Deep Learning for Super-Resolution Channel Estimation and DOA Estimation Based Massive MIMO System

Hongji Huang, *Student Member, IEEE*, Jie Yang, Hao Huang, Yiwei Song, and Guan Gui, *Senior Member, IEEE*

***Abstract*—The recent concept of massive multiple-input multiple-output (MIMO) can significantly improve the capacity of the communication network, and it has been regarded as a promising technology for the next-generation wireless communications. However, the fundamental challenge of existing massive MIMO systems is that high computational complexity and complicated spatial structures bring great difficulties to exploit the characteristics of the channel and sparsity of these multi-antennas systems. To address this problem, in this paper, we focus on channel estimation and direction-of-arrival (DOA) estimation, and a novel framework that integrates the massive MIMO into deep learning is proposed. To realize end-to-end performance, a deep neural network (DNN) is employed to conduct offline learning and online learning procedures, which is effective to learn the statistics of the wireless channel and the spatial structures in the angle domain. Concretely, the DNN is first trained by simulated data in different channel conditions with the aids of the offline learning, and then corresponding output data can be obtained based on current input data during online learning process. In order to realize super-resolution channel estimation and DOA estimation, two algorithms based on the deep learning are developed, in which the DOA can be estimated in the angle domain without additional complexity directly. Furthermore, simulation results corroborate that the proposed deep learning based scheme can achieve better performance in terms of the DOA estimation and the channel estimation compared with conventional methods, and the proposed scheme is well investigated by extensive simulation in various cases for testing its robustness.**

***Index Terms*—Massive MIMO, deep learning, channel estimation, DOA estimation, offline training.**

I. Introduction

IN THE future communication scenario, a lot of new protocols and requirements, comprising the basic transmission waveform, the increasing visualization of the network infrastructure, and the need for greatly increased energy efficiency, have been provided for the next generation communication [1]. In particular, the anticipated 1000-fold explosive data traffic growth by 2020 will bring about great challenges in the design of fifth generation communication networks (5G) [2]. Many advanced techniques have been proposed to meet this vital principle, such as massive multiple input multiple output (MIMO) [3]–[5], non-orthogonal multiple access (NOMA) [6], [7], and Millimeter-wave (mmWave) communication [8]. Massive MIMO systems are currently achieving great attention among the academic and industrial community, which is a remarkable technique for theoretically enhancing the capacity of a communication system by simply implementing additional antennas [9]. However, massive MIMO encounters some practical challenges, including the sophisticated channel modeling, the high-dimensional channel state information (CSI), the scheduling of numerous accessing users and the limited radio frequency (RF) chains, etc [10]. Consequently, the performance of the massive MIMO system is degraded without introducing efficient schemes, and its brief advantages may be overwhelmed to some extent.

In the past few years, a lot of works have been devoted to boosting the performance of massive MIMO. In [11], a powerful channel estimation method for 60-GHz indoor systems with the massive uniform rectangular array at base station is proposed, in which knowledge of channel covariances and eigenvalue decomposition of huge matrices are not required. Then, [12] provided a downlink joint spatial division multiplexing (JSDM) scheme, where a multiuser precoder was employed to restrain each user's beamforming vectors within the orthogonal complement of others' channel subspaces. Meanwhile, a beam division multiple access (BDMA) transmission scheme that synchronously serves multiple users via varied beams was developed based on the massive MIMO principles [13]. In order to facilitate bidirectional wireless functionality, a novel model that integrates massive MIMO into full-duplex (FD) communication in which the residual self-interference channels follow the Rician distribution and other channels are Rayleigh distributed was well investigated in [14]. Furthermore, for the sake of investigating 3D massive MIMO experimentally, the authors studied 2D grid antenna channel measurements in UMi, UMa, and O2I scenarios at both 3.5 and 6 GHz with 200 MHz bandwidth and the number of antennas elements varying from 32 to 256 [15]. Additionally, to lower the outage probability of massive MIMO, a cell-edge-aware (CEA) zero forcing (ZF) precoder that exploits the excessive spatial degrees of freedom to constrain inter-cell interference at the users is developed, and

Manuscript received May 11, 2018; revised June 27, 2018; accepted June 27, 2018. Date of publication June 29, 2018; date of current version September 17, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61671253, in part by the Jiangsu Specially Appointed Professor under Grant RK002STP16001, in part by the Innovation and Entrepreneurship of Jiangsu High-level Talent under Grant CZ0010617002, in part by the NUPTSF Grant NY215026, and in part by the "1311 Talent Plan" of Nanjing University of Posts and Telecommunications. The review of this paper was coordinated by Dr. K. Temma. *(Corresponding Author: Guan Gui.)*

The authors are with Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: b14111829@njupt.edu.cn; jyang@njupt.edu.cn; 1017010502@njupt.edu.cn; b15080234@njupt.edu.cn; guiguan@njupt.edu.cn).

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

Digital Object Identifier 10.1109/TVT.2018.2851783

this research indicates the necessity to limit the number of pilot contamination when processing the channel estimation [16].

A. Related Work and Motivation

As the potential gains of massive MIMO rely largely on the perfect CSI, researchers exploit many techniques to realize super-resolution channel estimation. In [17], a unified transmission framework for multiuser time division duplex (TDD)/frequency division duplex (FDD) massive MIMO systems was explored, in which uplink (UL)/downlink channel estimation performance for data transmission was studied. By making full use of structural characteristics of the mmwave beamspace channel, a support detection-based channel estimation strategy with low pilot overhead is employed to estimate the support of sparse beamspace channel [18]. Also, for the purpose of realizing high-precision channel estimation and direction of arrivals (DOA) estimation, frequency synchronization scheme [19], angle domain hybrid precoding and channel tracking method [20], and priori aided channel tracking scheme [21] were reported recently.

Although many recent papers have derived different schemes for super-resolution channel estimation and DOA estimation, these methods require high computational complexity due to the non-linear optimization and fail to leverage the sparsity structure based on massive MIMO. Since the channel characteristics are very complicated in the massive MIMO systems, conventional methods are incapable of capturing the changing of the channel condition in real time. In other words, the CSI acquirement is disturbed and channel/DOA estimation performance is degraded because of the sharply varying channel characteristics. In particular, Device-to-device (D2D)-enabled wireless networks lead to difficult assignment of the radio channels of the nodes induced by partially overlapping channels, which is a severe bottleneck for current channel methods [22]. Meanwhile, as the channel sparsity patterns in existing work are often assumed unknown, nonlinear reconstruction procedures are ineluctable. Thus, conventional techniques are not efficient and reliable enough to achieve super-resolution DOA estimation and signal detection. In recent decades, the promising techniques called machine learning (ML) can be incorporated into the massive MIMO system to realize auto-detection of the CSI. Deep learning [23] concept (i.e., a typical branch of the ML) which was proposed in 2006, is a very powerful tool to handle big data and solve complex nonlinear problems. Some previous work that incorporates deep learning into communication has been roughly studied among channel coding, MIMO, cognitive radio network (CRN) and offloading framework for mobile users [24]–[26]. Also, in [27], the authors provided a new algorithm based on two fast-convergent iterative procedures to design beamforming in multicast wirelessly powered network. Furthermore, deep learning has been applied in traffic control systems to optimize average delay and packet loss rate performance [28]–[30]. For instance, paper [31] designed a efficient deep learning based traffic load prediction algorithm to forecast future traffic load and congestion by realizing better channel assignment. In addition, unmanned aircraft systems (UASs) is an alternative choice for commercial use but it obstacles resource allocation issue[32], and deep learning can be applied to solve this problem. Importantly, brief advantages of the deep learning based communication schemes are demonstrated among the aforementioned work.

B. Main Contributions

In this work, a comprehensive study is conducted to optimize the channel estimation and DOA estimation fields of massive MIMO based on the deep learning technique. The main contributions of this paper are summarized as follows.

1) To the best of our knowledge, we first consider a framework that integrates the deep learning technique into the massive MIMO systems for DOA estimation and channel estimation by leveraging the spatial structure. To be specific, deep neural network (DNN) is adopted and this network is regarded as a blackbox (i.e., the DNN covers the whole massive MIMO system), in which different layers of the network can process specific functions. Thanks to the powerful recognition and representation abilities of the DNN, the sparsity characteristics of the complicated massive MIMO system are acquired through the training procedure.
2) In our work, two high-resolution schemes are proposed to realize DOA estimation and channel estimation in sparse case. Here, the DNN can obtain accurate and real-time CSI through offline training, which is dedicated to detect the channel characteristics and spacial structures in angle domain.
3) We provide performance analysis of the proposed deep learning methods of massive MIMO in different cases. Specifically, we simulate mean square error (MSE) of the arrival angle and bit error ratio (BER) for assessing the accuracy of the DOA estimation and the channel estimation. Also, extensive simulation results and comparison have demonstrated the efficiency and robustness of the proposed schemes.

