbol{\Omega}_{t-1}$ and $\mathbf{B}_t$. Similarly, Step 7 represents the individual optimization of nonlinear shrinkage parameters $\boldsymbol{\theta}_t$, while Step 8 represents the joint optimization of $\boldsymbol{\Omega}_{t-1}$, $\mathbf{B}_t$ and $\boldsymbol{\theta}_t$.

After overall trainable variables $\boldsymbol{\Omega}_{T-1}$ of T layers are optimized, we can obtain a trained GM-LAMP network to directly estimate the beamspace channel.

2) Online Estimation Phase: In this phase, we apply the trained GM-LAMP network to the beamspace channel estimation problem in mmWave massive MIMO systems, where the new measurements are fed into the trained GM-LAMP network to directly generate the corresponding estimates.

Finally, the normalized mean square error (NMSE) is used to evaluate the performance of the GM-LAMP network:

$$\text{NMSE} = \frac{\mathbb{E}\left\{\sum_{k=1}^{K}\left\|\hat{\tilde{\mathbf{h}}}_k - \tilde{\mathbf{h}}_k\right\|_2^2\right\}}{\mathbb{E}\left\{\sum_{k=1}^{K}\left\|\tilde{\mathbf{h}}_k\right\|_2^2\right\}}. \quad (32)$$

C. Insights From the Proposed GM-LAMP Network

From the discussion above, we can find that in the existing LAMP network [29], the soft threshold shrinkage function only utilizes the sparsity of the signal to be recovered. By contrast, the Gaussian mixture shrinkage function in the proposed GM-LAMP network is derived from the Gaussian mixture distribution, which can approximate the distribution of beamspace channel elements more accurately. With the help of more prior information, the GM-LAMP network is more suitable for the beamspace channel estimation problem.

In this article, we refine the existing LAMP network based on the prior distribution of the beamspace channel to improve the estimation accuracy. This idea can be extended to solve other sparse signal recovery problems in wireless communications with improved performances. If we know the prior distribution of sparse signals, e.g., the sparse active users in massive machine-type communications, the sparse active antennas in spatial modulation systems, and the sparse interfering BSs in ultra-dense networks [39], we can obtain a new shrinkage function based on the new distribution for the DNN, thus the performance can be improved.

Moreover, most existing DNNs, such as the fully connected network, have a generality for a large number of problems, but are not optimized for the specific problems to be solved. For specific problems, by leveraging the domain knowledge (besides the signal distribution considered in this article, e.g., other statistics like mean and variance, the inherent correlation of the signal, etc.), we can design some specialized DNNs for specific problems with better performance.

D. Computational Complexity Analysis

In this subsection, we provide the computational complexity analysis of the proposed GM-LAMP scheme and other existing schemes. Since both the LAMP network and the GM-LAMP network are constructed on the AMP algorithm, the computational complexity of the AMP algorithm, the LAMP network and the GM-LAMP network is the same, i.e., $\mathcal{O}(TMN)$. By contrast, the computational complexity of the OMP algorithm can be represented by $\mathcal{O}(SMN) + \mathcal{O}(S^3M)$, where S is the sparsity level of the beamspace channel vector [40].

IV. Simulation Results

In this section, we present the beamspace channel estimation performance comparison among the proposed GM-LAMP network, the existing LAMP network, and other conventional beamspace channel estimation schemes. In order to prove the effectiveness of our work, we provide the simulation results on the widely used Saleh-Valenzuela channel model and the publicly-available DeepMIMO dataset based on ray-tracing, respectively.

A. Simulation Setup

In our simulations, we consider that the BS equips a $N = 256$ lens antenna array and $N_{\mathrm{RF}}= 16$ RF chains. The number of single-antenna users is set to $K = 16$. The number of measurements is set to $M = 128$. The SNR for uplink channel estimation is defined as $1/\sigma_n^2$. Then, we generate the spatial channel samples according to the Saleh-Valenzuela channel model and the DeepMIMO dataset.

For the the Saleh-Valenzuela channel model in (1), we set the same channel parameters for each user k as follows [7]: 1) $L_k = 3$ path components; 2) $\beta_{k,l} \sim \mathcal{CN}(0,1)$ for $l = 1,2,3$; 3) $\theta_{k,l} \sim \mathcal{U}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$, $\theta_{k,l}^{\mathrm{azi}} \sim \mathcal{U}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ and $\theta_{k,l}^{\mathrm{ele}} \sim \mathcal{U}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ for $l = 1,2,3$. In order to train and test the GM-LAMP network, we generate 80000, 2000 and 2000 samples as the training, the validation and the testing set based on the above setup, respectively.

