IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 18, 2021, accepted April 4, 2021, date of publication April 7, 2021, date of current version April 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071590

Deep Learning-Based Channel Estimation for Massive-MIMO With Mixed-Resolution ADCs and Low-Resolution Information Utilization

JIN ZICHENG[1], GAO SHEN[1], LIU NAN[1], (Member, IEEE), PAN ZHIWEN[1,2], (Member, IEEE), AND YOU XIAOHU[1,2], (Fellow, IEEE)

[1]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[2]Purple Mountain Laboratories, Nanjing 211100, China

Corresponding author: Liu Nan (nanliu@seu.edu.cn)

This work was supported in part by the National Key Research and Development Project under Grant 2020YFB1806805 and Grant 2019YFE0123600, and in part by the European Union's Horizon 2020 Research and Innovation Programme through the Marie Skłodowska-Curie Grant under (TESTBED2 Project) Agreement 872172.

ABSTRACT In this paper, we propose two deep-learning based uplink channel estimation approaches that can utilize not only high-resolution-ADC-quantized but also low-resolution-ADC-quantized received pilot signals to improve estimation performance for mixed analog-to-digital converters (ADCs) massive multiple-input multiple-output (MIMO) systems. In each approach, low-resolution-ADC-quantized received pilot signals are utilized with one of three different schemes, i.e., High-resolution quantized pilot + All low-resolution quantized pilot(High + All), High-resolution quantized pilot + Argument of low-resolution quantized pilot (High + Arg) or High-resolution quantized pilot + Modulus of low-resolution quantized pilot (High + Mod). All three schemes include the intact quantized pilot signals at high-resolution antennas, but the quantized pilot signals at low-resolution ADCs are exploited differently in each scheme. Modified selective-input prediction deep neural network (Modified SIP-DNN) is developed to predict more realistic channels and test the effectiveness of the utilization scheme. To achieve further performance improvement, a deep neural network (DNN) based two-stage network is proposed where the recovering DNN (RC-DNN) in the first stage forms a coarse estimation for channels at antennas with low-resolution ADCs and the refining DNN (Ref-DNN) in the second stage outputs a refined estimation for channels at all antennas. Simulation results show that our proposed approaches outperform state-of-the-art channel estimation method especially when most antennas are equipped with low-resolution ADCs.

INDEX TERMS Massive MIMO, mixed-ADC, channel estimation, deep learning.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) can bring huge improvements in system throughput, spectral efficiency and overall signal coverage for wireless systems by equipping a large number of antennas at base stations [1]–[3]. However, using high-resolution analog-to-digital converters (ADCs) in MIMO systems leads to high hardware cost and power consumption. Hence, low-resolution ADCs (i.e., one to three bits) are considered as a solution to this problem. However, pure low-resolution ADCs degrade overall performance and cause error floor in linear multi-user detection [4], data rate loss in the high SNR regime [5] and challenging channel estimation [6].

Mixed-analog-to-digital converter (mixed-ADC) architecture where one-bit ADCs only partially replace conventional high-resolution ADCs is a good solution to this cost-performance conundrum [7], [8]. It is shown in [7] that compared with pure low-resolution ADC systems, channel estimation is more tractable in mixed-ADC architecture. A round-robin method is proposed in [8] to improve estimation performance for slowly and moderately varying fading channels where each antenna take turns to use the high-resolution ADC for channel estimation until all antennas are traversed. The linear minimum mean squared error (LMMSE) estimate of uplink channel and its variances of the independent zero mean elements is derived for

The associate editor coordinating the review of this manuscript and approving it for publication was Yunjie Yang.

mixed-ADC massive MIMO in [9]. The GTurbo-MMV algorithm in [10] exploits the joint sparsity of channel response and obtains good results and robustness in channel estimation as well as user detection. An algorithm that can simultaneously achieve good performance in uplink channel estimation and user activity detection in mixed-ADC distributed massive MIMO is proposed in [11].

Though the mixed-ADC architecture can reduce hardware cost and energy consumption, the low-resolution ADC part still limits the performance of the transceiver, including channel estimation performance. Recently, deep learning (DL) is adopted in channel estimation for massive MIMO [12] and obtains substantial success, especially when low-resolution ADC is deployed. For instance, [13] compares the estimation performance of DL approaches to generalized approximate message passing (GAMP) in massive MIMO with non-ideal one-bit ADCs and simulation results show that DL approaches are more robust to ADC impairments than GAMP approaches. References [14] develops a deep-learning based channel estimation framework for one-bit massive MIMO and observes that fewer pilots are required for the same channel estimation performance when more antennas are employed. Balevi and Andrews et al. [15] propose a two-stage estimation scheme for one-bit massive MIMO by exploiting deep neural network as well as convolutional neural network and obtain 5-10dB gain in channel estimation. In addition, [16] proposes a segment-average based one-bit massive MIMO channel estimation scheme that utilizes a deep neural network (DNN) to process the average of multiple pilot signal segments and the proposed scheme outperforms linear channel estimators. References [17] proposes a channel estimation approach for few-bit massive MIMO that utilizes DNN as an autoencoder to optimize the training signal and perform minimum mean squared error (MMSE) channel estimation concurrently and this approach outperforms the Bussgang-based linear MMSE channel estimator. References [18] exploits a generative adversarial network (GAN) to estimate channels from compressed pilot measurements for one-bit massive MIMO and the scheme achieves superior performance over sparse signal recovery methods.

DL is also demonstrated to be effective for channel estimation in mixed-ADC massive MIMO as well. To the best of our knowledge, [19] is the only work that adopts deep learning in channel estimation for mixed-ADC massive MIMO. The selective-input prediction deep neural network (SIP-DNN) proposed in [19] shows superiority in estimation performance over state-of-the-art linear and nonlinear channel estimation methods.

However, due to the severe quantization distortion, existing channel estimation methods for mixed-ADC massive MIMO usually do not utilize the pilot signals quantized by low-resolution ADCs. For instance, SIP-DNN completely discards the pilot signals received by low-resolution antennas (antennas that are connected with low-resolution ADCs), and round-robin method does not use low-resolution ADCs to quantize pilot signals at all. Avoid using pilot signals quantized by low-resolution ADCs may be viable when most antennas are high-resolution antennas (antennas that are connected with high-resolution ADCs). But for case with few high-resolution antennas, discarding these signals may cause degradation in channel estimation performance since these coarsely quantized pilot signals contain information that can assist channel estimation.

Inspired by previous observation, we propose deep-learning based channel estimation approaches that exploits useful information in not only high-resolution but also low-resolution received pilot signals with consideration of their adversarial impact in mixed-ADC massive MIMO. The contributions of this paper are summarized as follows:

1) Two deep-learning based approaches are proposed for channel estimation in mixed-ADC massive MIMO. In the proposed approaches, the pilot signals quantized by low-resolution ADCs are utilized in one of three ways: High + Arg, High + Mod and High + All. These three schemes all include the intact quantized pilot signals at high-resolution antennas, but the quantized pilot signals at low-resolution antennas are exploited in different manners. High + Arg utilizes the argument of quantized pilot signal at low-resolution antennas while High + Mod utilizes the modulus of low-resolution quantized pilot signals. Unlike High + Arg and High + Mod, High + All directly utilizes the quantized pilot signal at low-resolution antennas without any further processing.
2) Modified SIP-DNN approach is proposed based on SIP-DNN in [19] with modifications on input layer to fit the output of our low-resolution pilot utilization scheme. To achieve further performance improvement and alleviate the adversarial impact of low-resolution ADCs, we propose a two-stage channel estimation approach where RC-DNN in the first stage forms a coarse estimation for channels at low-resolution antennas and Ref-DNN in the second stage outputs a refined estimation for channels at all antennas.
3) Simulation results show that compared to SIP-DNN approach, which is state-of-the-art for channel estimation in mixed-ADC massive MIMO system, the proposed approaches achieve significant reduction in normalized mean squared error (NMSE) when the ratio of high-resolution antennas is low. Compared to SIP-DNN approach, the proposed approaches can also reduce the number of high-resolution antennas while maintaining approximately the same estimation performance.

Notations: In this article, upper and lower case boldface letter are used to denote matrices and vectors respectively. $\mathbf{I}_n$ is the identity matrix of size n. $\|\|, (\cdot)^T, \mathbb{E}\{\cdot\}$ denote the Euclidean norm, transpose and expectation. $|\chi|$ is the cardinality of set χ. $\mathcal{CN}\left(\mu, \sigma^2\right)$ is circular symmetric complex Gaussian distribution with mean μ and variance σ^2. $|a|$ and $\angle a$ denote the modulus and argument of the complex

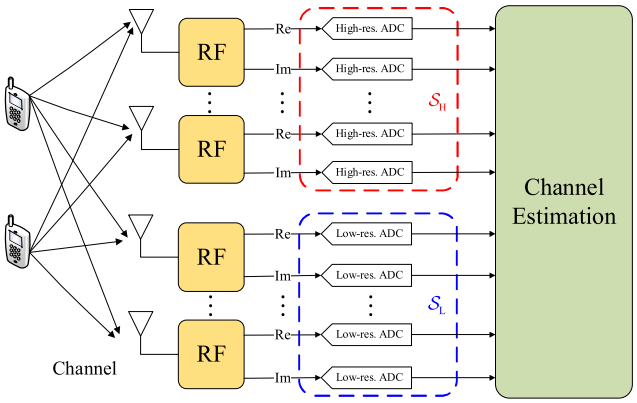


FIGURE 1. System model of a massive MIMO uplink with mixed-ADC.

number a. Re (a) is the real part of the complex number a and Im (a) is the imaginary part. $[\mathbf{w}]_i$ denotes the ith element of the vector $\mathbf{w}$.

