sensors

MDPI

Article

Deep Learning-Based Channel Estimation for mmWave Massive MIMO Systems in Mixed-ADC Architecture

Rui Zhang [1], Weiqiang Tan [1,*], Wenliang Nie [2], Xianda Wu [3] and Ting Liu [4,5]

[1] School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China; 2112006270@e.gzhu.edu.cn
[2] School of Electronic and Information Engineering, Chongqing Three Gorges University, Chongqing 404000, China; niewenliang@sanxiau.edu.cn
[3] School of Electronics and Information Engineering, South China Normal University, Foshan 528000, China; xiandawu@m.scnu.edu.cn
[4] School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China; liuting@nuist.edu.cn
[5] National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
* Correspondence: wqtan@gzhu.edu.cn

Abstract: Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems can significantly reduce the number of radio frequency (RF) chains by using lens antenna arrays, because it is usually the case that the number of RF chains is often much smaller than the number of antennas, so channel estimation becomes very challenging in practical wireless communication. In this paper, we investigated channel estimation for mmWave massive MIMO system with lens antenna array, in which we use a mixed (low/high) resolution analog-to-digital converter (ADC) architecture to trade-off the power consumption and performance of the system. Specifically, most antennas are equipped with low-resolution ADC and the rest of the antennas use high-resolution ADC. By utilizing the sparsity of the mmWave channel, the beamspace channel estimation can be expressed as a sparse signal recovery problem, and the channel can be recovered by the algorithm based on compressed sensing. We compare the traditional channel estimation scheme with the deep learning channel-estimation scheme, which has a better advantage, such as that the estimation scheme based on deep neural network is significantly better than the traditional channel-estimation algorithm.

Keywords: millimeter-wave; massive MIMO; channel estimation; deep learning; mixed resolution ADC; approximate message passing; compressed sensing

check for updates

Citation: Zhang, R.; Tan, W.; Nie, W.; Wu, X.; Liu, T. Deep Learning-Based Channel Estimation for mmWave Massive MIMO Systems in Mixed-ADC Architecture. *Sensors* **2022**, *22*, 3938. https://doi.org/10.3390/s22103938

Academic Editor: Davy P. Gaillot

Received: 20 April 2022
Accepted: 19 May 2022
Published: 23 May 2022

1. Introduction

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) technology can significantly improve data transmission rate with wider bandwidth and higher spectral efficiency, becoming one of the key technologies for the sixth generation (6G) wireless communication [1]. However, digital beamforming and full-resolution ADCs are not suitable in mmWave massive MIMO system, due to the power consumption of ADCs scales being exponential with the number of quantization bits, leading to high hardware cost, power consumption and system complexity. Therefore, the use of current high-speed and high-resolution ADCs (8–12 bits) for each antenna array would become a great burden to the base station (BS) [2–4]. Consequently, the use of low-cost and low-resolution ADCs (1–4 bits) is promoted as a potential solution to this problem.

1.1. Related Works

In order to reap the maximum benefits of mmWave massive MIMO, the researchers have proposed hybrid precoding, which involves a massive number of antennas attached with a few RF chains [5,6], which effectively improved the energy efficiency of systems.

In addition, the authors of [7] utilized the lens antenna array at the BS in the mmWave massive MIMO system, in which signals from different directions can be concentrated on different antennas, and the spatial channel can be converted into a beamspace channel. Thus, the mmWave beamspace channel is sparse and there are only a few main propagation path gains. By selecting a small number of main beams, the number of RF chains required by the massive MIMO system can be significantly reduced [8,9]. Unfortunately, the beam selection requires the BS to obtain the channel state information (CSI), and this is difficult to achieve, especially when the number of RF links is limited. Furthermore, the performance of hybrid precoding systems relies heavily on the precise control of the analog components, and also the selection of the optimal beam will be more difficult if the beam width is small.