Next, we introduce how to obtain the channel samples using the DeepMIMO dataset. The DeepMIMO is a parameterized dataset published for deep learning applications in mmWave and massive MIMO systems, described in detail in [41]. The channel samples generated by the DeepMIMO are based on the ray-tracing, which can capture the dependence on key environmental factors such as the environment geometry, operating frequency and so on. One main advantage of the DeepMIMO dataset is that it is completely defined by the ray-tracing scenario and the parameters set. In our simulations, we consider the DeepMIMO dataset with the outdoor ray-tracing scenario 'O1' working at the mmWave 28 GHz and with the parameters set in Table I. Based on the parameters setup in Table I, we can generate about 54000 channel samples between the BS 3 and the single-antenna users from the row R1000 to R1300. We split these 54000 channel samples into three parts: 50000 training samples, 2000 validation samples and 2000 testing samples.

It is noted that after obtaining spatial channel samples based on the Saleh-Valenzuela channel model and the DeepMIMO dataset, we can further generate the corresponding beamspace channel samples according to (4)-(8) and measurement samples according to (9)-(10).

For the proposed GM-LAMP network and the existing LAMP network, the number of the layers is set as $T = 8$, where the number of nodes for each layer is depended on the number of measurements M and the dimension of the beamspace channel N, i.e., $M + N$. The layer-by-layer training method described in detail in Section III-B is adopted to optimize overall trainable variables with the Adam optimizer. We use a mini-batch of 128 training samples for each updating. The training rate for individual optimization is set at 0.001, and for the joint optimization, the training rate decreases to 0.0005, 0.0001 and 0.00001 in turn when the validation errors stops decreasing. After training networks, we evaluate the performance of the trained networks on the test samples. What's more, in the GM-LAMP network, the number of the Gaussian component in the Gaussian mixture shrinkage function is set as $N_c = 4$. Therefore, the nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ of each layer t have 12 elements, which represent the probabilities, means and variances of four Gaussian components. Before the layer-by-layer training, we initialize the nonlinear shrinkage parameters $\boldsymbol{\theta}_0 = \{0.25, 0.25, 0.25, 0.25, 0, 0, 0, 0, 0, 0, 0, 0\}$, where the mean and variance of four Gaussian components are both set as 0 considering the sparsity of the beamspace channel. In the LAMP network, we initialize the nonlinear shrinkage parameter $\lambda_0 = 1$ following [29]. For the OMP-based channel estimation scheme, we consider that the sparsity level of the beamspace channel vector is $S = 24$. For the AMP-based channel estimation scheme, we set the number of iterations as $T = 10$ and the empirical shrinkage parameters as $\lambda_t = 1.1402$ for each iteration t, as did in [29].

TABLE I
THE ADOPTED DEEPMIMO DATASET PARAMETERS

Parameter	Value
Active BSs	3
Active users	From row R1000 to row R1300
Number of BS antennas	$(N_x, N_y, N_z) = (1, 256, 1); (1, 16, 16)$
Antenna spacing	0.5
Number of paths	3

The different training SNR settings have different effects on the performance of the GM-LAMP network. In order to find out what if the preferable SNR settings for the training phase, we provide the NMSE performance comparison of different trained GM-LAMP networks with different training settings, as shown in Fig. 3. Therein, we consider the Saleh-Valenzuela channel model for ULAs. The blue curves represent that all training samples are generated based on the same SNR, while the red curves represent that the training samples are generated based on multiple SNRs. For example, 'SSNR: 5 dB' means all measurement samples for training are generated when the SNR is 5 dB. 'MSNR: 0-10 dB' means that the SNR (in dB) of each sample is randomly drawn from the range $[0, 10]$. From Fig. 4, we can find that the trained network with the single SNR usually achieves good performance near the training SNR, but it is poor at other SNRs. In order to achieve good performance at both high SNRs and low SNRs, we train one network at SNRs between 0 dB and 10 dB for solving channel estimation problems at low SNRs ($0-10$ dB) and another network trained at SNRs between 10 dB and 20 dB for solving problems at high SNRs ($10 - 20$ dB) in the following simulation results.