II. SYSTEM MODEL

Consider a MIMO uplink system as shown in Fig. 1. M antennas at the BS are divided into two groups, where one group is equipped with high-resolution ADCs while the other group is connected with low-resolution ADCs. $\mathcal{S}_{\mathrm{H}}$ and $\mathcal{S}_{\mathrm{L}}$ represent the index sets of the antennas connected with high-res and low-res ADCs respectively and $\mathcal{S}_{\mathrm{H}} \cap \mathcal{S}_{\mathrm{L}} = \emptyset$, $\mathcal{S}_{\mathrm{H}} \cup \mathcal{S}_{\mathrm{L}} = \{1, 2, \cdots, M\}$. Denote

$$\eta = \frac{|\mathcal{S}_{\mathrm{H}}|}{M}$$

the ratio of antennas connected with high-resolution ADCs over the total number of antennas. One single antenna user is considered for simplicity but the proposed schemes can be easily generalized to MIMO scenario since the channels of different users can be isolated from the received signal. A pilot signal $\sqrt{P}x$ with P representing the transmit power is transmitted by the user to the BS equipped with a uniform linear array (ULA). Then the signal vector received at base station is

$$\mathbf{y} = (\sqrt{P}x) \cdot \mathbf{h} + \mathbf{z}, \tag{1}$$

where $\mathbf{z} \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I}_M)$ is the additive white Gaussian noise with zero mean and covariance $\sigma_0^2 \mathbf{I}_M$ and $\mathbf{h}$ is the multipath channel, and

$$\mathbf{h} = \sum_{l=1}^{L} \beta_l \mathbf{a}(\varphi_l), \tag{2}$$

where L is the number of paths. For the lth path, $\beta_l \sim \mathcal{CN}\left(0, \sigma_\beta^2\right)$ is its propagation gain with σ_β^2 as the average power gain, φ_l is its azimuth angle of arrival (AoA), and $\mathbf{a}(\varphi_l)$ is its response vector which will vary with different types of antenna array manifold. In our case of ULA, it can be expressed as

$$\mathbf{a}(\varphi_l) = \frac{1}{\sqrt{L}} \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\varphi_l)}, \cdots, e^{-j2(M-1)\pi \frac{d}{\lambda} \sin(\varphi_l)}\right]^T, \tag{3}$$

where d is the distance between two adjacent antennas in the ULA and λ is the corresponding wavelength of carrier.

The received pilot signal are then quantized by the ADCs in set $\mathcal{S}_{\mathrm{H}}$ and $\mathcal{S}_{\mathrm{L}}$ respectively. Since the ADCs connected with antennas in $\mathcal{S}_{\mathrm{H}}$ are high-resolution ones, the quantization distortion upon the signal would be diminutive and thus can be neglected. However, for signals received by antennas in $\mathcal{S}_{\mathrm{L}}$, the quantization distortion must be taken into consideration. Thus, let $\mathbf{r}$ denote the quantized received pilot signal, the signal output of ADCs connected to the mth antenna $[\mathbf{r}]_m$ could be expressed as

$$[\mathbf{r}]_m = \begin{cases} [\mathbf{y}]_m, & m \in \mathcal{S}_{\mathrm{H}} \\ \mathcal{Q}_{\mathrm{L}}(\mathrm{Re}([\mathbf{y}]_m)) + j\mathcal{Q}_{\mathrm{L}}(\mathrm{Im}([\mathbf{y}]_m)), & m \in \mathcal{S}_{\mathrm{L}} \end{cases} \tag{4}$$

where $\mathcal{Q}_{\mathrm{L}}(\cdot)$ denotes the element-wise quantization function for low-resolution ADC and $[\mathbf{r}]_m$ is the mth element of $\mathbf{r}$. We ignore the quantization error of the high-resolution ADC in (4) for antennas in set $\mathcal{S}_{\mathrm{H}}$.

In this paper, the low-resolution ADCs are considered to be uniform midrise ones with a fixed quantization step Δ. For a one-bit quantizer, we use sign function to denote its quantization [20]. We do not write this function explicitly for its simplicity. For a D-bit uniform midrise quantizer ($D \geq 2, D \in \mathbb{N}$) with quantization step Δ, its quantization function $\mathcal{Q}_{\text{D-bit}}(\cdot)$ can be described as

$$\mathcal{Q}_{\text{D-bit}}(V_{in}) = \begin{cases} (m - \frac{1}{2})\Delta & \text{if } (m-1)\Delta < V_{in} \leq m\Delta \quad \text{and} \\ & m \in \{-\frac{2^D}{2} + 2, \cdots, \frac{2^D}{2} - 1\} \\ \frac{2^D - 1}{2}\Delta & \text{if } V_{in} > (\frac{2^D}{2} - 1)\Delta \\ -\frac{2^D - 1}{2}\Delta & \text{if } V_{in} \leq (-\frac{2^D}{2} + 1)\Delta \end{cases} \tag{5}$$

where V_{in} is the input of the quantizer. Fig. 2 illustrates the quantization function of a 3-bit uniform midrise quantizer.

III. DEEP LEARNING-BASED CHANNEL ESTIMATION FOR MIXED-ADC MASSIVE MIMO

In the following section, two deep-learning based uplink channel estimation approaches, i.e., the modified SIP-DNN approach and the two-stage approach, are proposed for mixed-ADC massive MIMO. In both approaches, one of three schemes is employed to utilize the low-resolution-ADC-quantized received pilot signals, i.e., High-resolution quantized pilot + All low-resolution quantized pilot (**High + All**), High-resolution quantized pilot + Argument of low-resolution quantized pilot (**High + Arg**) or High-resolution quantized pilot + Modulus of low-resolution quantized pilot (**High + Mod**). Modified SIP-DNN approach is derived from

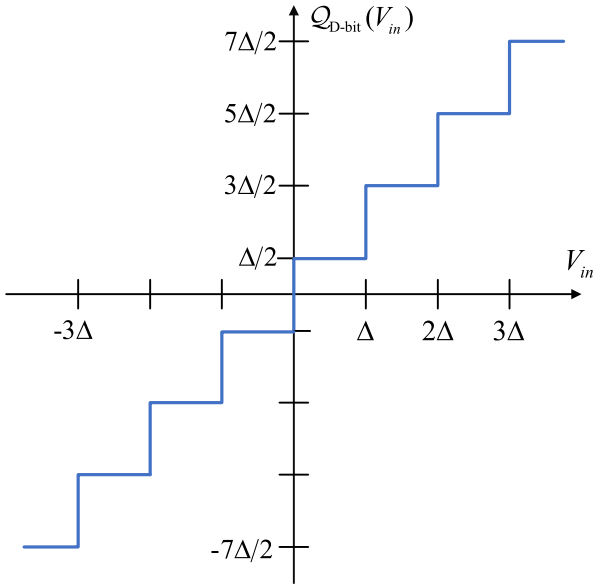


FIGURE 2. A 3-bit uniform midrise quantizer.

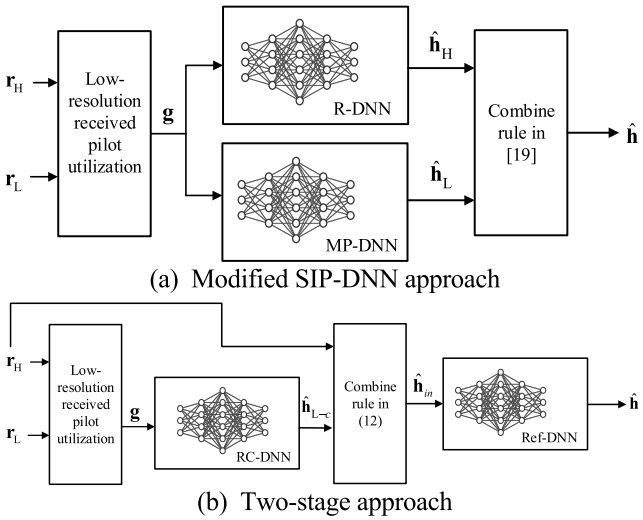


FIGURE 3. Deep learning-based channel estimation.

the SIP-DNN approach in [19] with modifications on the input vector and size of input layers. We also propose a two-stage approach where RC-DNN at the first stage uses the output of low-resolution ADCs to achieve a coarse estimation of channels at low-resolution antennas while Ref-DNN uses the output of RC-DNN and $\mathbf{r}_H$ to yield a more accurate estimation of channel at all antennas. Fig. 3 shows the algorithm detail of both approaches.

A. LOW-RESOLUTION RECEIVED PILOT UTILIZATION SCHEMES

Let $\mathbf{r}_H$ and $\mathbf{r}_L$ denote the sub-vector of the quantized received signal vector $\mathbf{r}$ for antennas in $\mathcal{S}_H$ and $\mathcal{S}_L$ respectively. To exploit low-resolution information, $\mathbf{r}_L$ must be utilized in channel estimation. However, it is also essential to include $\mathbf{r}_H$ in channel estimation since $\mathbf{r}_H$ mainly determines the accuracy of the estimation [21]. Therefore, denote $\mathbf{g}$ the output vector of our utilization scheme, we devise three schemes to utilize the pilot signals at low resolution antennas in both approaches as below.