The remainder of this paper is organized as follows. In Section II, we propose a scheme that incorporates DNN into the massive MIMO system, in which different activation functions are employed in different layers. Then, to achieve high-resolution channel estimation, a novel approach is provided based on the deep learning in Section III. Thereafter, we describe a deep learning based algorithm for DOA estimation in Section IV. Numerical results for evaluating the performance of the proposed schemes are provided in Section V, which is followed by conclusions in Section VI.

Notations: Vectors are defined by boldface small letters, while matrices are noted by boldface capital letters; superscripts $(\cdot)^{\star}$, $(\cdot)^{T}$, $(\cdot)^{H}$ and $\|\cdot\|$ represent conjugate, transpose, Hermitian and the Frobenius norm operator, respectively. Also, $\bigoplus$ and $\mathbb{C}^{m\times n}$ are denoted as the Kronecker product and the vector space of all $m \times n$ complex matrices. Furthermore, $\mathbf{E}[\cdot]$ is given as the expectation operator. Additionally, diag$(\boldsymbol{x})$ and $j = \sqrt{-1}$ represent a diagonal matrix with main diagonal of $\boldsymbol{x}$ and the imaginary unit, respectively.

II. System Model and Problem Formulation

A. Massive MIMO System Model

We consider a typical massive MIMO UL system, where one base station (BS) with a uniform linear array (ULA) of N_t antennas and K single-antenna users are designed. Here, the

BS is assumed to have no information on each path of the user. Furthermore, we introduce the classical narrowband ray-based channel model [33], and the UL model of k-th user can be given as

$$\mathbf{h}_k = \sum_{i=1}^{P} g_{k,i}\mathbf{a}_t(\theta_{k,i}) = \mathbf{A}_{t,k}\mathbf{g}_k, \tag{1}$$

where P and $g_{k,i}$ are denoted as the number of resolvable paths from the BS to the k-th user and the complex gain of the i-th path of the k-th user, respectively. $\theta_{k,i}$ is noted as the physical DOA of the i-th path at the k-th user. Also, the steering vector $\mathbf{a}_t(\theta_{k,i})$ is defined as the array response of the i-th path at the BS. For a ULA, $\mathbf{a}_t(\theta_{k,i})$ can be expressed as

$$\mathbf{a}_t(\theta_{k,i}) = \frac{1}{\sqrt{N_t}}\left[1, e^{-j2\pi\frac{d}{\lambda}\sin\theta_{k,i}}, \ldots, e^{-j2\pi\frac{d}{\lambda}(N_t-1)\sin\theta_{k,i}}\right]^T, \tag{2}$$

Here, d represents the antenna spacing, while λ is defined as the wavelength of the carrier frequency. Meanwhile, $\mathbf{A}_{t,k} = [\mathbf{a}_t(\theta_{k,1}), \mathbf{a}_t(\theta_{k,2}), \ldots, \mathbf{a}_t(\theta_{k,P})] \in \mathbb{C}^{N_t\times P}$ and $\mathbf{g}_k = [g_{k,1}, g_{k,2}, \ldots, g_{k,P}]^T \in \mathbb{C}^{P\times 1}$. Furthermore, the uplink channel matrix is $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_k] \in \mathbb{C}^{N_t\times K}$.

In order to limit user-to-user interference, we design a digital precoding based on the concept of zero-forcing (ZF) approach. Assuming the source vector as $\mathbf{s} \in \mathbb{C}^{K\times 1}$ and the digital precoding as $\mathbf{P} \in \mathbb{C}^{N_t\times K}$, the transmitted signal vector can be given by

$$\mathbf{x} = \mathbf{P}\mathbf{s}. \tag{3}$$

Then, the received signal at the BS can be formulated as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4}$$

where $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_K)$ is additive white Gaussian noise (AWGN). Also, it is noted that the source vector is normalized with $\mathbf{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$.

From (1), we observe that the complex gain information and the DOA information are two key characteristics that model the channel matrix. To achieve super-resolution channel estimation, the proposed deep learning based framework first obtains the physical DOA information (i.e., DOA estimation), and the DOA information is incorporated into the complex gain information to realize the channel estimation. This estimation mechanism makes full use of the sparsity characteristics of the massive MIMO which is dedicated by the state-of-the-art deep learning concept.

B. Review of Deep Learning

The modern term "deep learning" that goes beyond the neuroscientific perspective inspired by the ML technique, is considered as a better principle of learning multiple levels of composition [34]. The simplest deep learning models is linear model, which is expressed as

$$f(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} v_i w_i, \tag{5}$$

where $\mathbf{w}$ represents the weight of the network, and this model is designed to take n input values as $\{v_1, v_2, \ldots, v_n\}$. Also, $f(\cdot)$ represents the output of the model, which can classify two different branches by identifying whether $f(\mathbf{v}, \mathbf{w})$ is positive or negative. Then, motivated by the basic idea of more computational units can facilitate intelligent interaction manners, many works develop new models that consist of multiple units. The modern convolutional network for processing images is designed [35], which has been verified as an effective tool to deal with complex learning tasks. Lately, a large breeds of deep learning architectures are based on a model neuron named the rectified linear unit.

Another milestone of the deep learning was the successful use of back-propagation (BP) to train DNN with internal representations [36]. Inspired by this movement, the long short-term memory (LSTM) algorithm [37] was developed and it elevates the performance of the natural language processing (NLP). Also, Kernel machines [38] and graphical methods [39] both perform well in many important tasks.

In recent decades, Geoffrey Hinton exhibited that a kind of neural network called deep belief network can be optimized using a strategy named greedy layer-wise pretraining [23], and it brings a breakthrough in deep learning area among research community. Other research groups quickly showed that the same method can be adopted to train many other kinds of DNN [40] and systematically enhance generalization on test examples. Synchronously, some works have emphasized that DNN outperforms competing artificial intelligence (AI) systems based on other ML methods as well as hand-designed functionality. In our work, we firmly believe that the promising DNN can boost the performance of massive MIMO in terms of channel estimation and DOA estimation.

III. Deep Learning Based DOA Estimation Scheme

In this section, we propose a framework where deep learning is integrated into the massive MIMO system as an end-to-end method for super-resolution DOA estimation and signal detection. Some conventional methods have been presented for DOA estimation, such as estimation of signal parameters via rotational invariance technique (ESPRIT) [41] and multiple signal classification (MUSIC) [42]. However, paper [43] indicated that complicated eigen-decomposition plays an essential role in these subspace based methods, which is prohibited for massive MIMO system. Different from the scheme proposed in [43], the proposed approach regards the whole system as a DNN and achieves end-to-end learning, which promotes the performance of the DOA estimation. The powerful learning capacity of deep learning can deal with this problem through mapping operations and training, in which the massive MIMO system can be regarded as a black box.

A. DNN Learning Framework

In the deep learning area, DNN is considered as one of the most popular generative models. According to the well-known *universal approximation theorem* [44], it is noted that a feed-forward network with a single hidden layer processed by multilayer perception mechanism can approximate continuous functions on compact subsets of R^n. As a multilayer processor, the DNN is capable of dealing with many non-convex and non-linear issues. Based on the massive MIMO system, high computation complexity is a key limit in terms of channel esti-

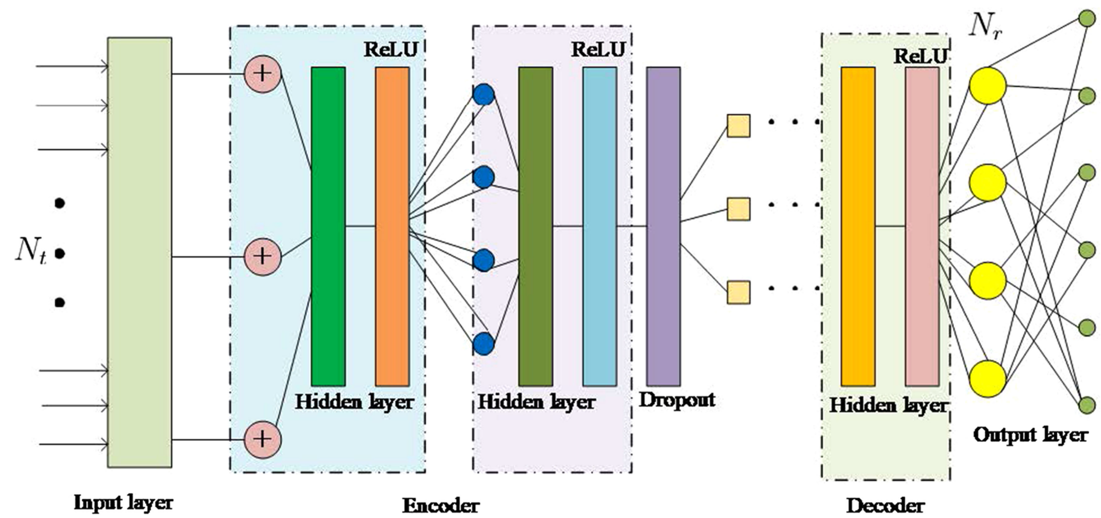

Fig. 1. DNN based massive MIMO framework (All the circles and squares represent neurons.)

mation and DOA estimation. Through non-linear operation and propagation, this problem can be settled with the DNN. Also, the multilayer perception mechanism and special training policy promote the DNN to be a commendable tool to leverage the sparsity characteristics of the massive MIMO.