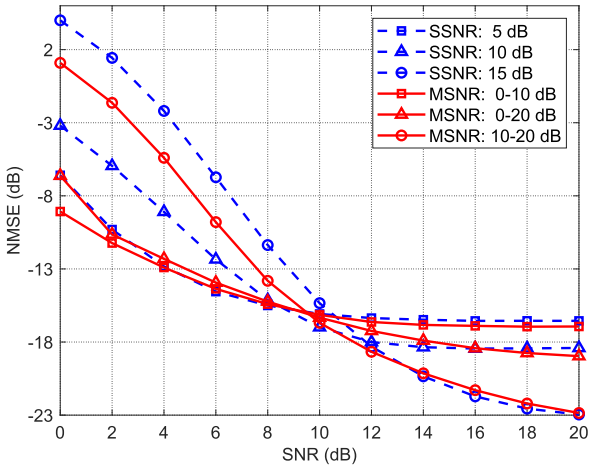

Fig. 3. NMSE performance comparison of different trained GM-LAMP networks with different training settings.

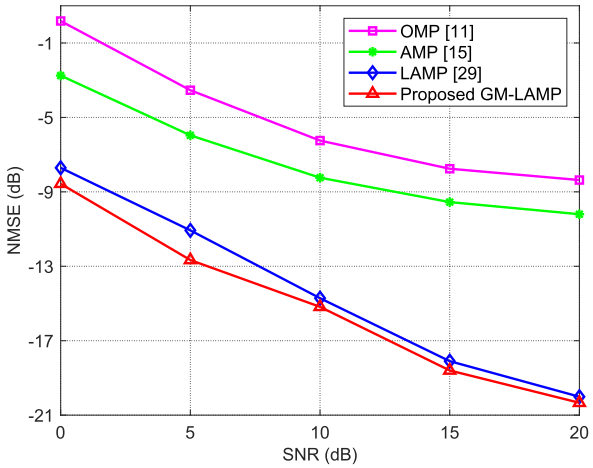

Fig. 5. NMSE performance comparison for UPAs based on the Saleh-Valenzuela channel model.

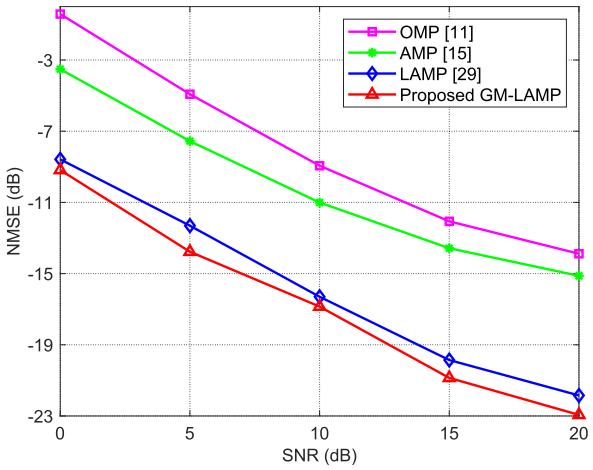

Fig. 4. NMSE performance comparison for ULAs based on the Saleh-Valenzuela channel model.

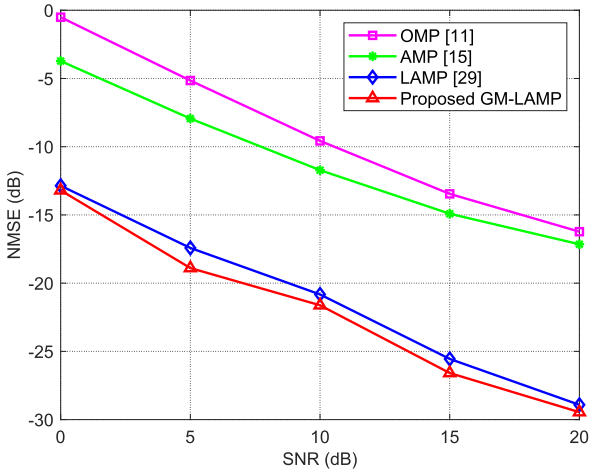

Fig. 6. NMSE performance comparison for ULAs based on the DeepMIMO dataset.

B. Simulation Results on the Saleh-Valenzuela Channel Model

In this subsection, we provide the beamspace channel estimation performance comparison of the OMP algorithm [11], the AMP algorithm [15], the LAMP network [29] and the proposed GM-LAMP network based on the Saleh-Valenzuela channel model.