1) HIGH + ALL

In this scheme, all the quantized pilot signals, no matter quantized by high-resolution or low-resolution ADCs, are utilized directly without further processing. Hence,

$$[\mathbf{g}_{\text{All}}]_m = [\mathbf{r}]_m, \quad \forall m \in \mathcal{S}_H, \mathcal{S}_L \tag{6}$$

2) HIGH + ARG

For a low-resolution antenna, this scheme utilizes the argument of its quantized pilot signal. The quantized pilot signals at high-resolution antennas are utilized directly without further processing. Thus the output of this scheme corresponding to the mth antenna is

$$[\mathbf{g}_{\text{Arg}}]_m = \begin{cases} [\mathbf{r}]_m, & m \in \mathcal{S}_H \\ \angle\,[\mathbf{r}]_m\,, & m \in \mathcal{S}_L \end{cases} \tag{7}$$

3) HIGH + MOD

For a low-resolution antenna, this scheme utilizes the modulus of its quantized pilot signal. The quantized pilot signals at high-resolution antennas are utilized directly without further processing. Thus the output of this scheme corresponding to the mth antenna is

$$[\mathbf{g}_{\text{Mod}}]_m = \begin{cases} [\mathbf{r}]_m, & m \in \mathcal{S}_H \\ |[\mathbf{r}]_m|\,, & m \in \mathcal{S}_L \end{cases} \tag{8}$$

Fig. 4 shows the proposed low-resolution received pilot utilization scheme. There are two reasons for selecting argument/modulus as the way to exploit low-bit information. The first reason is that calculating argument and modulus are two of the most common methods to process complex numbers and since the real and imaginary part of $[\mathbf{r}]_m$ is available for $\forall m \in \{1, 2, \cdots, M\}$, it does not require too much resources to determine $\angle\,[\mathbf{r}]_m$ and $|[\mathbf{r}]_m|$ for $m \in \mathcal{S}_L$. Another reason is that involving argument/modulus of ach element in $\mathbf{r}_L$ in estimation may open up new angles for us to further investigate the channel estimation problem in mixed-ADC massive MIMO system.

After processing, the output of the utilization scheme $\mathbf{g}$ (which is either $\mathbf{g}_{\text{All}}$, $\mathbf{g}_{\text{Arg}}$ or $\mathbf{g}_{\text{Mod}}$ depending on scheme selection) is exploited as inputs for the deep-learning estimation network in modified SIP-DNN approach in Section B as well as the RC-DNN in the two-stage approach in Section C.

B. MODIFIED SIP-DNN APPROACH

SIP-DNN is a deep-learning-based channel estimation approach proposed in [19]. It achieves significant improvement in estimation accuracy by utilizing two DNNs in parallel. The estimation approach in [19] only uses the quantized

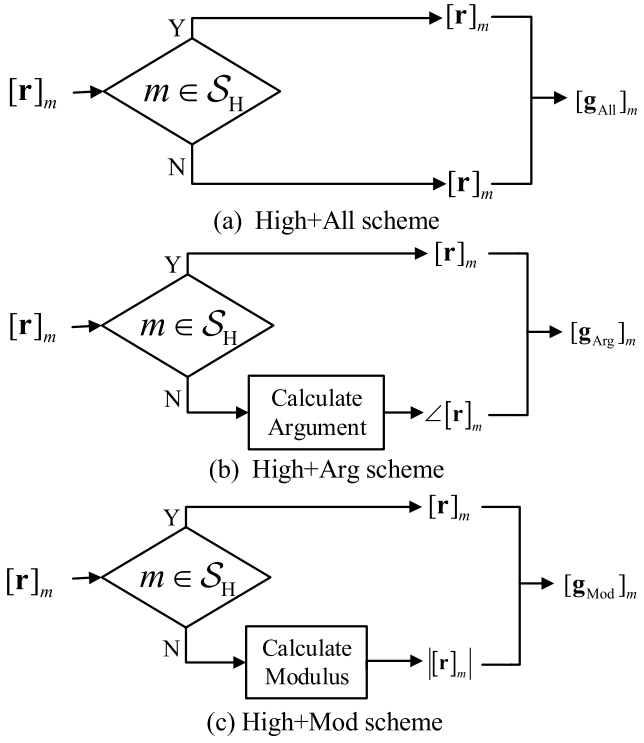


FIGURE 4. Low-resolution received pilot utilization schemes.

received signals at high-resolution antennas as input for the SIP-DNN network. We modify this approach by using the output of our low-resolution utilization scheme $\mathbf{g}$ as input for deep-learning estimation network and change the input layer size for both DNNs.

We propose this approach mainly to test the effectiveness of the proposed utilization schemes. If the modified approach outperforms the original SIP-DNN approach under same test conditions in simulation, the effectiveness of the proposed utilization schemes can be proved credibly since the only difference between two approaches is the input. In addition, by devising the modified approach, we could demonstrate the applicability and practicality of the proposed utilization schemes since deploying the schemes only requires simply changing the input and adjusting the size of the input layer.

Fig. 3(a) shows the modified SIP-DNN approach. Denote $\hat{\mathbf{h}}_{\mathrm{H}}$ and $\hat{\mathbf{h}}_{\mathrm{L}}$ the estimated channel corresponding to high-resolution antennas and low-resolution antennas respectively. $\hat{\mathbf{h}}_{\mathrm{H}}$ and $\hat{\mathbf{h}}_{\mathrm{L}}$ can be expressed as below

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{H}} &= g_{\mathrm{R},S}(\mathbf{U}_S g_{\mathrm{R},S-1}(\cdots g_{\mathrm{R},2}(\mathbf{U}_2\mathbf{g})))\\ \hat{\mathbf{h}}_{\mathrm{L}} &= g_{\mathrm{MP},T}(\mathbf{V}_T g_{\mathrm{MP},T-1}(\cdots g_{\mathrm{MP},2}(\mathbf{V}_2\mathbf{g})))\end{aligned} \tag{9}$$

where S and T represent the total numbers of neural layers, $g_{\mathrm{R},s}$ and $g_{\mathrm{MP},t}$ are the activation function for sth layer and tth layer, respectively. $\mathbf{U}_s$ and $\mathbf{V}_t$ denote the corresponding weight matrices with respect to R-DNN and MP-DNN for $\forall s = 2,\ldots,S$ and $\forall t = 2,\ldots,T$. Table 2 shows the architecture of the modified SIP-DNN under different η with $M = 64$.

C. TWO-STAGE APPROACH

Inspired by the fact that neural networks have been widely used to recover information in image processing and the excellent performance of DNN in recovering channel information at antennas with low-bit ADCs, we believe that improvements in performance can be obtained by recovering the channel information at the antennas with low-resolution ADCs with decent accuracy first and then further processing the combined signal of recovered channels at low-resolution antennas and information at high-resolution antennas.

Denote $\mathbf{h}_{\mathrm{L}}$ the real channels from user equipment to the BS antennas in $\mathcal{S}_{\mathrm{L}}$. The basic idea of the proposed two-stage approach can be summarized as follows:

1) RC-DNN at the first stage utilizes the output of Low-resolution received pilot utilization schemes $\mathbf{g}$ to obtain a coarse estimation of $\mathbf{h}_{\mathrm{L}}$, i.e. $\hat{\mathbf{h}}_{\mathrm{L}-c}$.
2) $\hat{\mathbf{h}}_{\mathrm{L}-c}$ is combined with $\mathbf{r}_{\mathrm{H}}$ to form the input vector $\hat{\mathbf{h}}_{in}$ for Ref-DNN, which is the DNN at second stage.
3) Ref-DNN uses $\hat{\mathbf{h}}_{in}$ to produce a refined estimation of $\mathbf{h}$, i.e. $\hat{\mathbf{h}}$. Fig. 3(b) shows the detailed structure of our two-stage approach.

1) RC-DNN

For RC-DNN, its output $\hat{\mathbf{h}}_{\mathrm{L}-c}$ can be expressed as

$$\hat{\mathbf{h}}_{\mathrm{L}-c} = c \cdot g_{\mathrm{RC},T}(\mathbf{W}_T g_{\mathrm{RC},T-1}(\cdots g_{\mathrm{RC},2}(\mathbf{W}_2\mathbf{g} + \mathbf{b}_{\mathrm{RC},2}) + \cdots) + \mathbf{b}_{\mathrm{RC},T}) \tag{10}$$

where $T \geq 2, T \in \mathbb{N}$ is the total number of layers in RC-DNN. For $t \in \{2,\cdots,T\}$, $g_{\mathrm{RC},t}(\cdot)$ denotes the activation function for RC-DNN. $\mathbf{b}_{\mathrm{RC},t}$ and $\mathbf{W}_t$ denotes the bias vector and the weight matrix of tth layer respectively. c is a scaling factor to make sure all of the target data fall into the range of the activation function at the output layer.

The weight and bias matrices of RC-DNN are updated through offline training, then saved at base station for uplink channel estimation. The training objective of RC-DNN is to minimize the MSE between the coarse estimation and the actual channel over all training samples, which is

$$\mathrm{MSE}_{RC} = \frac{1}{N_{\mathrm{tr}}}\sum_{n=1}^{N_{\mathrm{tr}}}\left\|\mathbf{h}_{\mathrm{L},n} - \hat{\mathbf{h}}_{\mathrm{L}-c,n}\right\|^2 \tag{11}$$

where N_{tr} is the number of training samples. $\mathbf{h}_{\mathrm{L},n}$ and $\hat{\mathbf{h}}_{\mathrm{L}-c,n}$ denote the true channels at low-resolution antennas and their coarse estimate respectively for the nth training sample. Back propagation algorithm is exploited to minimize the training objective.