Here, let us introduce the DNN briefly. DNN, which can be regarded as a deeper version of artificial neural networks (ANN), is designed as a neural network with many hidden layers. In particular, there are multiple neurons implementing in each hidden layer, as well as an output with weighted sum of these neurons operated by a nonlinear function. In order to realize recognition and representation operation, the DNN is processed by activation. In general, the Sigmoid function and the ReLU function are the most universal choices in the nonlinear operation, which can be given by

$$f_{\text{Sigmoid}}(x) = \frac{1}{1+e^{-x}}, f_{\text{ReLU}}(x) = \max\,(0, x), \tag{6}$$

where x is denoted as the argument of the function. Assuming that the output of the DNN is $\mathbf{o}$, and $\mathbf{x}$ represents the input signal of the massive MIMO, we obtain

$$\mathbf{o} = f(\mathbf{x}, w) = f^{(n-1)}\left(f^{(n-2)}\left(\ldots f^{1}(\mathbf{x})\right)\right), \tag{7}$$

Here, n and w are the amount of layers of the DNN and the weights of the DNN, respectively.

As shown in Fig. 1, in the proposed DNN framework, we design a L-dimension input layer, which is equal to the length of each training sequence. It is a dense layer with 256 neurons, serving as the input of the transmitted signal vectors $\mathbf{x}$ and propagating characteristics to the first hidden layer. The second hidden layer and the third hidden layer are designed to encode and learn the features of the signals based on the output of the input layer comprising 512 neurons and 400 neurons, which is employed for determining how many bits should be encoded into a given block and achieving optimized antennas selection by capturing the system sparsity. For the sake of averting overfitting, we deploy a dropout layer with retaining probability p. Then, we design the next layer as noise layer with 256 neurons for corrupting the processed signal by the AWGN. It simply adds AWGN or other distortion to the network so that different output based on the same input is obtained. Thereafter, the

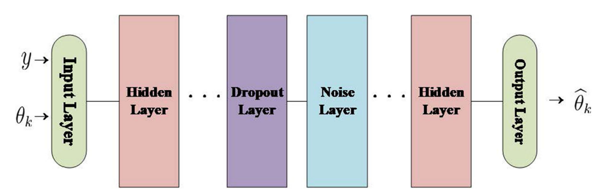

Fig. 2. Deep learning based learning framework for DOA estimation.

last hidden layer is designed with 128 neurons and 64 neurons, which acts as a decoder. The decoder is adopted for reconstructing the original input by minimizing the difference between the input and the output of the DNN. Additionally, the output layer is a linear layer, providing the estimated information based on the massive MIMO system. Besides, it is worth indicating that the ReLU function is adopted as the activation function in the input layer, noise layer and all the hidden layers, whereas the output layer is processed by the Sigmoid function.

B. Learning Policy for DOA Estimation

The proposed deep learning based framework can be regarded as a mapping function, and a learning policy is proposed for learning the sparsity characteristics of the massive MIMO UL system, as shown in Fig. 2. The network framework presented in Fig. 2 is as same as the proposed DNN model. In the first stage, we consider an offline learning scheme for training the DNN. As indicated above, the corresponding received signal vector $\mathbf{y}$ can be acquired based on the fixed channel matrix $\mathbf{H}$ in different direction, which is used as the training samples. To be specific, every time, once a transmitted signal vector is added in the channel, the corresponding $\mathbf{y}$ is obtained in a special direction. Based on this manner, the received signals can be obtained in all direction. In the past decades, many channel models that well depict the channel characteristics have been proposed, which make it possible to realize DOA estimation in almost all the channel condition. Thus, we can obtain received signals $\mathbf{y}$ of the massive MIMO system in all direction based on different wireless channel. Synchronously, the physical DOA θ_k can be generated randomly to form a training dataset with the received signal vectors $\mathbf{y}$, which is the training samples of the DNN. In the second stage, online deployment is conducted by giving a specific channel model and the estimated DOA can be obtained without requiring iterations. Apparently, the essential part of the learning policy is the first stage, which will be illustrated in detail as follows. For the purpose of obtaining the estimated DOA θ_k, a loss function based on the mean square error concept is given by

$$l_2 = \mathbb{E}\left\{\left\|\theta_k - \widehat{\theta}_k\right\|^2\right\} = \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\left\|\theta_k - \widehat{\theta}_k\right\|^2, \tag{8}$$

where N is noted as the number of samples, and $\widehat{\theta}_k$ is the prediction. To enhance the generalization ability of the proposed scheme, Eq. (8) can be reformulated as

$$loss = \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\left\|\theta_k - \widehat{\theta}_k\right\|^2 + \beta\Omega(\psi) = \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\left\|\theta_k - \widehat{\theta}_k\right\|^2 + \frac{\beta}{2}w^T w. \tag{9}$$

Here, β is defined as a hyperparameter that weights the relative contribution of the norm penalty term, and Ω is the standard objective function. Also, ψ is noted as the parameter of the model. Specifically, w represents all the weights that must be affected by a norm penalty, while variable ψ includes w and unregularized parameters.

Then, we adopt the stochastic gradient decent (SGD) algorithm with momentum to optimize the loss function based on the proposed DNN framework. Here, the SGD iteration can be expressed as

$$\psi^{m+1} = \psi^m + v, \tag{10}$$

where v represents as the velocity which accumulates the gradient element. Also, m is the iteration number, and ψ^0 is denoted as the initial parameter. To be specific, the update process of v can be formulated as

$$v = \alpha v - \epsilon g = \alpha v - \epsilon\frac{1}{N}\nabla_{\theta_k}\sum_{j=1}^{N}\left\|\theta_k - \widehat{\theta}_k\right\|^2, \tag{11}$$

where α is denoted as the momentum parameter, while ϵ is noted as the learning rate. Meanwhile, g represents the gradient element. In order to evaluate the performance of the proposed deep learning based DOA estimation scheme, the mean square error (MSE) is introduced in the proposed scenario. According to the principles of the MSE, we obtain

$$\mathrm{MSE_d} = \frac{1}{N}\sum_{j=1}^{N}\left\|\theta_k - \widehat{\theta}_k\right\|^2. \tag{12}$$

C. Robustness Analysis of the Proposed DOA Estimation Scheme

In addition, as a deep learning based scheme, robustness of the proposed algorithm is required to be investigated. One question is whether or not the proposed deep learning based scheme relies on deep networks with multiple hidden layers. Also, it is still confused about the impact induced by regularization item on the performance of the DNN. To answer these questions, without changing the implementation of the noise layer, we perform experiments on less hidden layers of the DNN and regularization item are adopted for fair comparison, which are illustrated in Table I.

TABLE I
TEST ACCURACY FOR DIFFERENT NUMBER OF HIDDEN LAYERS AND REGULARIZATION TERM

Hidden Layers	Method	Test Accuracy(%)
3	DNN	99.905
	DNN + L^2 Regularization	99.905
4	DNN	99.955
	DNN + L^2 Regularization	99.955
5	DNN	99.982
	DNN + L^2 Regularization	99.990
6	DNN	99.990
	DNN + L^2 Regularization	99.998

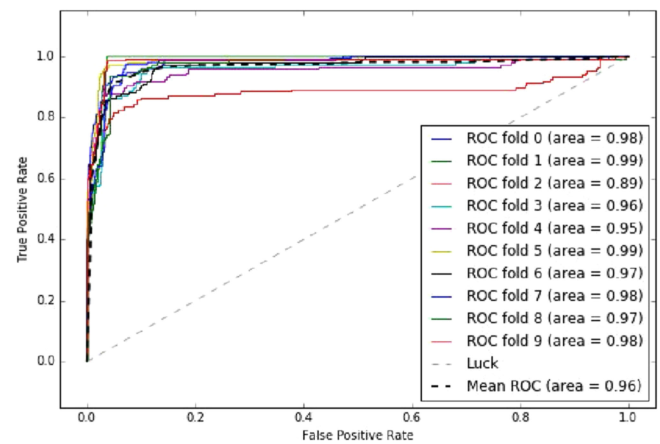

Fig. 3. AUC curve for the proposed DNN when the number of hidden layers is 4.