Fig. 4 shows the NMSE performance comparison of four different schemes mentioned above against different SNRs, where the ULA is considered. We can observe that compared with the other three existing schemes under investigation, the proposed GM-LAMP network enjoys lower estimation errors in all considered SNR regions. In particular, we can observe that the NMSE performance of the OMP algorithm and the AMP algorithm is poor, whereas the two DL based schemes (i.e., the LAMP network and the GM-LAMP network) can achieve better NMSE performance. Moreover, thanks to considering the prior distribution of the beamspace channel, the proposed GM-LAMP network has better channel estimation accuracy than the LAMP network.

In Fig. 5, we further compare the NMSE performance of four different schemes for the 16×16 UPA. We can observe that the conventional OMP algorithm and AMP algorithm cannot achieve satisfactory estimation accuracy. By contrast, the proposed GM-LAMP network can still outperform the other three schemes when the antenna array is UPA.

C. Simulation Results on the DeepMIMO Dataset

In this subsection, we provide the beamspace channel estimation performance comparison of the proposed GM-LAMP network and the other three existing schemes based on the DeepMIMO dataset.

Fig. 6 and Fig. 7 respectively show the NMSE performance comparison of four different schemes against different SNRs for the 256×1 ULA and the 16×16 UPA. We can observe that the proposed GM-LAMP network can still achieve better beamspace channel estimation accuracy based on the more practical DeepMIMO dataset.

D. Other Simulation Results

Based on the above setup that the number of antennas at the BS is $N = 256$ and the number of measurements is

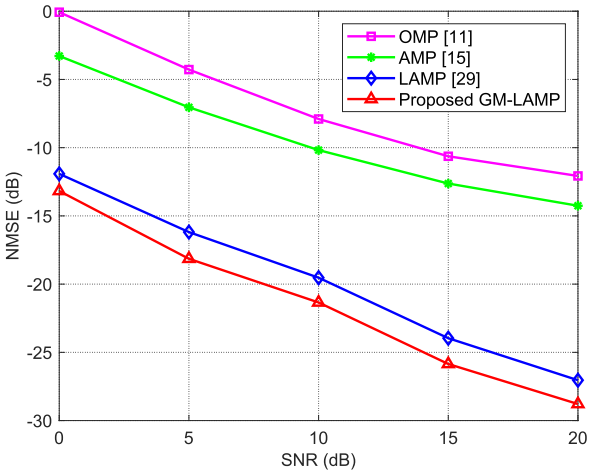

Fig. 7. NMSE performance comparison for UPAs based on the DeepMIMO dataset.

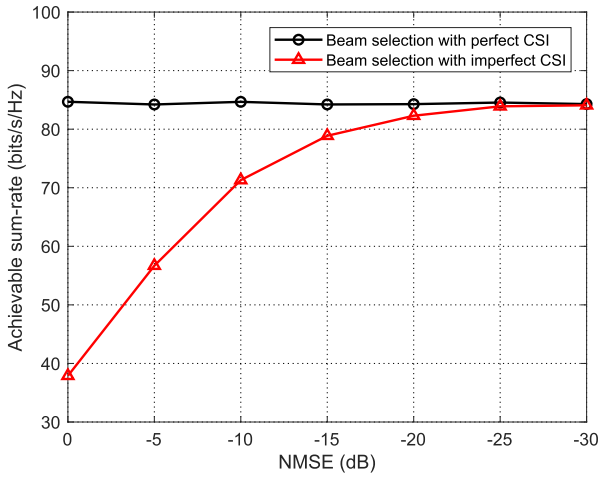

Fig. 9. Sum-rate for beam selection against different NMSE for the beamspace channel estimation.

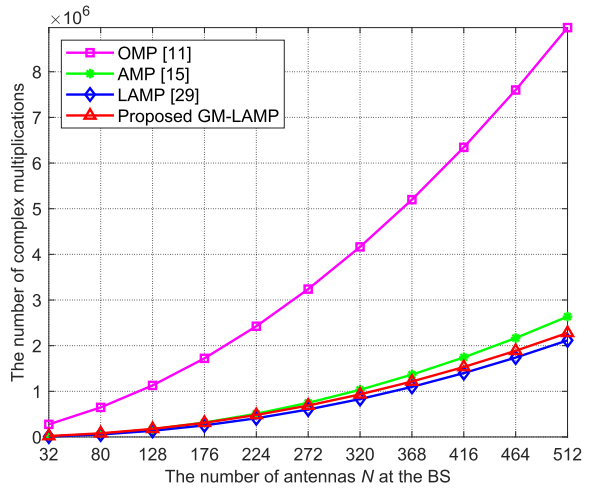

Fig. 8. The number of complex multiplications against the number of antennas N.