For RC-DNN, the nth offline training sample has the form of

$$\left(\mathbf{c}_n, \frac{\mathbf{h}_{\mathrm{L},n}}{c}\right)$$

where $\mathbf{c}_n$ is the input for RC-DNN for the nth sample and

$$\frac{\mathbf{h}_{\mathrm{L},n}}{c}$$

is the target data that RC-DNN is trying to approximate in offline training.

2) FORMULATING THE INPUT VECTOR FOR REF-DNN

As mentioned before, $\mathbf{r}_{\mathrm{H}}$ and $\hat{\mathbf{h}}_{\mathrm{L}-c}$ need to be combined into one vector before being processed by Ref-DNN. The combining principle can be described as

$$\left[\hat{\mathbf{h}}_{in}\right]_m = \begin{cases} [\mathbf{r}_{\mathrm{H}}]_{u_m}, & \text{if } m \in \mathcal{S}_{\mathrm{H}} \\ \left[\hat{\mathbf{h}}_{\mathrm{L}-c}\right]_{v_m}, & \text{if } m \in \mathcal{S}_{\mathrm{L}} \end{cases} \tag{12}$$

where u_m and v_m are the indexes in $\mathcal{S}_{\mathrm{H}}$ and $\mathcal{S}_{\mathrm{L}}$ for the mth antenna, respectively.

3) REF-DNN

Although $\hat{\mathbf{h}}_{\mathrm{L}-c}$ is obtained in the first stage, more processing is required to achieve a more accurate estimation of $\mathbf{h}_{\mathrm{L}}$. In addition, an estimation of $\mathbf{h}_{\mathrm{H}}$ also remains to be acquired since $\mathbf{r}_{\mathrm{H}}$ is still noisy and cannot be regarded as accurate estimation.

For Ref-DNN, its output $\hat{\mathbf{h}}$ can be expressed as

$$\hat{\mathbf{h}} = c \cdot g_{\mathrm{Ref},K}(\mathbf{P}_K g_{\mathrm{Ref},K-1}(\cdots g_{\mathrm{Ref},2}(\mathbf{P}_2\hat{\mathbf{h}}_{in} + \mathbf{b}_{\mathrm{Ref},2}) + \cdots) + \mathbf{b}_{\mathrm{Ref},K}) \tag{13}$$

where $K \geq 2, K \in \mathbb{N}$ is the total number of layers in Ref-DNN. For $k \in \{2, \cdots, K\}$, $g_{\mathrm{Ref},k}(\cdot)$ denotes the activation function for Ref-DNN, $\mathbf{b}_{\mathrm{Ref},k}$ and $\mathbf{P}_k$ denotes the bias vector and the weight matrix of kth layer. c is the same scaling factor used in RC-DNN. $\hat{\mathbf{h}}$ is the also final output of this approach, which is a refined estimation of the real channel $\mathbf{h}$.

The weight and bias matrices of Ref-DNN are updated through back propagation in offline training and saved at base station. The training objective of Ref-DNN is to minimize the MSE between $\hat{\mathbf{h}}$ and $\mathbf{h}$ for all training samples, which is

$$\mathrm{MSE}_{\mathrm{Ref}} = \frac{1}{N_{\mathrm{tr}}} \sum_{n=1}^{N_{\mathrm{tr}}} \left\| \mathbf{h}_n - \hat{\mathbf{h}}_n \right\|^2 \tag{14}$$

$\mathbf{h}_n$ and $\hat{\mathbf{h}}_n$ represent the true channels and the output channel estimation for the nth sample. For Ref-DNN, the nth offline training sample has the form of

$$\left(\hat{\mathbf{h}}_{in,n}, \frac{\mathbf{h}_n}{c}\right)$$

where $\hat{\mathbf{h}}_{in,n}$ is the input for Ref-DNN for the nth sample, which is formed by $\hat{\mathbf{h}}_{\mathrm{L}-c,n}$ and $\mathbf{h}_{\mathrm{H},n}$.

$$\frac{\mathbf{h}_n}{c}$$

is the target data that Ref-DNN is trying to approximate in offline training.

TABLE 1. Output vector size for the low-res received pilot utilization schemes.

η	Input layer size	
	Schemes	
	High+Arg/Mod	High+All
0.1	70	128
0.2	77	
0.3	83	
0.4	90	
0.5	96	

TABLE 2. Architecture of modified SIP-DNN.

Neural layer	Size of R-DNN					Activation
	η					
	0.1	0.2	0.3	0.4	0.5	
Input layer	70	77	83	90	96	
Dense layer 1	20	50	80	100	128	ReLU
Dense layer 2	40	100	160	160	200	ReLU
Dense layer 3	20	50	80	100	128	ReLU
Output layer	12	26	38	52	64	tanh

Neural layer	Size of MP-DNN					Activation
	η					
	0.1	0.2	0.3	0.4	0.5	
Input layer	70	77	83	90	96	
Dense layer 1	20	50	80	100	128	ReLU
Dense layer 2	60	100	160	160	200	ReLU
Dense layer 3	220	140	135	150	128	ReLU
Output layer	116	102	90	76	64	tanh

IV. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed estimation approaches. In our simulation, the number of antennas is set as $M = 64$ in all experiments. The number of paths is set as. $L = 8$.. The ADCs are arranged in a block pattern where high-resolution and low-resolution antennas are placed in two blocks without any overlap. For example, when $\eta = 0.5$, $\mathcal{S}_{\mathrm{H}}$ will be $\mathcal{S}_{\mathrm{H}} = \{1, 2, \cdots, 32\}$ while $\mathcal{S}_{\mathrm{L}} = \{33, 34, \cdots, 64\}$. For each path, the average power gain is set as $\sigma_\beta^2 = 1$, $\frac{d}{\lambda}$ is set as 0.5 and the AoA is selected randomly from $\frac{2\pi}{20} \times [0, 1, \cdots, 19]$.

Table 1 shows the output size for the low-resolution received pilot utilization schemes. Table 2 shows the detailed structure of the modified SIP-DNN approach under High + Arg scheme and all the η in the test. Both MP-DNN and R-DNN are densely connected and have one input layer, three hidden layers and one output layer. In addition, in order to facilitate comparison with the original SIP-DNN and test the effectiveness of our scheme, all other aspects of this approach is identical to the original SIP-DNN approach

TABLE 3. Architecture of two-stage approach.

Neural layer	Size of RC-DNN η					Activation
	0.1	0.2	0.3	0.4	0.5	
Input layer	70	77	83	90	96	
Dense layer 1	20	50	80	100	128	ReLU
Dense layer 2	60	100	160	160	200	ReLU
Dense layer 3	220	140	135	150	128	ReLU
Output layer	116	102	90	76	64	tanh

Neural Layer	Size of Ref-DNN	Activation
Input layer	128	
Dense layer 1	192	ReLU
Dense layer 2	192	ReLU
Output layer	128	tanh

including offline training objectives and vector combining principle.

The detailed structure for the two-stage approach is shown in Table 3. The RC-DNN architecture we design for $M = 64$ consists of one input layer, three hidden layers with rectified linear unit (ReLU) activation function and one output layer with hyperbolic tangent (tanh) activation function. The Ref-DNN architecture includes one input layer, two hidden layers with 192 neurons each layer and one output layer with tanh activation function. All neurons in RC-DNN and Ref-DNN are densely connected.

100 sets of AoAs are generated randomly for training and validation and another 10 sets of random AoAs are generated for testing. 1000 samples are generated for each AoA, which means 100000 samples in total for training and validation and 10000 samples for the training set. The 100000 samples are first shuffled to ensure the intrinsic channel structure can be studied without interference of clusters in data and then divided into the training set and validation set by the ratios of 90% and 10%. The number of epochs, learning rate, batch size are as 100, 1×10^{-3} and 128 respectively. Adam optimizer is used for training. The scaling factor, i.e. c, is set as 3. The transmit power for the pilot signal, i.e. P, is set as 1. Normalized MSE (NMSE) is exploited to evaluate the channel estimation performance. NMSE can be described as

$$\text{NMSE} = \eta\mathbb{E}\left\{\frac{\left\|\mathbf{h}_{\text{H}} - \hat{\mathbf{h}}_{\text{H}}\right\|^2}{\|\mathbf{h}_{\text{H}}\|^2}\right\} + (1-\eta)\mathbb{E}\left\{\frac{\left\|\mathbf{h}_{\text{L}} - \hat{\mathbf{h}}_{\text{L}}\right\|^2}{\|\mathbf{h}_{\text{L}}\|^2}\right\}$$

in modified SIP-DNN approach and

$$\text{NMSE} = \mathbb{E}\left\{\frac{\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|^2}{\|\mathbf{h}\|^2}\right\}$$

TABLE 4. Training and testing time for all approaches.