In parallel, there is a common solution to replace high-resolution ADCs with low-resolution ADC (1–4 bits), which means the deployment of pure low-resolution ADCs at the BS. However, low-resolution ADCs are prone to severe non-linear distortion, which inevitably causes several problems, including high pilot overhead for channel estimation [10], the system performance loss and signal detection [11]. In order to balance the cost and performance of the system, a mmWave massive MIMO system with a mixed ADC architecture was proposed in [12,13], which replaces the low-resolution ADCs with the partial high-resolution ADCs on the original basis. As reported in [14], channel estimation in a mixed-resolution ADC architecture is easier to process than in a pure low-resolution ADC system. The authors of [15] derived a closed-form approximation of the reachable rate of massive MIMO uplink under Rician fading channels with mixed-ADC. The work in [16] analyzed the sum-rate performance of the multi-user massive MIMO relaying system equipped with mixed-ADC architecture in the BS. Authors of [17] proposed the channel estimation algorithm for uplinking massive MIMO systems with mixed-ADC architecture. For ease of understanding, the advantages and limitations of the two solutions mentioned above are summarized in Table 1.

Table 1. Multiple solutions to the problem of overwhelming hardware costs and power consumption in mmWave massive MIMO system.

Solutions	Advantages	Limitations
Lens antenna arrays	Significant reduction in the number of RF chains by adopting a hybrid precoding scheme.	Beam selection requires the BS to obtain the CSI of beamspace, and system performance is vulnerable to bandwidth.
Low-resolution ADCs	Reduce power consumption by reducing the high resolution of the ADCs to low resolution.	Low-resolution ADCs are prone to severe non-linear distortion.

Different from using mixed-ADC architecture to reduce cost, another research direction focuses on using the sparsity of the mmWave channel to estimate the channel by using some classical schemes based on compressed sensing (CS) [18–21]. Specifically, the work in [18] used the orthogonal matching pursuit (OMP) algorithm to detect the dominant entries of multiple channel paths. The authors of [19] proposed a simultaneous weighted orthogonal matching pursuit (SWOMP) channel estimation scheme. Furthermore, the authors of [20] proposed stage wise orthogonal matching pursuit. Using the sparse characteristics of the mmWave channel, the work in [21] designed a compressive sampling matching pursuit, which aims to reduce system complexity. Unfortunately, the estimation accuracy of the greedy algorithm is not ideal in the low signal-to-noise ratio (SNR) range. As a powerful sparse signal recovery algorithm, the approximate message passing (AMP) algorithm was proposed in [22], which can be used to estimate the beamspace channel, but it is difficult to find the optimal solution for the shrinkage parameters of the algorithm.

Deep learning (DL) has achieved great success in speech recognition and image processing. The advantages of DL are expected to bring changes to the communication system. Compared with traditional methods, DL can reveal the internal characteristics of end-to-end collected data or signals, so as to better solve various complicated problems encountered in wireless communication [23–27]. Inspired by the powerful learning ability of deep neural networks (DNNs), some methods based on DL have been applied to channel estimation and achieved good results. For example, the work in [28] applied to channel estimation in the wireless downlink transmission system, which is superior to the traditional scheme in estimation accuracy and energy acquisition. The use of neural networks can help improve channel estimation performance was also demonstrated in [29], where the method was designed by the structure of the minimum mean squared error (MMSE) channel estimator. In addition, the authors of [30] proposed a framework based on DL for direction-of-arrival estimation and the massive MIMO system. In [31], the authors proposed an iterative channel estimation scheme, in which a denoising neural network is used to update the estimated channel in each iteration. The authors in [32] proposed a two-stage massive MIMO channel estimation process based on DL, including pilot-aided and data-aided estimation stages. In [33], a deep neural network based on long short-term memory (LSTM) was introduced to develop a more effective CSI feedback channel compression and recovery method. To improve the estimation accuracy, the authors of [34] developed a learned AMP (LAMP) network for channel estimation. Compared with the original AMP algorithm, the estimation accuracy of both algorithms is improved.

1.2. Contributions

In this paper, we investigate channel estimation for mmWave massive MIMO system with lens antenna array, in which we use a mixed (low/high) resolution ADC architecture. To the best of our knowledge, there is no previous work exploring a mixed-ADC architecture of mmWave massive MIMO system and how to estimate CSI based on the DL method. We first study that the mixed-ADC architecture is practically useful because it can provide performance comparable to the ideal high-resolution ADC architecture, while reducing the complexity and power consumption of signal processing. Specifically, the main contributions of this work can be summarized as follows:

- We study the DL-based channel estimation for a mmWave massive MIMO system with mixed-ADC architecture, where most antennas are equipped with low-resolution ADC and the rest of the antennas use high-resolution ADC. In terms of spectral efficiency, the results showed that the mixed-ADC architecture is practically useful and it can provide performance comparable to the ideal high-resolution ADC architecture.
- We compare the performance of traditional beamspace channel estimation algorithms and DL-based schemes in mixed-ADC architecture. By applying the sparsity of the mmWave channel, the DL-based channel estimation scheme performs significantly better than the conventional channel estimation algorithm and its complexity outperforms the conventional estimation algorithm.
- We evaluate the effect of different quantization bits of low-resolution ADC in the mixed-ADC architecture on the channel estimation and the sum rate. The results show that when the number of quantization is about 4 bit, the estimation error with the mixed-ADC architecture is small compared to that with the high-resolution ADC.