$M = 128$, the number of complex multiplications required by four different algorithms can be calculated. The OMP algorithm requires about 2.9×10^6 complex multiplications, and the AMP algorithm requires about 6.6×10^5 complex multiplications. By contrast, the LAMP network and the GM-LAMP network require about 5.3×10^5 and 6.1×10^5 complex multiplications, respectively. In order to show the computational complexity of the proposed scheme and the existing schemes more clearly, we provide the number of complex multiplications comparison against the number of antennas N at the BS, as shown in Fig. 8. It is noted that the number of measurements is set to $M = N/2$.

From Fig. 8, we can find that the conventional OMP algorithm requires more complex multiplications than AMP-type algorithms. What's more, thanks to the powerful learning of DNNs, the LAMP network and the GM-LAMP network can converge faster than the AMP algorithm. Therefore, the number of complex multiplications required by the LAMP network and the GM-LAMP network is smaller than that required by the AMP algorithm [29]. It is noted that the proposed GM-LAMP network employs a more complex Gaussian mixture shrinkage function considering the prior distribution of beamspace channel elements, which slightly increases the number of complex multiplications but achieves better estimation performance compared with the existing LAMP network.

Next, we will evaluate the impact of the NMSE for the beamspace channel estimation on beam selection. In this article, we adopt the interference-aware (IA) beam selection scheme proposed in [9], where the downlink SNR for beam selection is defined as $1/\sigma_d^2$ with σ_d^2 representing the power of the receiving noise at the user side. Fig. 9 shows the sum-rate achieved by the IA beam selection against the NMSE for the beamspace channel estimation. In our simulations, we follow [42] to model the estimated beamspace channel (imperfect CSI) as

$$\hat{\tilde{\mathbf{H}}} = \tilde{\mathbf{H}} + \mathbf{E}, \tag{33}$$

where $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_K]$ represents perfect CSI for K users, $\mathbf{E}$ is the error matrix with entries following the distribution independent and identically distributed (i.i.d.) $\mathcal{CN}(0, \mathrm{NMSE})$. Besides, we consider the IA beam selection with perfect CSI as our benchmark.

In Fig. 9, the ULA based on the Saleh-Valenzuela channel model is considered. We provide the sum-rate comparison achieved by the IA beam selection between imperfect CSI and perfect CSI with the downlink SNR of 10 dB. From Fig. 4, we can observe that the proposed GM-LAMP network can achieve an NMSE of about -23 dB with the reduced pilot overhead by half at the SNR of 20 dB. Fig. 9 shows that compared with the case with perfect CSI, the sum-rate loss caused by the beam selection with imperfect CSI is less than 5% when NMSE is about -23 dB.

In order to show the convergence of the proposed GM-LAMP network, we further provide the simulation result about the NMSE performance against the number of layers. Here, we also consider the ULA based on the Saleh-Valenzuela channel model. The convergence results with different SNRs are presented in Fig. 10, which shows that the GM-LAMP network can reach convergence about at layer $\mathrm{T} = 8$.

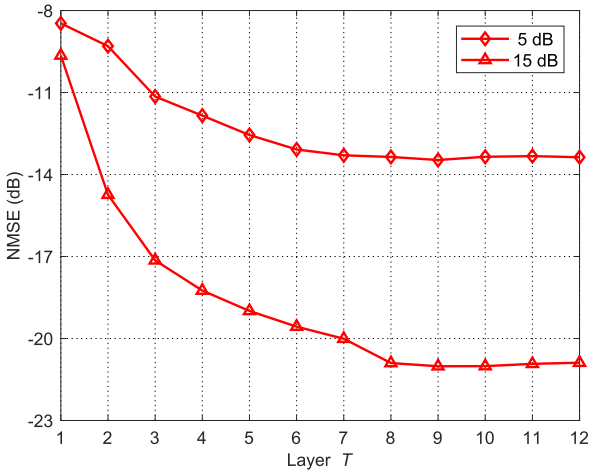

Fig. 10. NMSE performance against the number of layers for the GM-LAMP network.

It is noted that since the orthogonal pilot transmission strategy is adopted among multiple users, pilots from different users can be distinguished without any inter-user interferences. Consequently, the NMSE performance for the beamspace channel estimation has nothing to do with the number of users.