Modified SIP-DNN approach			
	High+Arg	High+Mod	High+All
Total training time (s)	139.063	139.8848	140.6308
Testing time per sample (s)	0.163562	0.16456	0.165546
Two-stage approach			
	High+Arg	High+Mod	High+All
Total training time (s)	260.3546	262.5354	268.5028
Testing time per sample (s)	0.285237	0.281221	0.287227
Original SIP-DNN			
Total training time (s)	137.718		
Testing time per sample (s)	0.161567		

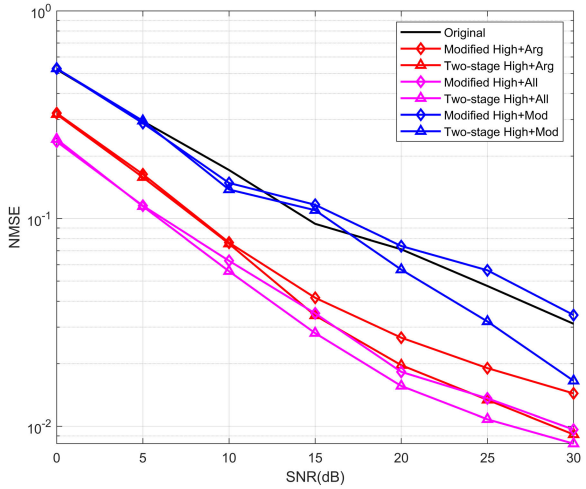
in two-stage approach. SIP-DNN from [19] serves as benchmark for comparison.

A. ANALYSIS WITH $\eta = 0.2$

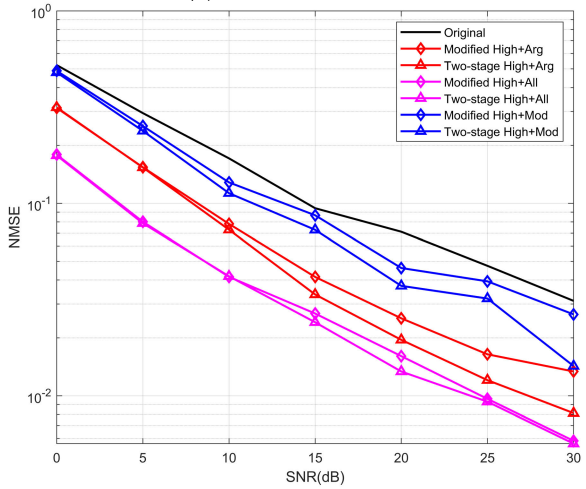
Fig. 5 illustrates the NMSE of the original SIP-DNN approach and the two proposed approaches with different low-resolution pilot utilization schemes for SNRs from 0dB to 30dB. 1∼3bit ADCs are selected as low-resolution ADCs. As we can see in Fig. 5, both proposed approaches, especially the modified SIP-DNN, outperform the original SIP-DNN in nearly all scenarios when $\eta = 0.2$. This shows the effectiveness of the low-resolution received pilot utilization schemes and that our approach could effectively utilize the low-resolution-ADC-quantized pilot signals to improve estimation performance.

From Fig. 5 we can also notice that when both approaches apply the same low-resolution pilot utilization scheme, two-stage approach outperforms modified SIP-DNN approach. This indicates that the two-stage approach can take better advantage of the information in the low-resolution-ADC-quantized received pilot signals. We speculate that this is because two-stage approach allows the received pilot signals at low-resolution antennas to undergo more processing, thereby improving the estimation accuracy for channels at low-resolution antennas. Compared to $\mathbf{r}_{\text{L}}$, the recovered $\hat{\mathbf{h}}_{\text{L}-c}$ at the output of first stage has more accurate modulus information and reduced distortion, thus the DNN in second stage could utilize the additional information to further improve the estimation accuracy.

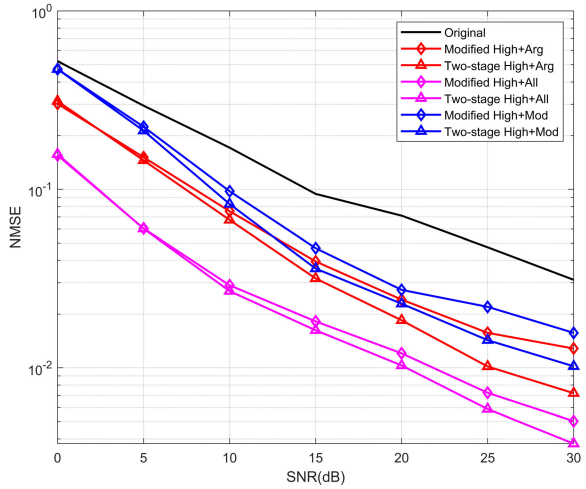
In addition, it is noticeable in Fig. 5 that High + All scheme achieves the best performance when other factors are the same. This is understandable since the calculation of modulus or argument is lossy, therefore High + All scheme retains more information at the network input than the other two schemes. What interests us is that High + Arg outperforms High + Mod. This indicates that the argument of received pilot signals at low-resolution antennas is more useful than their modulus at improving estimation performance. Since the observations in [21] show that the addition of low-bit antenna information, though not helpful for estimating the



(a) With 1-bit ADC

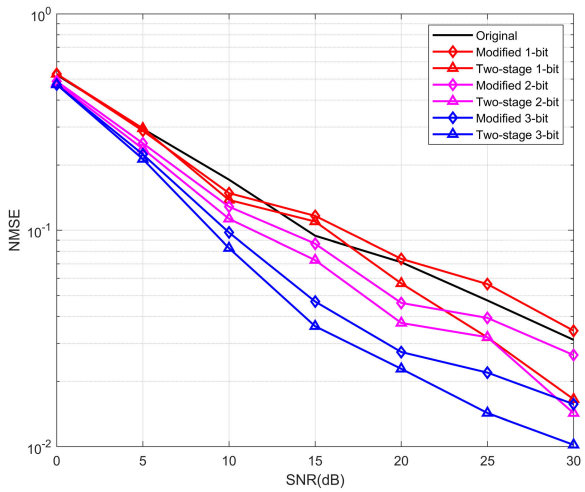


(b) With 2-bit ADC

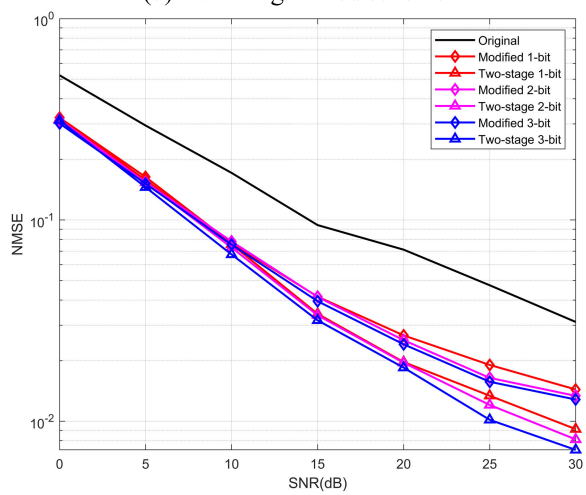


(c) With 3-bit ADC

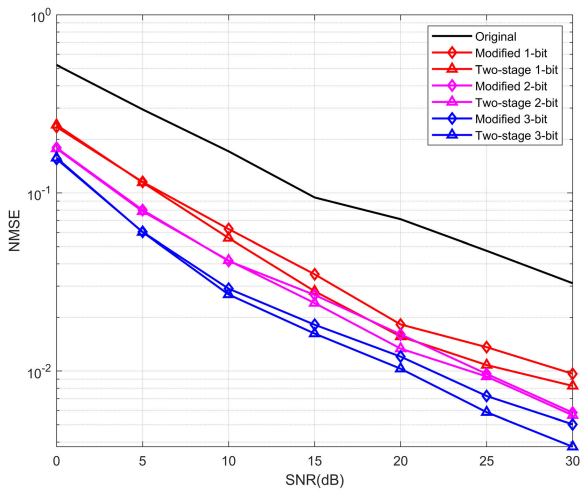
FIGURE 5. NMSE versus SNR with $\eta = 0.2$ and 1∼3bit ADCs used as low-resolution ADCs.



(d) With High+Mod scheme



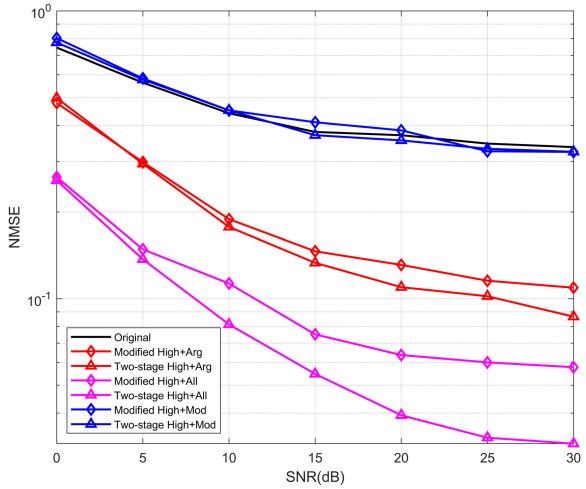
(e) With High+Arg scheme



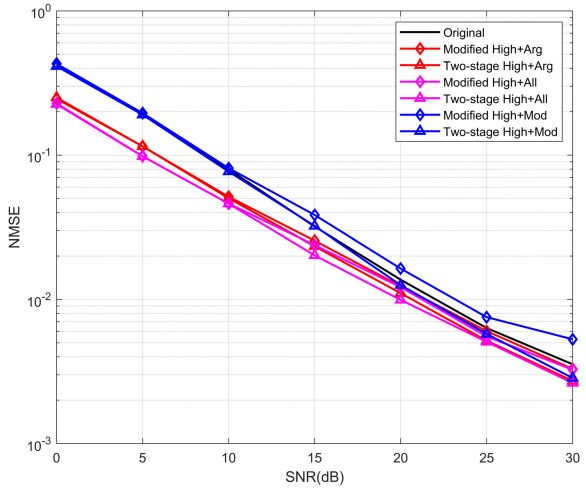
(f) With High+All scheme

FIGURE 5. *(Continued.)* NMSE versus SNR with $\eta = 0.2$ and 1∼3bit ADCs used as low-resolution ADCs.