As we can see from Table I, the deep learning based scheme can achieve further improvement with the aids of the L^2 regularization terms method, for the reason that the regularization term can hinder the overfitting risk and promote generalization performance. Also, it is observed that the test accuracy is not being degraded when reducing the number of hidden layers, which demonstrates that the proposed deep learning based scheme is robust and high-efficiency. On the other hand, the area under the receiver operating characteristic (AUC) curve is a common index for evaluating a model. As exhibited in Fig. 3, it can be observed that the AUC is 0.96, indicating that the proposed scheme does not occur overfitting and its testing accuracy is high. Hence, the proposed deep learning based scheme is very effective and reliable in DOA estimation in the UL massive MIMO system.

Then, to assess the computational complexity of the proposed deep learning based scheme, the operation time during the second stage (i.e., testing stage) is presented in the cases with different number of the hidden layers. It can be observed from Table II that the computational complexity of the proposed scheme is low and this index is almost invariant with

TABLE II
OPERATION TIME FOR DIFFERENT NUMBER OF HIDDEN LAYERS AND REGULARIZATION TERM

Hidden Layers	Method	Operation time(s)
3	DNN	0.0191
	DNN + L^2 Regularization	0.0192
4	DNN	0.0194
	DNN + L^2 Regularization	0.0194
5	DNN	0.0201
	DNN + L^2 Regularization	0.0200
6	DNN	0.0204
	DNN + L^2 Regularization	0.0205

the changing of the hidden layers. Meanwhile, the operation time performance is not been degraded when adding regularization term. Importantly, the operation time in the conventional approaches is longer than the proposed scheme, such as the ADMA user scheduling method (0.582 s) [20], and PA channel tracking scheme (0.807 s) [21]. Hence, the robustness of the proposed deep learning based scheme is beyond our expectation.

D. DOA Tracking Based Deep Learning technique During Data Transmission

During the UL transmission, we can track the DOA information and spatial signatures of the massive MIMO system for reducing training overhead. As a deep learning based scheme, the channel characteristics of the massive MIMO system have been captured after training stage. The wireless channel model can be treated as constant spanning, which consists of many data blocks or data groups. In the proposed scheme, the training sequences can be separated into multiple data blocks in our proposed DNN based strategy. In this way, the error of the DOA estimation can be tracked in each data block.

To realize DOA tracking, the block index is denoted as b, and the scheduled received data of b-th block can be formulated as

$$\mathbf{y}^{(b)} = \mathbf{H}\mathbf{x}^{(b)} + \mathbf{n}^{(b)}, \tag{13}$$

where $\mathbf{x}^{(b)}$ is the UL transmitted data of the b-th block. After that, noting that T as the time interval between one block and the next block, the DOA rate of the b-th block can be given by

$$\dot{\theta}_{k,i}^{(b)} = \frac{\widehat{\theta}_{k,i}^{(b)} - \widehat{\theta}_{k,i}^{(b-1)}}{T}. \tag{14}$$

Here, $\widehat{\theta}_{k,i}^{(b)}$ is denoted as the estimated DOA of the b-th block of the k-th user. We assume the number of samples transmitted during T as M, $\widehat{\theta}_{k,i}^{(b)}$ can be calculated by $\widehat{\theta}_{k,i}^{(b)} = \frac{1}{T}\sum_{j=1}^{M} \widehat{\theta}_{k,i,j}$. Similarly, all the DOA of all users can be tracked with the aids of the proposed deep learning based scheme.

Based on Eq. (13), we further illustrate the method for DOA tracking in the deep learning based massive MIMO system. Initially, the DOA rate of the 0-th block is set as $\dot{\theta}_{k,i}^{(0)} = 0$. Then, the estimated DOA of 1-st block can be obtained by processing the proposed deep learning DOA estimation scheme. According to Eq. (13), the DOA rate can be formulated in each block. Let us consider the acceleration manner, the DOA acceleration of the b-th block of the k-th user can be given as

$$\begin{aligned}\ddot{\theta}_{k,i}^{(b)} &= \frac{\dot{\theta}_{k,i}^{(b)} - \dot{\theta}_{k,i}^{(b-1)}}{T} \\ &= \sqrt{\frac{2 \times \left(\widehat{\theta}_{k,i}^{(b)} - \widehat{\theta}_{k,i}^{(b-1)}\right)}{T^2}}.\end{aligned} \tag{15}$$

Therefore, combined with the proposed deep learning based scheme, the DOA tracking of each block can be realized. Based on the analysis above, the DOA tracking and the users' spatial signatures can be detected automatically until the distance between two signatures is less than the guard interval. Although the proposed deep learning based scheme has been trained based on almost all the channel models, it is still required to conduct additional small-scale training at times because some rare terrible distortion has not been considered in existing channels models. Apparently, DOA tracking is an effective mean for preventing unexpected great distortion to the channel, which can ensure the precision of the DOA estimation and signal detection of the proposed framework. With the aids of the DOA tracking method, once the massive MIMO encounters unexpected distortion, we can implement additional small-scale training in time to correct the DOA estimation manner, bringing great benefits in improving the precision of the DOA estimation.

IV. DEEP LEARNING BASED CHANNEL ESTIMATION SCHEME

In this section, we propose a deep learning based scheme in massive MIMO system to achieve super-resolution channel estimation. We consider the sparse case, in which the covariance of the channel matrix is low-rank. To be specific, based on the learning scheme proposed in Section III-A, the channel estimation method is presented based on the deep learning framework in the massive MIMO system.

A. Complex Gain Estimation

It is observed that the channel matrix is determined by the DOA information and the complex gain. In Section III, the high-resolution DOA estimation is achieved and how to realize super complex gain estimation is becoming a remaining issue for channel estimation. Let the j-th estimated output signal vector be denoted by $\widehat{\mathbf{y}}_j$, the loss function is given as

$$loss_{\text{signal}} = \frac{1}{N}\sum_{j=1}^{N} \|\mathbf{x}_j - \widehat{\mathbf{y}}_j\|^2, \tag{16}$$

where x_j represents the j-th transmitted signal vector. The aim of our model is to minimize the difference between the output of the DNN and the transmitted signals. Interestingly, the proposed DNN learning framework for DOA estimation is adopted in complex gain estimation, which is used to extract the useful characteristics of the input signals. Then, for the purpose of eliminating the effect of the random noise, the estimated output signal is reformulated as

$$\begin{aligned}\bar{\mathbf{y}}_j &= \widehat{\mathbf{y}}_j - \widehat{\mathbf{n}}_j \\ &= \widehat{\mathbf{H}}_j\widehat{\mathbf{z}}_j.\end{aligned} \tag{17}$$

Algorithm 1: DNN Based Training Algorithm for Channel Estimation.

Input: Environment simulator, transmitted signal vector $\mathbf{x}$.
Output: DNN.
1: Start the environment simulator to generate wireless channel, and mix specific man-made noise or distortion into the channel.
2: Generate a set of training sequences, in which every 24 bits of the transmitted signal $\mathbf{x}$ are grouped.
3: Design the proposed DNN framework. Also, set the learning rate and the loss rate, as well as the weight as $w = 0$. Furthermore, set the error threshold as $\gamma = 10^{-6}$.
4: **while** error $\geq \gamma$**:** Train the DNN based on the given sequences according to the proposed learning policy by the SGD.
5: Update the weight w and the output of each layer of the DNN.
6: **end while**
7: **return:** DNN.