V. Conclusion

In this article, we have proposed a prior-aided GM-LAMP network to solve the beamspace channel estimation problem in mmWave massive MIMO systems. Specifically, we first derive a new shrinkage function by exploiting the Gaussian mixture prior distribution of beamspace channel elements. Different from the original shrinkage function in the existing LAMP network, the derived Gaussian mixture shrinkage function can embody more prior information of the beamspace channel besides sparsity. Then, by integrating the LAMP network with the Gaussian mixture shrinkage function, a GM-LAMP based beamspace channel estimation scheme is developed. To verify the performance of our work, we provide simulation results on the Saleh-Valenzuela channel model and the ray-tracing based DeepMIMO dataset, respectively. Simulation results show that compared with the existing LAMP network and other conventional beamspace channel estimation schemes, the proposed GM-LAMP network considering the prior distribution can achieve better estimation accuracy with a low pilot overhead. We can find by leveraging the domain knowledge of the problems to be solved, the general DNN can be redesigned to improve the performance for the specific problems. For future work, we will follow the idea of the proposed GM-LAMP network to solve the channel estimation problem in terahertz (THz) communications by considering THz channel features.

References

[1] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. New York, NY, USA: Academic, 2016.

[2] B. Ai, A. F. Molisch, M. Rupp, and Z.-D. Zhong, "5G key technologies for smart railways," *Proc. IEEE*, vol. 108, no. 6, pp. 856–893, Jun. 2020.

[3] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[4] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1194–1206, Jun. 2014.

[5] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[6] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[7] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 3679–3684.

[8] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2222, Jun. 2015.

[9] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[10] L. Yang, Y. Zeng, and R. Zhang, "Channel estimation for millimeter-wave MIMO communications with lens antenna arrays," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3239–3251, Apr. 2018.

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

[12] J. Tao, C. Qi, and Y. Huang, "Regularized multipath matching pursuit for sparse channel estimation in millimeter wave massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 169–172, Feb. 2019.

[13] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.

[14] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[15] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. Inf. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[16] X. Zou, F. Li, J. Fang, and H. Li, "Computationally efficient sparse Bayesian learning via generalized approximate message passing," in *Proc. IEEE Int. Conf. Ubiquitous Wireless Broadband (ICUWB)*, Nanjing, China, Oct. 2016, pp. 1–4.

[17] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[18] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.

[19] B. Wu *et al.*, "An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 8, pp. 1289–1300, Dec. 2017.

[20] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.

[21] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided SCMA," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 720–723, Apr. 2018.

[22] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[23] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[24] J.-M. Kang, C.-J. Chun, and I.-M. Kim, "Deep-learning-based channel estimation for wireless energy transfer," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2310–2313, Nov. 2018.

[25] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1228–1231, Aug. 2019.

[26] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[27] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Nanjing, China, Oct. 2019, pp. 266–271.
[28] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.
[29] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.
[30] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.
[31] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[32] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
[33] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
[34] J. J. Fuchs, "On sparse representation in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2002.
[35] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[36] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[37] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
[38] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
[39] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.
[40] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation for hybrid architecture based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
[41] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.
[42] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

Xiuhong Wei received the B.E. degree from the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China, in 2019. She is currently pursuing the M.S. degree in electronic engineering with Tsinghua University, Beijing, China. Her research interests include massive MIMO, millimeter-wave communications, and AI-based wireless communications. She received the National Scholarship in 2016, 2017, and 2018.

Chen Hu (Student Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree in electronic engineering. His research interests include mmWave massive MIMO and reconfigurable intelligent surfaces, with the emphasis on channel estimation. He received the Freshman Scholarship of Tsinghua University in 2012, the Excellent Thesis Award of Tsinghua University in 2016, and IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Award in 2018.

Linglong Dai (Senior Member, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree (Hons.) from the China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, in 2011. From 2011 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. His current research interests include massive MIMO, millimeter-wave communications, THz communications, NOMA, reconfigurable intelligent surface (RIS), and machine learning for wireless communications. He has coauthored the book *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, 2016). He has authored or coauthored over 60 IEEE journal articles and over 40 IEEE conference papers. He also holds 19 granted patents. He received five IEEE Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017-Fall, and the IEEE ICC 2018, the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the *Electronics Letters* Best Paper Award in 2016, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2018, the China Communications Best Paper Award in 2019, and the IEEE Communications Society Leonard G. Abraham Prize in 2020. He is an Area Editor of IEEE COMMUNICATIONS LETTERS and an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is dedicated to reproducible research and has made a large amount of simulation code publicly available.