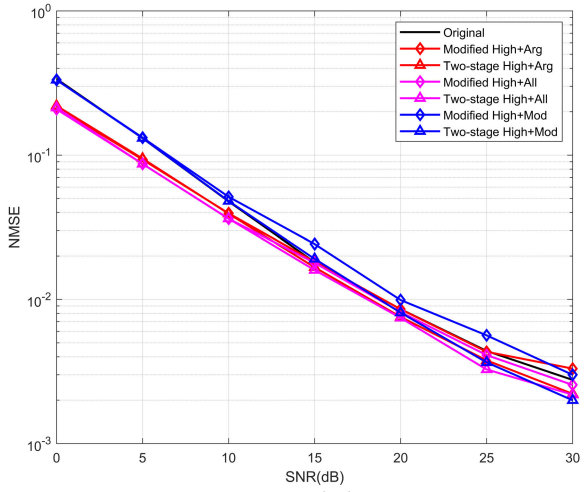
amplitude, can provide useful information for the estimation of AoA, this indicates that argument can provide more AoA related information than modulus and thus can help improve the estimation performance. In addition, by observing Fig. 5(d) to Fig. 5(f), we can notice that though all three schemes will benefit from the bit increase, High + Mod is most-benefited and while High + Arg is the least-influenced scheme regarding the bit increase.



(a) $\eta = 0.1$

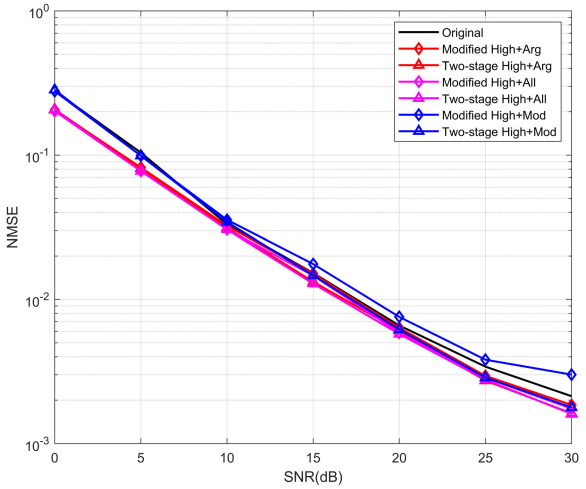


(b) $\eta = 0.3$



(c) $\eta = 0.4$

FIGURE 6. NMSE versus SNR with one-bit ADC as low-resolution ADC and different η.



(d) $\eta = 0.5$

FIGURE 6. *(Continued.)* NMSE versus SNR with one-bit ADC as low-resolution ADC and different η.

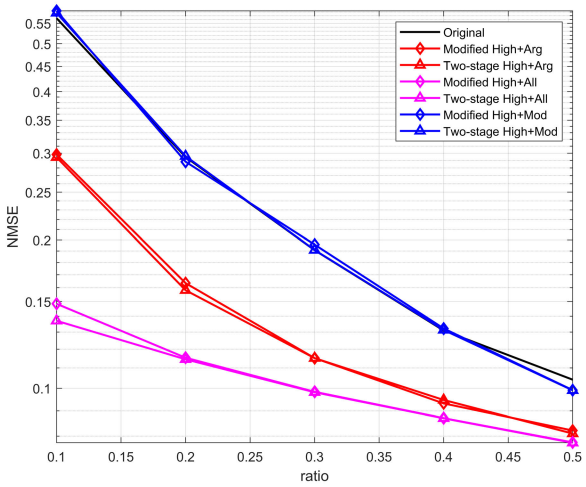
The improvement in performance of our proposed approaches comes with additional complexity. Table 4 shows the amount of time required to complete training and testing for modified SIP-DNN, original SIP-DNN and two-stage approach. As is shown in the table, the two-stage approach need almost twice the time to achieve an estimation compared to the modified and original SIP-DNN approach. This is because in SIP-DNN, the estimation of channels at high-resolution antennas and low-resolution antennas can be performed in parallel while in two-stage approach the received pilot signals must be pipelined through RC-DNN and Ref-DNN sequentially to form a channel estimate. However, both the two-stage approach and modified SIP-DNN approach are sufficient for practical usage with this testing time.

B. ANALYSIS WITH DIFFERENT RATIO

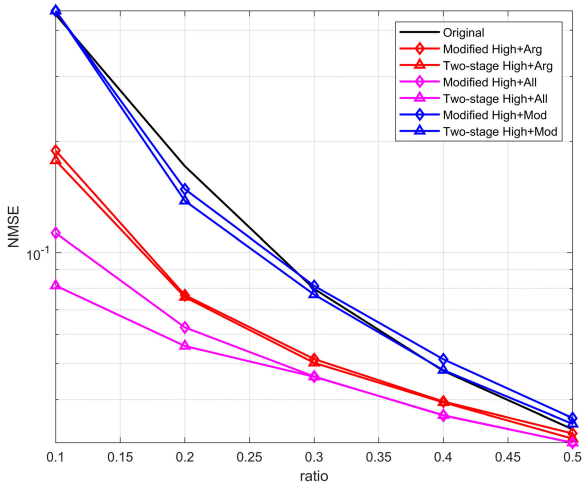
Fig. 6 shows the NMSE versus SNR for the original SIP-DNN, modified SIP-DNN approach and two-stage approach with all three schemes when $\eta = 0.1, 0.3, 0.4, 0.5$ respectively. One-bit ADCs are used as low-resolution ADCs in Fig. 6. Compare Fig. 6 with Fig. 5, we can notice that with the increase of η, the performance gain of our proposed approaches decreases. This is because the gain of our method is due to the addition of information at low-resolution antennas and as the proportion of high-resolution antennas increases, the proportion of low-resolution antennas correspondingly decreases, leading to the reduction in performance gain. This also corroborates our previous statement that High + Arg scheme is a good alternative to High + All scheme since Fig.6 also shows that the performance gap between High + Arg and High + All scheme decreases as the number of high-resolution antennas increases and when $\eta \geq 0.3$, the High + Arg and High + All schemes can achieve similar estimation performance.

C. ANALYSIS UNDER DIFFERENT SNR

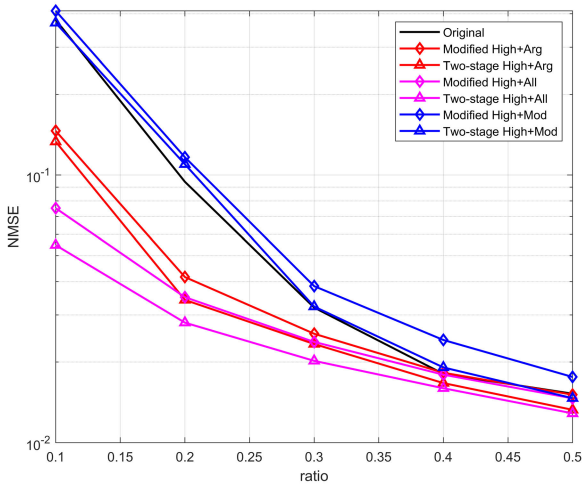
Fig. 7 shows the NMSE performance versus the ratio of high-resolution antennas under SNR from 5 to 30 dB for original SIP-DNN, modified SIP-DNN approach and two-stage



(a) SNR=5dB

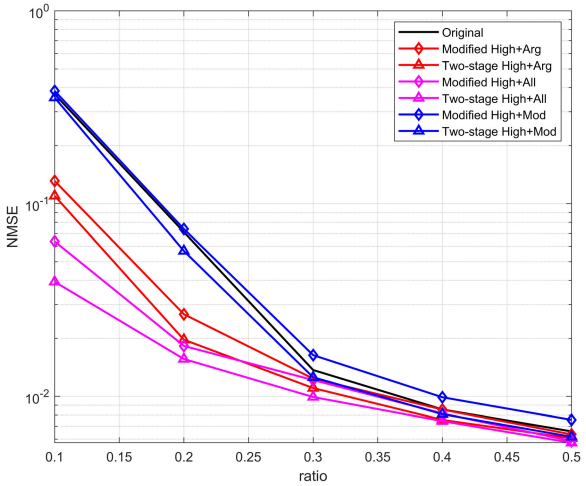


(b) SNR=10dB

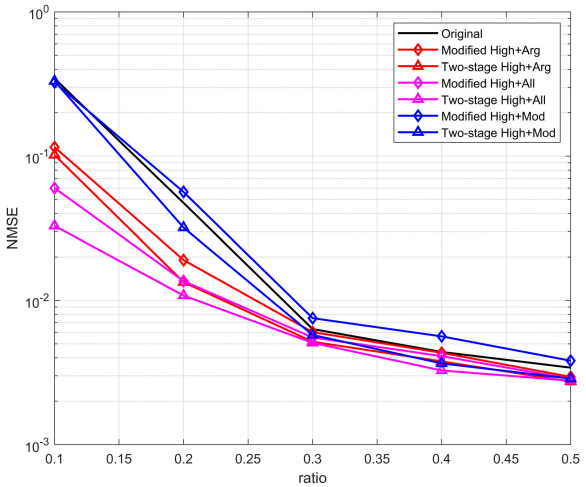


(c) SNR=15dB

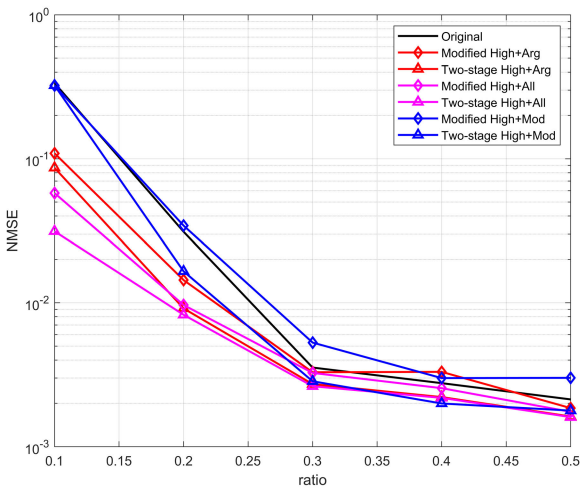
FIGURE 7. NMSE versus η with one-bit ADC as low-resolution ADC and different SNRs.