Here, $\widehat{\mathbf{n}}_j$ and $\widehat{\mathbf{H}}_j$ represent the estimated noise and channel matrix of j-th sample, respectively. Also, $\widehat{\mathbf{z}}_j$ is defined as j-th unit test signal (i.e., the strength and magnitude of $\widehat{\mathbf{z}}_j$ have no impact on the estimated output signal). Hence, the j-th estimated channel matrix can be given by

$$\bar{\mathbf{y}}_j = \widehat{\mathbf{H}}_j = \widehat{\mathbf{A}}_{t,j}\widehat{\mathbf{G}}_j. \tag{18}$$

In Eq. (18), $\widehat{\mathbf{A}}_{t,j}$ can be obtained based on the proposed DOA estimation scheme since it is formed by the physical DOA in massive MIMO. Moreover, the estimated complex gain matrix $\widehat{\mathbf{G}}$ is represented as

$$\widehat{\mathbf{G}} = \begin{bmatrix} \widehat{g}_{1,1} & \widehat{g}_{2,1} & \cdots & \widehat{g}_{K,1} \\ \widehat{g}_{1,2} & \widehat{g}_{2,2} & \cdots & \widehat{g}_{K,2} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{g}_{1,P} & \widehat{g}_{2,P} & \cdots & \widehat{g}_{K,P} \end{bmatrix}. \tag{19}$$

Then, with the aids of algebraic transformation, we obtain

$$\widehat{\mathbf{G}}_j = \widehat{\mathbf{A}}_{t,j}^{-1}\bar{\mathbf{y}}_j. \tag{20}$$

Apparently, the estimated complex gain of the k-th user $\widehat{\mathbf{g}}_k = [\widehat{g}_{k,1}, \widehat{g}_{k,2}, \ldots, \widehat{g}_{k,P}]^T$ is obtained. Therefore, the complex gain is estimated based on the proposed DNN in the massive MIMO system.

B. The Proposed Deep Learning based Channel Estimation

We consider a channel estimation method in the massive MIMO system. For the sake of enhancing the performance of the channel estimation, we acquire the DOA information and the complex gain information independently, which is motivated by the fact that less variables reduce estimation bias. As a DNN based strategy, the learning approach is divided into training stage and testing stage generally, which is illustrated in **Algorithm 1** and **Algorithm 2**.

Thereafter, in order to evaluate the performance of the proposed deep learning based channel estimation scheme, the MSE

Algorithm 2: DNN Based Testing Algorithm for Channel Estimation.

Input: Environment simulator, DNN, estimated DOA.
Output: $\widehat{\mathbf{H}}$.
1: Load the DNN which has been trained thoroughly.
2: Start the environment simulator to generate wireless channel, and mix specific man-made noise or distortion into the channel.
3: Process the DNN.
4: Update the output $\widehat{\mathbf{y}}$ of the DNN.
5: Compute $\widehat{\mathbf{g}}_k$ by equation (17), (18), and (20).
6: Calculate the estimated channel vector $\widehat{\mathbf{h}}_k$ of k-th user as follows

$$\widehat{\mathbf{h}}_k = \widehat{\mathbf{A}}_{t,k}\widehat{\mathbf{g}}_k, \tag{21}$$

7: Obtain the estimated channel matrix $\widehat{\mathbf{H}}$ based on the $\widehat{\mathbf{h}}_k$.
8: **return:** $\widehat{\mathbf{H}}$.

is introduced in the proposed framework. According to the principles of the MSE, we obtain

$$\mathrm{MSE_c} = \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\frac{\left\|\mathbf{h}_{k,j} - \widehat{\mathbf{h}}_{k,j}\right\|^2}{\|\mathbf{h}_{k,j}\|^2}. \tag{22}$$

V. Simulation Results and Analysis

In this section, the performance of the DOA estimation and the channel estimation of the proposed deep learning based schemes are well investigated. The powerful *Tensorflow* is introduced to design and process the DNN. Also, we deploy a GPU-based server, which is equipped with 4 Nvidia GPUs and its CPU is E5-2683 v3. For implementing the proposed deep learning based schemes on GPUs, we treat the whole mini-batch as a "big training example", where the operation of the proposed DNN framework for DOA estimation and channel estimation is on the basis of the averaged values of all training samples in the mini-batch. Without loss of generality, we consider a typical massive MIMO system, in which the BS is equipped with $N_t = 128$ antennas and $K = 32$ randomly distributed single-antenna users are linked by $P = 100$ resolvable paths. Meanwhile, $d = \frac{\lambda}{2}$ and the complex channel gain follows $\mathbf{n} \sim \mathcal{CN}(0, 1)$. Furthermore, the DOA $\theta_{k,i}$ is randomly distributed in the space $[-\frac{\pi}{2}, \frac{\pi}{2}]$, and all reported simulation results are averaged over 50000 iterations. To collect the training samples, we transmit unit signals to the system model in different direction and obtain the corresponding received signals. To be specific, each group of received signals are obtained in a specific direction (i.e., each DOA represents one direction), and the sampling interval is 0.01° in the range of $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Hence, the training samples are formed by the direction information $\theta_{k,i}$ and the corresponding received signals. Additionally, the training set comprises 150000 examples in the simulation, while a validation set with 20000 samples for each SNR.

A. DOA Estimation

Fig. 4 shows the MSE performance of the DOA estimation against SNR of the proposed deep learning (DL) based massive MIMO scheme with different length L of training sequences,

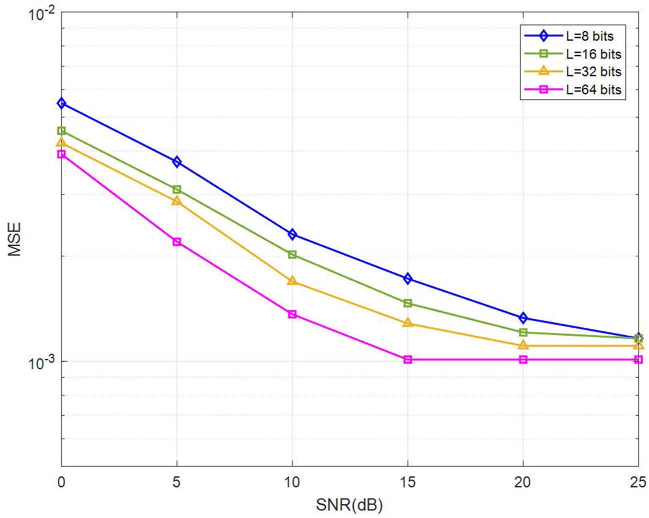

Fig. 4. MSE performance of the DOA estimation with different length L of sequences of the proposed DL based scheme.

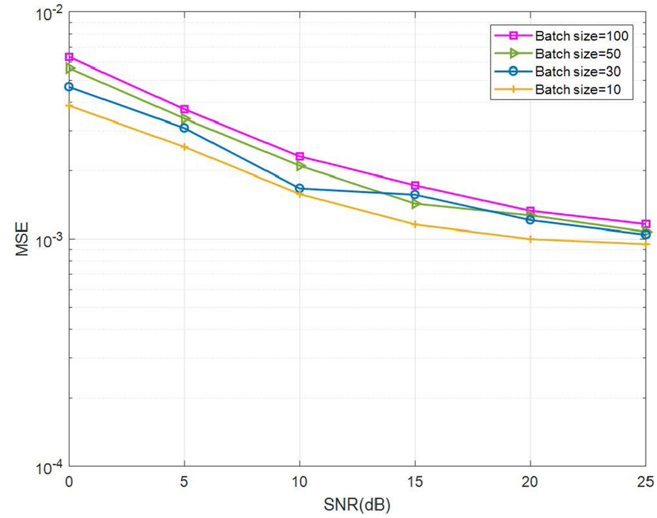

Fig. 6. MSE performance of the DOA estimation of the proposed DL based scheme in the case of "batch size = 100," "batch size = 50," "batch size = 30," and "batch size = 10."

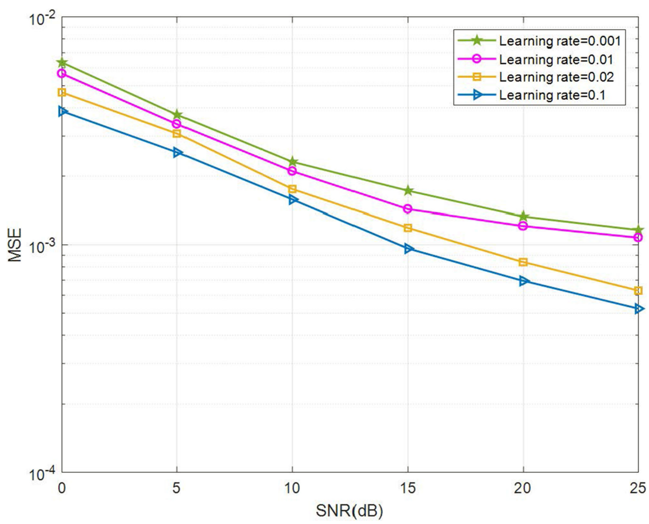

Fig. 5. MSE performance of the DOA estimation of the proposed DL based scheme when the learning rate are set as 0.001, 0.01, 0.02, and 0.1.

in which $L = 8$ bits, $L = 16$ bits, $L = 32$ bits, and $L = 64$ bits are considered. It can be seen from Fig. 4 that the MSE of the DOA estimation is reducing with the increasing SNR, and it becomes stable gradually until the SNR is large enough. Also, we can see from the simulation results that the MSE performance can be enhanced when adopting longer training sequences. This result is dedicated by the fact that longer length of training sequence can stir the effectiveness of the offline training process of the DNN, which can capture more complete information about the wireless channel. Additionally, it can further be observed that smaller SNR is required in the $L = 64$ bits case to urge the MSE performance to become stable, indicating that the DOA estimation can be improved when introducing longer training sequences.