Notation: We use boldface letters to denote vectors and capitals to denote matrices. For matrix $\mathbf{A}$, $\mathbf{A}_{ij}$ is the (i,j) th entry of $\mathbf{A}$. $\mathbf{A}^T, \mathbf{A}^H$ represent $\mathbf{A}$'s transpose, complex conjugate. Moreover, a circularly symmetric complex Gaussian random vector with zero mean and covariance matrix $\mathbf{R}$ is denoted $v \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, $\mathbf{I}_n$ is the identity matrix of size n, $\otimes$ is the Kronecker product and $\|\cdot\|_2$ is the Euclidean norm.

2. System Model

We consider a time division duplex (TDD) mmWave massive MIMO system, in which the BS is equipped with N antennas and N_{RF} RF chains, serving K single antenna users simultaneously In this section, we first introduce the traditional mmWave massive MIMO channel and the space channel after adding the lens antenna array. In the second section, we introduce the mixed resolution ADC architecture. Finally, we express the beamspace channel estimation problem as the sparse signal recovery problem.

2.1. Beamspace Channel Model

We start with the traditional massive MIMO mmWave channel, and in this work, we adopt the Saleh–Valenzuela channel model, which is widely used in the frequency domain. The channel vector with the size of $N \times 1$ between the kth $(k = 1, 2, \ldots, K)$ user and the BS can be expressed as

$$\mathbf{h}_k = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \beta_{k,l} \mathbf{a}(\phi_{k,l}, \theta_{k,l}) = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \mathbf{c}_{k,l}, \tag{1}$$

where N is the total number of antennas, L_k is the number of resolvable paths and $\beta_{k,l}$, $\phi_{k,l}$ and $\theta_{k,l}$ are the complex gain, azimuth and elevation on the lth path, respectively. $\mathbf{a}(\phi_{k,l}, \theta_{k,l})$ is the $N \times 1$ array steering vector, which depends on the array geometry, and $\mathbf{c}_{k,l} = \beta_{k,l} \mathbf{a}(\phi_{k,l}, \theta_{k,l})$ denotes the lth path component. For a typical uniform planar arrays (UPAs) with $N_1 \times N_2$ $(N = N_1 \times N_2)$ antennas, the array steering vector can be expressed as

$$\mathbf{a}(\phi, \theta) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi d \sin(\phi) \sin(\theta) \mathbf{n}_1 / \lambda} \right] \otimes \left[e^{-j2\pi d \cos(\theta) \mathbf{n}_2 / \lambda} \right], \tag{2}$$

where $\mathbf{n}_1 = [0, 1, \cdots, N_1 - 1]^T$ and $\mathbf{n}_2 = [0, 1, \cdots, N_2 - 1]^T, \lambda$ is the wavelength of the carrier, and d is the antenna spacing usually satisfying $d = \lambda/2$ in mmWave communications. Then, we can define $\psi_{\mathrm{a}} = d \sin(\phi) \sin(\theta) / \lambda$ and $\psi_{\mathrm{e}} = d \cos(\theta) / \lambda$, respectively, as the spatial angles for UPAs configuration [35].