(d) SNR=20dB



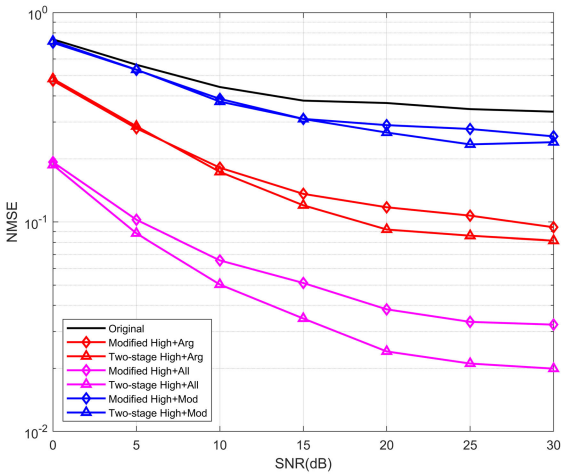
(e) SNR=25dB



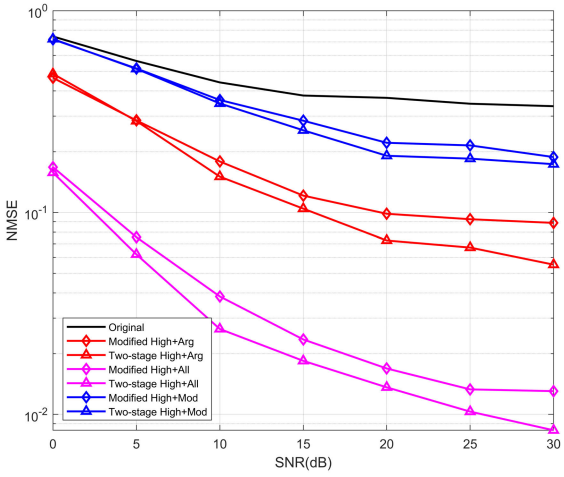
(f) SNR=30dB

FIGURE 7. *(Continued.)* NMSE versus η with one-bit ADC as low-resolution ADC and different SNRs.

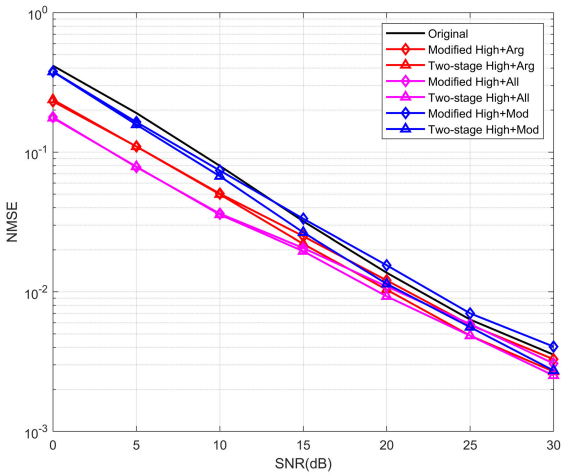
approach with all three schemes. One-bit ADCs are used as low-resolution ADCs in Fig. 7. Fig. 7 shows that compared to the original SIP-DNN approach, the proposed approaches can reduce the number of high-resolution antennas while maintaining approximately the same estimation performance. For instance, under High + All scheme and SNR = 10dB, the estimation performance of modified SIP-DNN approach at $\eta = 0.15$ is approximately the same as the performance



(a) 2-bit ADC used, $\eta = 0.1$

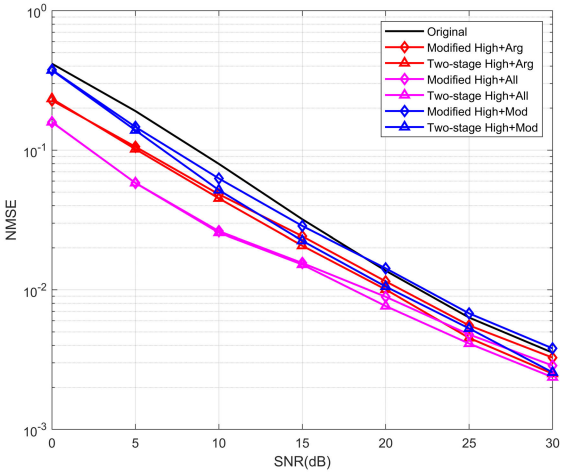


(b) 3-bit ADC used, $\eta = 0.1$

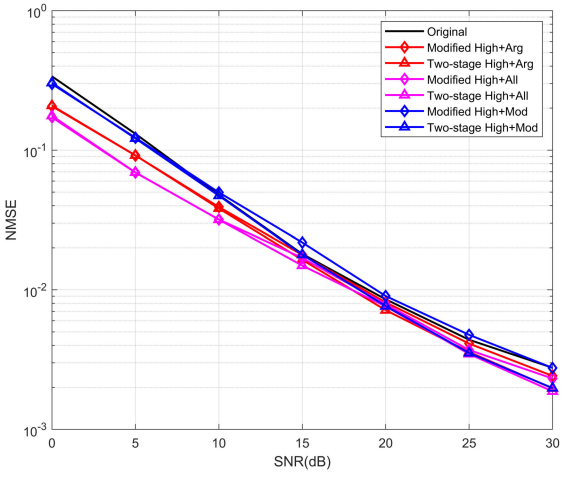


(c) 2-bit ADC used, $\eta = 0.3$

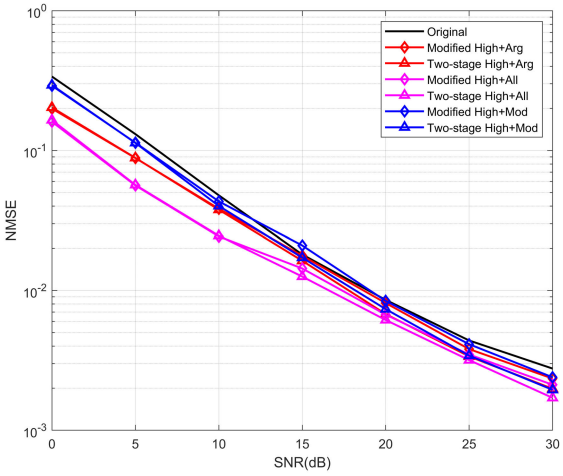
FIGURE 8. NMSE versus SNR with 2-bit/3-bit ADC as low-resolution ADC and different η.



(d) 3-bit ADC used, $\eta = 0.3$



(e) 2-bit ADC used, $\eta = 0.4$

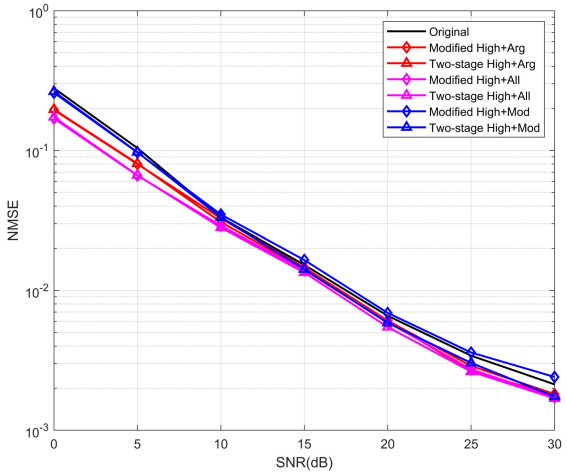


(f) 3-bit ADC used, $\eta = 0.4$

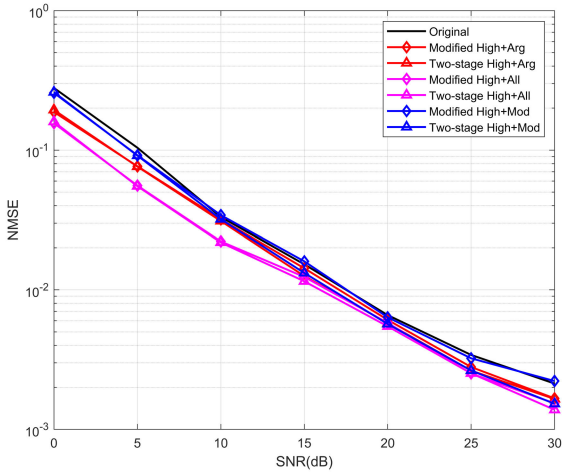
FIGURE 8. *(Continued.)* NMSE versus SNR with 2-bit/3-bit ADC as low-resolution ADC and different η.

of original SIP-DNN approach at $\eta = 0.3$, which means our estimation approach could reduce the ratio of high-resolution antenna by 0.15 without damaging estimation performance. The ratio of high-resolution antennas can be further reduced by implementing two-stage approach. In Fig. 7(b), while both using High + All scheme, the two-stage approach can reduce η by 0.05 on the basis of modified SIP-DNN approach.