The performance comparison in terms of the MSE of the DOA estimation against SNR is shown in Fig. 5, where the learning rate are set as 0.001, 0.01, 0.02, and 0.1, respectively. Here, the length of the training sequence is initialized as 32 bits. It can be seen from Fig. 5 that the MSE performance in the case of "learning rate = 0.1" outperforms that of other cases in terms of MSE, which implies that introducing larger learning rate can elevate the performance of the DOA estimation based on the proposed scheme in the massive MIMO system. By contrast, we must emphasize that too large learning rate will lead to a higher validation error, which can be reflected in the exposition in Fig. 5 where larger learning rate makes the curve steeper. This implies that selecting an appropriate learning rate is a significant issue for boosting the performance of the DL based DOA estimation approach.

Fig. 6 provides a MSE comparison of the DOA estimation in massive MIMO system against SNR based on the proposed DL based approach, where "batch size = 100", "batch size = 50", "batch size = 30", and "batch size = 10" are investigated. Initially, the length of the training sequence is set as 32 bits. It can be observed from Fig. 6 that the MSE is reducing as SNR increases, and MSE can be further constrained when setting smaller batch size in general. Meanwhile, we can find that the MSE performance of the DOA estimation is tending to be more stable in the case of larger batch size. This favorable result is attributed to the fact that a smaller batch size leads to slower convergence while deploying too large batch size in the training procedure causes very large epochs. Thus, it is a trade off between the MSE and the stability of the proposed DL based scheme when modifying the batch size for optimizing the network.

Fig. 7 exhibits the MSE performance of the DOA estimation against the SNR of the proposed DL based scheme, SBEM scheme [17], PA channel tracking scheme [21], ADMA user scheduling scheme [20], and compressed sensing scheme [45], respectively. We can observe that the MSE performance can be improved with higher SNR among all the methods. Particularly, the proposed DL based massive MIMO scheme has about one order of magnitude reduction in terms of the MSE performance compared with the PA channel tracking method, which benefits from the DL based technique that can realize end-to-end optimization whereas enormous channel power leakage and

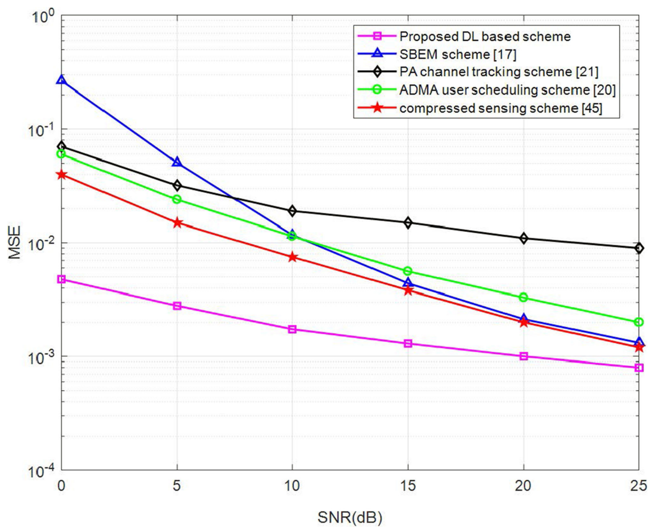

Fig. 7. Comparison of the MSE performance of the DOA estimation of the proposed DL based scheme, SBEM scheme, PA channel tracking scheme, ADMA user scheduling scheme, and compressed sensing scheme.

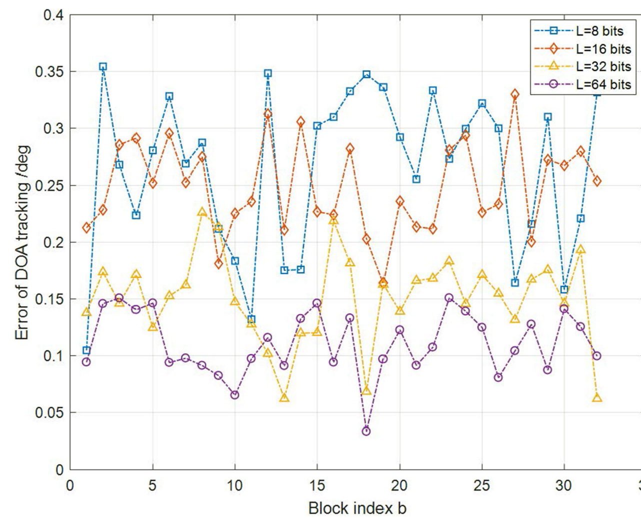

Fig. 8. Error of the DOA tracking against block index b in the cases of $L = 8$ bits, $L = 16$ bits, $L = 32$ bits, and $L = 64$ bits training sequences.

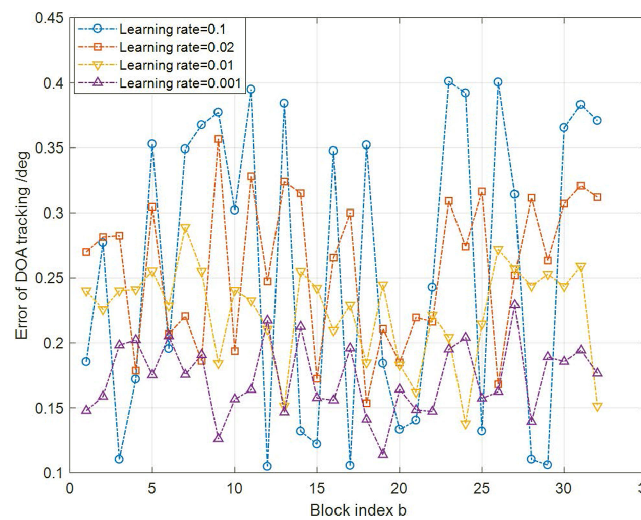

Fig. 9. Error of the DOA tracking against block index b when the learning rate are 0.001, 0.01, 0.02, and 0.1.

tremendous performance loss are induced by the selection of the single spatial support with the maximum amplitude for transmission of the method in [21]. Also, the MSE performance of the DOA estimation of the proposed DL based scheme outperforms that of the SBEM scheme, the ADMA user scheduling approach, and the compressed sensing scheme, for the reason that the DOA can be obtained directly in angle domain and the sparsity features are fully leveraged with the aids of the powerful DL technique and the excellent generalization ability of the learning mechanism of the DNN. Additionally, the significant enhancement of the MSE performance of the DOA estimation is not required for high computational complexity in our proposed DL based scheme.

In this example, we investigate the performance of the DOA tracking method based on the proposed DL based scheme in the massive MIMO system, where the $L = 8$ bits, $L = 16$ bits, $L = 32$ bits, and $L = 64$ bits training sequences are simulated. The block index is changed from 1 to 32 and the SNR is set as 10 dB. As shown in Fig. 8, the DOA tracking performance is improved with the increasing length of the training sequences, for the reason that longer training sequence is more capable of learning the structure of the DOA and the entire training power is proportional to L. Meanwhile, we can find that the vibration of the error of the DOA tracking is curtailed when applying longer training sequences, indicating that the stability of the DOA tracking of the proposed scheme can be optimized by applying longer training sequences. In addition, the DOA tracking performance in the proposed DL based scheme outperforms the conventional ADMA user scheduling scheme [20] in terms of error rate and stability.

The performance comparison in terms of error of the DOA tracking against the block index b is provided in Fig. 9, where SNR is set as 10 dB. Here, the DOA tracking performance is studied with different learning rate. We imply from Fig. 9 that the DOA tracking performance can be improved by using larger learning rate at the cost of increasing jitter error. Also, the vibration of the error of the DOA tracking with "learning rate = 0.001" is smaller than that of other cases with different learning rate. That means smaller learning rate can boost the learning and recognition abilities of the DL based schemes for obtaining smooth results in the massive MIMO system, which contributes to detect the spatial structure of the channel matrix in angle domain. Consider the simulation results presented above, we conclude that the proposed DL based scheme is capable of optimizing the DOA estimation performance in terms of accuracy and robustness.

In this comparison, the MSE performance against training epochs is investigated, where the SNR is set as 10 dB and the length of the training sequence is set as $L = 16$ bits. As shown in Fig. 10, it is observed that the MSE performance is optimized as the training epochs increases, illustrating that the proposed algorithm is approaching to the best solution. Besides, we can further see from Fig. 10 that the MSE curve becomes smooth and the MSE performance is stable when the training epochs is about 20, which verifies the convergence of the proposed DOA estimation method based on the DL based framework.