Traditional channels in the spatial domain can be converted to beam spatial channels by using a lens antenna array. In fact, the lens antenna array plays the role of the spatial discrete Fourier transform (DFT) matrix U of size $N \times N$. For UPA configuration, the matrix U can be expressed as

$$\mathbf{U} = \left[\mathbf{a}(\psi_{a,1}, \psi_{e,1}), \cdots, \mathbf{a}(\psi_{a,1}, \psi_{e,N_2}), \cdots, \mathbf{a}(\psi_{a,N_1}, \psi_{e,1}), \cdots, \mathbf{a}(\psi_{a,N_1}, \psi_{e,N_2}) \right]^H, \tag{3}$$

where $\psi_{a,n} = \frac{1}{N_1}\left(n - \frac{N_1+1}{2}\right)$ for $n = 1, 2, \cdots, N_1$ and $\psi_{e,n} = \frac{1}{N_2}\left(n - \frac{N_2+1}{2}\right)$ for $n = 1, 2, \cdots, N_2$ are the spatial direction corresponding to the azimuth and elevation predefined by the lens antenna array, respectively. Therefore, the beamspace channel vector $\overline{\mathbf{h}}_k$ with a size of $N \times 1$ between the kth user and BS with N antennas can be written as

$$\overline{\mathbf{h}}_k = \mathbf{U}\mathbf{h}_k = \sqrt{\frac{N}{L}} \sum_{l=1}^{L} \overline{\mathbf{c}}_{k,l} \tag{4}$$

where $\overline{\mathbf{c}}_{k,l} = \mathbf{U}\mathbf{c}_{k,l}$ represents the lth path component in the beamspace channel.

2.2. Mixed Resolution ADC Architecture

In this section, we consider a mixed resolution ADC architecture on the part of the lens antenna array. We divide the antenna into two parts, where the $N_0 = \sigma N$ antennas are connected to the high-resolution ADC, and the $N_1 = (1 - \sigma)N$ antenna is connected to lower ADC. In order to facilitate calculation, the antenna number N_0 and N_1 in the simulation are integers, and the coefficient σ is also limited to a certain rational number. In this experiment, σ is set to 0.25, so we can redivide the channel matrix and rewrite the above Formula (4) as

$$\tilde{\mathbf{h}}_k = \begin{bmatrix} \overline{h}_{k,0} \\ \overline{h}_{k,1} \end{bmatrix} = \mathbf{U} \begin{bmatrix} \mathbf{h}_{k,0} \\ \mathbf{h}_{k,1} \end{bmatrix}, \tag{5}$$

where $\overline{\mathbf{h}}_{k,0}$ represents the channel matrix associated with N_0 antennas connected to a high resolution ADC, and $\overline{\mathbf{h}}_{k,1}$ represents the channel matrix associated with N_1 antennas connected to a low resolution ADC.

Due to the reciprocity of the TDD channel, after the pilot sequence is transmitted to BS, the user can obtain the downlink channel through the estimated uplink channel. This paper adopts the widely used orthogonal pilot transmission strategy, and the channel estimation of each user is independent, so the subscript k in Formula (4) can be omitted.

Assuming that all users transmit known pilot sequences through instance Q for channel estimation, the measured signal $\mathbf{y}_q$ with a size of $N \times 1$ passing through the RF chains at the qth instant can be expressed as

$$\mathbf{y}_q = \mathbf{A}_q \tilde{\mathbf{h}} s_q + \overline{\mathbf{n}}_q, \quad q = 1, 2, \cdots, Q, \tag{6}$$

where $\mathbf{A}_q$ is the $N_{\mathrm{RF}} \times N$ beam-selection network, s_q is the pilot transmitted symbol at the qth instant, $\overline{\mathbf{n}}_q = \mathbf{A}_q \mathbf{n}_q$ is the effective noise vector, where $\mathbf{n}_q \sim \mathcal{CN}\left(0, \sigma_n^2 \mathbf{I}_N\right)$ is the $N \times 1$ noise vector with σ_n^2 representing the noise power.

After the pilot transmission of Q instances, we can obtain the $QN_{\mathrm{RF}} \times 1$ overall measurement vector $\mathbf{y}$ by assuming $s_q = 1$ for $q = 1, 2, \cdots, Q$ as

$$\mathbf{y} = \left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_Q\right]^T = \mathbf{A}\tilde{\mathbf{h}} + \mathbf{n}, \tag{7}$$

where $\mathbf{A} = \left[\mathbf{A}_1^T, \mathbf{A}_2^T, \cdots, \mathbf{A}_Q^T\right]^T$ is the $QN_{\mathrm{RF}} \times N$ overall combining matrix, and $\mathbf{n} = \left[\overline{\mathbf{n}}_1^T, \overline{\mathbf{n}}_2^T, \cdots, \overline{\mathbf{n}}_Q^T\right]^T$ is the effective noise vector for Q instants.