Another observation from Fig. 7 is that when the number of high-resolution antennas exceeds a certain proportion, the performance gain brought by adding quantized received pilot



(g) 2-bit ADC used, $\eta = 0.5$



(f) 3-bit ADC used, $\eta = 0.5$

FIGURE 8. ***(Continued.)*** **NMSE versus SNR with 2-bit/3-bit ADC as low-resolution ADC and different η.**

signals at low-resolution antennas becomes much less significant. From Fig. 7 we observe that this certain proportion is related to SNR and decreases as SNR increases. When SNR = 10dB, the performance gain brought by the addition of low-bit information will not disappear until $\eta = 0.5$ but when SNR = 15dB and SNR = 20dB, the performance gain will disappear after $\eta = 0.4$ and $\eta = 0.3$. The performance gain brought by adding received pilot signals at low-resolution antennas is still decreasing for $\eta \leq 0.3$ when SNR = 25dB and SNR = 30dB, indicating this certain proportion will keep decreasing as SNR increases. We hope to investigate the properties of this certain proportion further in future studies.

Though involving the quantized received pilot signals at low-resolution antennas in channel estimation does not significantly improve estimation accuracy when SNR is high or the ratio of high resolution antennas exceeds a certain proportion, it can significantly improve the estimation performance at low SNR and can effectively reduce the ratio of high-resolution antennas needed at base station. And from our observation, if the estimation approach is well-designed, the addition of low-resolution signals will not hinder the overall estimation performance.

D. ANALYSIS WITH 2-BIT ANT 3-BIT ADCS AS LOW-RESOLUTION ADCS

An important issue related to practical applications is whether it is necessary to use 2-bit or 3-bit ADCs instead of 1-bit ADCs as low-resolution ADCs in a mixed-ADC MIMO system. We mainly inspect this issue from the perspective of channel estimation performance. Fig. 8 and Fig. 5 show that under the condition of using High + All scheme and $\eta \leq 0.2$, equipping 3-bit ADCs as low-resolution ADCs can reduce NMSE by 50 percent compared to when 1-bit ADCs are used. Similar phenomenon can be observed when High + Mod scheme is employed. However, when High + Arg scheme is employed, the performance improvement brought by the bit increase is not obvious except when SNR is high. Therefore, when $\eta \leq 0.2$, it is preferable to use 2-bit ADCs or 3-bit ADCs for High + All or High + Mod scheme but for High + Arg scheme, it is not necessary to use 2-bit or 3-bit ADC as low-resolution ADC. When $\eta \geq 0.3$, it can be observed from Fig. 8 that increasing the number of bits of the low-resolution ADC has little effect on the performance for all schemes, thus using 2-bit or 3-bit ADC as low-resolution ADC in this scenario is unnecessary. However, when the cost and energy consumption conditions permit, using 2-bit or 3-bit ADCs as low-resolution ADCs can improve channel estimation performance. In this case, using 2bit and 3bit ADCs is better.

V. CONCLUSION

In this paper, two-stage approach and modified SIP-DNN approach are proposed for uplink channel estimation in mixed-ADC massive MIMO systems. By adopting the three low-resolution received pilot utilization schemes we proposed, the DNN in both approaches could exploit not only high-resolution-ADC-quantized but also low-resolution-ADC-quantized received pilot signals to improve estimation accuracy. Simulation results show that the proposed approaches outperform state-of-the-art SIP-DNN method and reduces the number of high-resolution ADCs required at base stations without performance degradation.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
[2] 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.
[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[4] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.
[5] J. Mo and R. W. Heath, Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2014, pp. 1–5.
[6] 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.

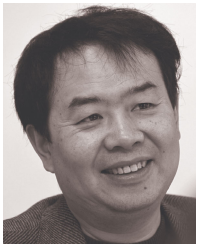
[7] N. Liang and W. Zhang, "Mixed-ADC massive MIMO uplink in frequency-selective channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.
[8] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.
[9] H. Pirzadeh and A. L. Swindlehurst, "Spectral efficiency under energy constraint for mixed-ADC MRC massive MIMO," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1847–1851, Dec. 2017, doi: 10.1109/LSP.2017.2761356.
[10] T. Liu, S. Jin, C.-K. Wen, M. Matthaiou, and X. You, "Generalized channel estimation and user detection for massive connectivity with mixed-ADC massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3236–3250, Jun. 2019, doi: 10.1109/TWC.2019.2912370.
[11] J. Yuan, Q. He, M. Matthaiou, T. Q. S. Quek, and S. Jin, "Toward massive connectivity for IoT in mixed-ADC distributed massive MIMO," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1841–1856, Mar. 2020, doi: 10.1109/JIOT.2019.2957281.
[12] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, "Deep CNN-based channel estimation for mmWave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 989–1000, Sep. 2019, doi: 10.1109/JSTSP.2019.2925975.
[13] M. Y. Takeda, A. Klautau, A. Mezghani, and R. W. Heath, Jr., "MIMO channel estimation with non-ideal ADCS: Deep learning versus GAMP," in *Proc. IEEE 29th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Pittsburgh, PA, USA, Oct. 2019, pp. 1–6, doi: 10.1109/MLSP.2019.8918799.
[14] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Deep learning for massive MIMO with 1-bit ADCs: When more antennas need fewer pilots," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1273–1277, Aug. 2020, doi: 10.1109/LWC.2020.2987893.
[15] E. Balevi and J. G. Andrews, "Two-stage learning for uplink channel estimation in one-bit massive MIMO," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1764–1768, doi: 10.1109/IEEECONF44664.2019.9048915.
[16] R. Zhu and G. Zhang, "A segment-average based channel estimation scheme for one-bit massive MIMO systems with deep neural network," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Xi'an, China, Oct. 2019, pp. 81–86, doi: 10.1109/ICCT46805.2019.8947071.
[17] D. H. N. Nguyen, "Neural network-optimized channel estimator and training signal design for MIMO systems with few-bit ADCs," *IEEE Signal Process. Lett.*, vol. 27, pp. 1370–1374, 2020, doi: 10.1109/LSP.2020.3012794.
[18] E. Balevi, A. Doshi, A. Jalal, A. Dimakis, and J. G. Andrews, "High dimensional channel estimation using deep generative networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 18–30, Jan. 2021, doi: 10.1109/JSAC.2020.3036947.
[19] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1989–1993, Nov. 2019.
[20] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017, doi: 10.1109/TWC.2017.2691318.
[21] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Gridless channel estimation for mixed one-bit antenna array systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8485–8501, Dec. 2018.

JIN ZICHENG received the bachelor's degree in information engineering from Southeast University, Nanjing, China, in 2018. He is currently pursuing the master's degree with Southeast University. His research interest includes channel estimation, MIMO, and machine learning.

GAO SHEN received the M.S. degree in communication and information system from Shandong University, Jinan, China, in 2015. She is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. From October 2017 to 2019, she was a Visiting Student with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Her research interests include machine learning for wireless communications, cache-enabled wireless networks, and massive MIMO.

LIU NAN (Member, IEEE) received the B.Eng. degree in electrical engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2001, and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 2007. From 2007 to 2008, she was a Postdoctoral Scholar with the Wireless Systems Laboratory, Department of Electrical Engineering, Stanford University. She became a Professor with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing, China, in 2009. Her research interests include network information theory for wireless networks, SON algorithms for next-generation cellular networks, and energy-efficient communications. She is an Associate Editor of *China Communications*.

PAN ZHIWEN (Member, IEEE) has been with the National Mobile Communications Research Laboratory, Southeast University, as an Associate Professor, since 2000, and a Professor, since 2004. From 2000 to 2001, he was involved in the research and standardization of 3G, and since 2002, he has been involved in the investigations on key technologies for IMT-A and 5G. He has published over 50 articles recently, and holds over 50 patents. His research interests include self-organizing networks, wireless networking, and radio transmission technology for wireless communications.



YOU XIAOHU (Fellow, IEEE) has been with the National Mobile Communications Research Laboratory, Southeast University, since 1990, where he was a Professor and the Chair Professor of the Cheung Kong Scholars Program, and served as the Director for the Laboratory. From 1999 to 2002, he was a Principal Expert of the C3G Project, responsible for organizing China's 3G mobile communications research and development activities. From 2001 to 2006, he was a Principal Expert of the National 863 Future Project. He has contributed over 50 IEEE journal articles and two books in the areas of adaptive signal processing, neural networks, and their applications to communication systems, and holds over 80 patents. His research interests include mobile communication systems, signal processing, and its applications. His current research interests include wireless and mobile communication systems, and modern digital signal processing. He was a Premier Foundation Investigator of the China National Science Foundation. He has been the Section Chair of the IEEE Nanjing Section, since 2010. He served as the General Co-Chair for the IEEE WCNC 2014.