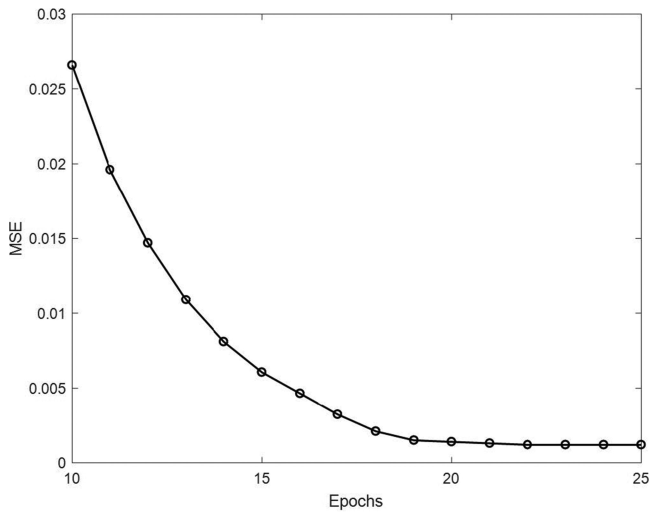

Fig. 10. MSE performance against training epochs.

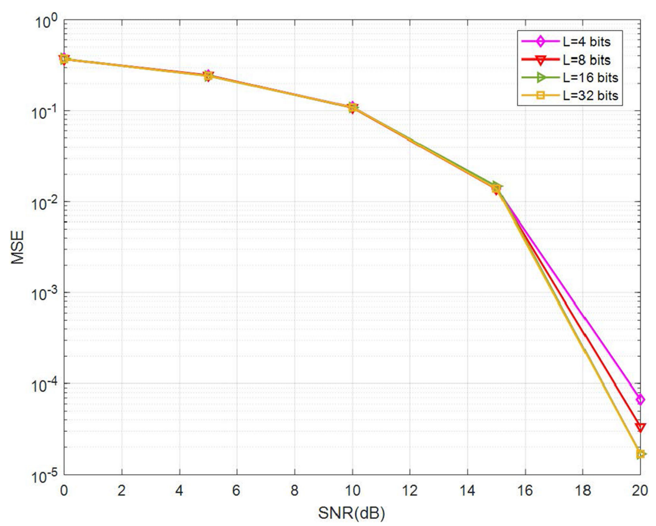

Fig. 11. The MSE performance of the channel estimation in the case of $L = 4$ bits, $L = 8$ bits, $L = 16$ bits, and $L = 32$ bits training sequences.

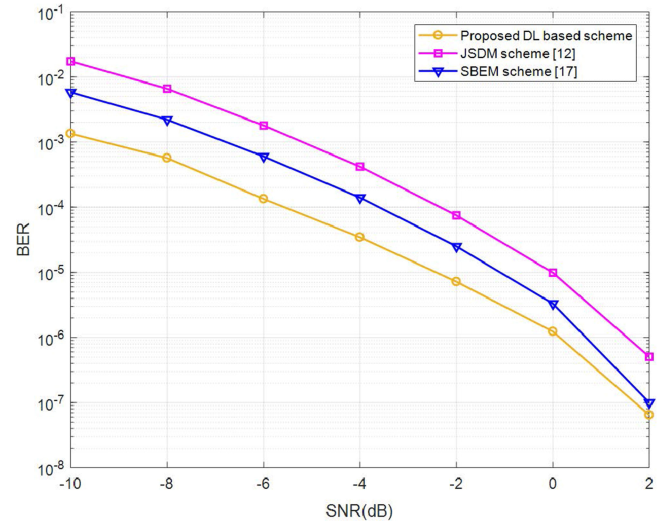

Fig. 12. Comparison of the BER performance of the proposed DL based scheme, the JSDM scheme, and the SBEM scheme.

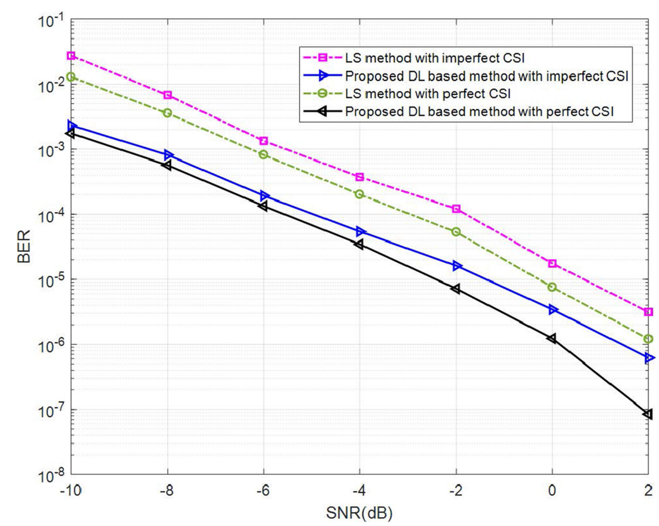

Fig. 13. BER against the SNR of the channel estimation in the proposed DL based scheme and the LS scheme, where imperfect CSI and perfect CSI cases are considered.

B. Channel Estimation

Fig. 11 depicts the MSE performance of the channel estimation based on the proposed DL based scheme in the massive MIMO system, in which the learning rate is set as 0.01. As the SNR increases, we can observe that the MSE performance is improved, which is similar to the results in the DOA estimation presented above. Also, it can be seen from Fig. 11 that the performance of the channel estimation is optimized when employing longer training sequences, which means that the accuracy of the signal detection is enhanced. The reason is that the entire training power is proportional to the length L of the training sequences and the DNN can be trained in a better manner. Furthermore, since the the curve for the $L = 16$ bits case match closely with that of the $L = 32$ bits case, the improvement of the MSE performance of the channel estimation can be negligible when L increases from 16 bits to 32 bits. Hence, we can conclude that the length of the training sequence can be set as 16 bits for avoiding additional overhead in practice.

In the next comparison, we investigate the BER performance against the SNR of the proposed DL based scheme, the JSDM scheme[12], and the SBEM scheme[17], where $L = 12$ bits and the batch size is 50. It is clearly seen from Fig. 12 that the BER performance is improved when the SNR increases among all the methods for channel estimation in the massive MIMO system. In particular, we can find that the BER performance of the proposed DL based scheme outperforms that of the JSDM scheme and the SBEM scheme, which can achieve lower than 10^{-7} BER when the SNR is 2 dB. We deem that the proposed DL based scheme is a better approach for channel estimation since it can achieve lower BER compared with the state-of-the-art methods. Additionally, this result further indicates that DL is a promising tool for ensuring the reliability of the massive MIMO system.

Fig. 13 displays the BER performance against the SNR of the proposed DL based scheme and the traditional least squares

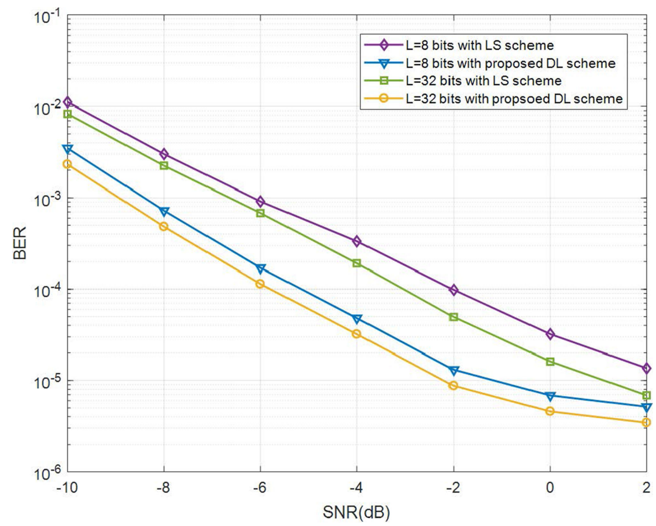

Fig. 14. BER against the SNR of the channel estimation in the proposed scheme and the LS scheme, in which $L = 8$ bits and $L = 32$ bits training sequences are considered.