2.3. Problem Formulation

According to Formula (7), we can now recover channel $\tilde{\mathbf{h}}$ from $\mathbf{y}$ and $\mathbf{A}$. Due to the limited scattering at mmWave frequency, there are only a few propagation paths, and the beam space channel $\tilde{\mathbf{h}}$ is approximately sparse. This problem can be solved by the sparse signal recovery algorithm in compressed sensing (CS), in which matrix $\mathbf{A}$ in Formula (7) can be regarded as the sensing matrix in CS.

$$\min \|\tilde{\mathbf{h}}\|_0, \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{h}}\|_2 \leq \varepsilon, \tag{8}$$

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

Because the l_0-norm minimization problem is a NP-hard problem in practice, so in many cases, the l_0-norm optimization problem will be converted to a higher-dimension norm problem, such as replacing l_0-norm with l_1-norm for the convex optimization problem. At present, some traditional greedy algorithms are commonly used to solve this problem; for example, the OMP [18] and compressive sampling matching pursuit (CoSaMP) [21]. However, using these greedy algorithms it is difficult to find the global optimal solution and the estimation accuracy is not ideal.

3. Traditional Channel Estmation Algorithems

3.1. Compressed Sensing

Compressed sensing is a new technology for finding sparse solutions of underdetermined linear systems. By using the characteristics of signal sparsity, compared with Nyquist theory, the technology can restore the original signal to be recognized from fewer measured values. The theory mainly includes three parts: signal-sparse representation, reconstruction conditions and signal recovery algorithm. Suppose a signal $\boldsymbol{x} \in \mathbb{R}^{N \times 1}$ can be sparsely represented by an orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$, i.e,

$$\boldsymbol{x} = \mathbf{\Psi}\mathbf{s}, \tag{9}$$

where $\mathbf{s}$ is a sparse column vector, the sparse coefficient of $\boldsymbol{x}$ is K and K is far less than N. Consider a linear measurement process; $\mathbf{x}$ is represented by an orthonormal basis $\mathbf{\Psi}$ and a coefficient vector $\mathbf{s}$, and the measured value $\mathbf{y} \in \mathbb{R}^{M\times 1}(M << N)$ can be written as

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{x} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s} = \mathbf{A}\mathbf{s}, \tag{10}$$

where $\mathbf{\Phi}$ represents the measurement matrix, and $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$ is the sensing matrix.

In general, taking the sparse of the reconstructed signal in a certain transform domain as a priori information, the original signal is observed with the measurement matrix, and the complete measured signal is reconstructed from the observed value combined with the reconstruction algorithm.

3.2. OMP (Orthogonal Matching Pursuit)

Orthogonal Matching Pursuit is one of the classic algorithms in the field of compressed sensing. It is the basis of many commonly used efficient algorithms. This algorithm has the characteristics of simplicity and efficiency. The essence of the OMP algorithm is to select the columns in the sensor matrix by greedy iteration so that the selected column is most related to the current redundant vector in the process of each iteration, subtract the relevant part from the original signal vector, and iterate repeatedly until the number of iterations reaches the sparsity K, and then stop the iteration.

We define $\mathbf{A} \in \mathbb{C}^{M\times N}$ as a sensing matrix and $\boldsymbol{a}_i$ is the ith column of $\mathbf{A}$, observation vector $\mathbf{y} \in \mathbb{C}^{M\times 1}$ and sparsity K, the initialization residual $\mathbf{r}_0 = \mathbf{y}$. The initial iterated index collection $\Lambda_0 = \varnothing$, and initial value of iteration $k = 1$. To solve the first problem, we attempt to look for an iterative index as follows:

$$\lambda_k = \arg \max_{i=1,\cdots,N} |\langle \mathbf{r}_{k-1}, \boldsymbol{a}_i \rangle|. \tag{11}$$

Adds the index of the most relevant dictionary element found to the index set, while the set of reconstructed atoms in the sensing matrix is updated:

$$\Lambda_k = \Lambda_{k-1} \cup \{\lambda_k\}, \tag{12}$$

$$\mathbf{A}_{\Lambda_k} = \mathbf{A}_{\Lambda_{k-1}} \cup \boldsymbol{a}_{\Lambda_k}, \tag{13}$$

Update the residual:

$$\mathbf{r}_k = \mathbf{y} - \mathbf{A}_{\Lambda_k}\left(\mathbf{A}_{\Lambda_k}^T \mathbf{A}_{\Lambda_k}\right)^{-1} \mathbf{A}_{\Lambda_k}^T \mathbf{y} = \mathbf{y} - \mathbf{A}_{\Lambda_k}\mathbf{s}_k. \tag{14}$$

After executing K cycles, the reconstructed sparse coefficient $\mathbf{s}_k$ can be obtained. For more specific OMP algorithm flow, please refer to Algorithm 1.