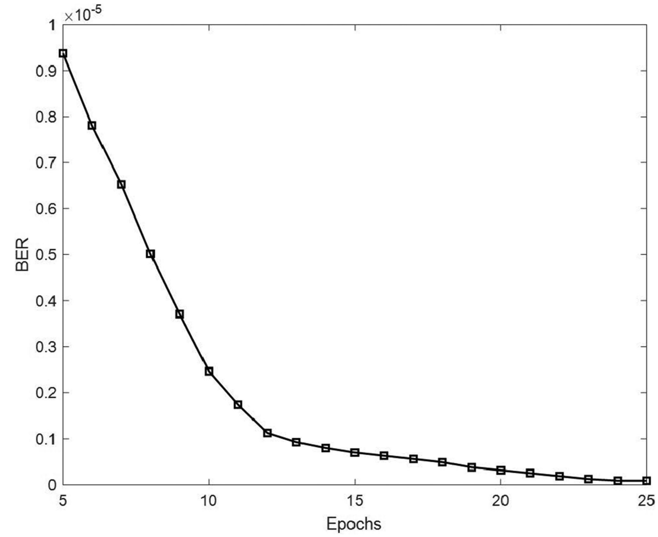

Fig. 15. BER performance against training epochs.

(LS) method, in which imperfect CSI and perfect CSI cases of the methods are investigated. Initially, the length of the training sequence is set as $L = 12$ bits. We can see from Fig. 13 that with better CSI, the BER performance of the proposed DL based scheme and the LS scheme is improved. Synchronously, it is intuitive that the proposed DL based scheme outperforms the LS approach in terms of the BER performance, which still achieves better BER performance with imperfect CSI in comparison with the LS method with perfect CSI and imperfect CSI. The reason lies on the fact that prior statistics of the wireless channel in the massive MIMO system have not been learned in the LS method. In contrast, the proposed DNN has learned the features of the channel matrix by the offline learning and the online learning procedures.

Finally, we consider the BER performance of the channel estimation based on the proposed DL based scheme and the LS scheme in different length of the training sequences, where the learning rate is set as 0.01. As shown in Fig. 14, we can find that the BER performance can be optimized with longer training sequences. Meanwhile, the BER achieves better performance in the proposed DL based scheme, whose performance would be jeopardized in the traditional LS method because uncertain channel has not been detected previously. Particularly, the BER performance of the DL based scheme with $L = 8$ bits training sequence is still better than that of the LS method with $L = 32$ bits training sequence. Together with the results presented in Fig. 13, we can conclude that the proposed DL based massive MIMO framework is robust enough in terms of the length of the training sequences and the various channel conditions. Thus, it comes to the conclusion that the high-resolution channel estimation can be realized with the aids of the well-designed DNN in the massive MIMO system.

In addition, we evaluate the convergence of the proposed Algorithm 1 and Algorithm 2 in Section IV, in which the SNR is set as -5 dB and the length of the training sequence is $L = 12$ bits. As shown in Fig. 15, the BER performance is becoming stable after 24 times of training epochs approximately, which demonstrates the convergence of the proposed Algorithm 1 and Algorithm 2 in the massive MIMO system.

Finally, so as to realize better performance of the DOA estimation and channel estimation based on the proposed schemes, we briefly give some suggestions. According to the simulation results of the proposed methods for DOA estimation and channel estimation, the length of the training sequence can be selected as 12 bits and the learning rate is recommended as 0.01. Also, the batch size is chosen as 50. This deployment strategy is to restrain the overhead in the network training stage without degrading the super performance of the proposed schemes.

VI. Conclusion

In this paper, we have proposed a deep learning based schemes for achieving super-resolution DOA estimation and channel estimation in the massive MIMO system, in which multiple antennas at the BS and single-antenna users are integrated into a DNN. At first, we construct a typical massive MIMO system model and analyze its characteristics. Then, a novel DNN is developed to learn the statistics of the channel model and capture the sparsity features in angle domain, in which multiple hidden layers for strengthening the ability of representation and recognition are included in the network. In order to realize super-resolution DOA estimation and channel estimation, two schemes based on the developed DNN are proposed. To suppress the error of the DOA estimation induced by the spacial distribution of the channel matrix, we derive an algorithm to achieve DOA estimation through the DOA directly with the aids of the powerful deep learning. Afterwards, using the the results obtained in the DOA estimation method, the channel estimation strategy is formulated after running the proposed complex gain estimation method. Simulation results have demonstrated that the proposed scheme can achieve better performance in terms of DOA estimation, which performs better in the MSE performance and the DOA tracking compared with previous methods. Also, it is verified that the proposed scheme can realize high-resolution channel estimation. Our research exploits a new way for wireless communication, which has proved the effectiveness of the framework that employing the deep learning in the massive MIMO system.

References

[1] J. G. Andrews, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] H. Xie, B. Wang, F. Gao, and S. Jin, "A full-space spectrum-sharing strategy for massive MIMO cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2537–2549, Oct. 2016.

[5] Y. Peng, F. Al-Hazemi, R. Boutaba, F. Tong, I. S. Hwang, and C. H. Youn, "Enhancing energy efficiency via cooperative MIMO in wireless sensor networks: state of the art and future research directions," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 47–53, Nov. 2017.

[6] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annu. Symp. Pers. Indoor Mobile Radio Commun.*, London, U.K., Sep. 2013, pp. 611–615.

[7] H. Huang, J. Xiong, J. Yang, G. Gui, and H. Sari, "Rate region analysis in a full-duplex-aided cooperative nonorthogonal multiple-access system," *IEEE Access*, vol. 5, pp. 17869–17880, Aug. 2017.

[8] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[9] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[10] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[11] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.

[12] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[13] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[14] A. Shojaeifard, K. K. Wong, M. Di Renzo, G. Zheng, K. A. Hamdi, and J. Tang, "Massive MIMO-enabled full-duplex cellular networks," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4734–4750, Nov. 2017.

[15] J. Zhang, Z. Zheng , J. Xi , Y. Zhang , X. Zhao, and G. Gui, "3D MIMO for 5G NR: Several observations from 32 to massive 256 antennas based on channel measurement," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 62–70, Mar. 2018.

[16] H. H. Yang, G. Geraci, T. Q. S. Quek, and J. G. Andrews, "Cell-edge-aware precoding for downlink massive MIMO cellular networks," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3344–3358, Jul. 2017.

[17] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD Massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[18] 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.

[19] W. Zhang, F. Gao, S. Jin, and H. Lin, "Frequency synchronization for uplink massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 235–249, Jan. 2018.

[20] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[21] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[22] F. Tang, Z. M. Fadlullah, N. Kato, F. Ono, and R. Miura, "AC-POCA: Anticoordination game based partially overlapping channels assignment in combined UAV and D2D-based networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1672–1683, Feb. 2018.

[23] Hinton, Geoffrey E., S. Osindero, and Y. W. The, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, pp. 1527–1554, 2006.

[24] S. Yu, X. Wang, and R. Langar, "Computation offloading for mobile edge computing: A deep learning approach," in *Proc. IEEE 28th Annu. Int. Sympo. Personal, Indoor, Mobile Radio Commun.*, Montreal, QC, 2017, pp. 1–6.

[25] B. Mao *et al.*, "Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Trans. Comput.*, vol. 66, no. 11, pp. 1946–1960, Nov. 2017.

[26] S. Ayoubi *et al.*, "Machine learning for cognitive network management," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 158–165, Jan. 2018.

[27] S. Wang, M. Xia, and Y. C. Wu, "Multicast wirelessly powered network with large number of antennas via first-order method," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3781–3793, Jun. 2018.

[28] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2432–2455, May 2017.

[29] F. Tang *et al.*, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 154–160, Feb. 2018.

[30] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

[31] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An intelligent traffic load prediction based adaptive channel assignment algorithm in SDN-IoT: A deep learning approach," *IEEE Int. Things J.*, to be published, doi: 0.1109/JIOT.2018.2838574.

[32] D. Takaishi, Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Virtual cell-based resource allocation for efficient frequency utilization in unmanned aircraft systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3495–3504, Apr. 2018.

[33] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 322, pp. 533–536, 1986.

[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[38] B. E. Boser, "A training algorithm for optimal margin classifiers," in *Proc. ACM 5th Workshop Comput. Lern. Theory*, Pittsburgh, 1992, pp. 144–152.

[39] M. I. Jordan, *Learning in Graphical Models*. Cambridge, MA, USA: MIT Press, 1999.

[40] M. A. Ranzato, C. Poultney, S. Chopra, and Y. Lecun, "Efficient learning of sparse representations with an energy-based model," in *Proc. Neural Inf. Proc. Syst.*, 2006, pp. 1137–1144.

[41] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, no. 7, pp. 1044–1046, Jul. 1986.

[42] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[43] L. Cheng, Y.-C. Wu, J. Zhang, and L. Liu, "Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5897–5909, Nov. 2015.

[44] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, 1989, vol. 2, no. 5, pp. 359–366, 1989.

[45] Z. Yang, L. Xie, and C. Zhang, "Off-Grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

Authors' photograph and biography not available at the time of publication.