3.3. AMP (Approximate Message Passing)

The approximate message passing (AMP) algorithm is an iterative compressed sensing approach based on a probability graph to predict the next iteration through state evolution and to de-noise through soft threshold iteration. In this section, we will describe how the AMP algorithm estimates the beamspace channel, as shown in Algorithm 2.

In Algorithm 2, α in Step 6 is a tuning parameter and usually takes a fixed value during iteration, the term $b_t \mathbf{v}_{t-1}$ in Step 3 is called Onsager Correction, which is introduced into the AMP algorithm to accelerate the convergence. The key step of the AMP algorithm is step 6, in which the estimated channel $\hat{\mathbf{h}}_{t+1}$ is obtained by soft threshold shrinking function η_{st} during the t^{th} iteration. The shrinkage function η_{st} is non-linear element-wise operation, due to the sparsity of the channel, the channel vector $\hat{\mathbf{h}}_{t+1}$ updated in each iteration will be more sparse. For the i th element $r_{t,i} = |r_{t,i}|e^{j\omega_{t,i}} (i = 1, 2, \cdots, N)$ of input vector $\mathbf{r}_t$, we have

$$\begin{aligned}[\boldsymbol{\eta}_{\text{st}}(\mathbf{r}_t; \alpha, \lambda_t)]_i &= \eta_{\text{st}}\left(|r_{t,i}|e^{j\omega_{t,i}}; \alpha, \lambda_t\right) \\ &= \max(|r_{t,i}| - \alpha\lambda_t, 0)e^{j\omega_{t,i}},\end{aligned} \tag{15}$$

where $\omega_{t,i}$ is the phase of complex-valued element $r_{t,i}$, α is the fixed parameter in the T iteration, and λ_t is updated with iteration process in Step 5. In addition, b is obtained by calculating the element-wise derivatives of the shrinkage function at the input vector r in Step 7.

Algorithm 1 OMP Algorithm in mmWave Channels

```
Require:
    sensing matrix $\mathbf{A}$, measurement vector $\mathbf{y}$, and the threshold $\delta$
 1: $\Lambda_0 = \varnothing, \mathbf{r}_0 = \mathbf{y}, \mathbf{r}_{-1} = 0$, and $k = 1$
 2: while$\|\mathbf{r}_{k-1} - \mathbf{r}_{k-2}\|_2^2 > \delta$ and $k \leq k_{\max}$ do
 3:     $\lambda_k = \underset{i=1,\cdots,N}{\operatorname{argmax}} \left|\left\langle \mathbf{A}^H(i), \mathbf{r}_{k-1} \right\rangle\right|$
 4:     $\Lambda_k = \Lambda_{k-1} \cup \{\lambda_k\}$
 5:     $\mathbf{A}_{\Lambda_k} = \mathbf{A}_{\Lambda_{k-1}} \cup \mathbf{a}_{\Lambda_k}$
 6:     $\mathbf{s}_k = \left(\mathbf{A}_{\Lambda_k}^T \mathbf{A}_{\Lambda_k}\right)^{-1} \mathbf{A}_{\Lambda_k}^T \mathbf{y}$
 7:     $\mathbf{r}_k = \mathbf{y} - \mathbf{A}_{\Lambda_k}\mathbf{s}_k$
 8:     $k = k + 1$
 9: end while
10: return $\mathbf{s}_k$
```

Algorithm 2 AMP Algorithm in mmWave Channels

```
Require:
    sensing matrix $\mathbf{A}$, measurement vector $\mathbf{y}$, and the number of iterations $T$
 1: Initialize $\hat{\mathbf{h}}_0 = 0, \mathbf{v}_{-1} = 0$, and $b_0 = 0$
 2: for $t = 0, 1, \cdots, T - 1$ do
 3:     $\mathbf{v}_t = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_t + b_t\mathbf{v}_{t-1}$
 4:     $\mathbf{r}_t = \hat{\mathbf{h}}_t + \mathbf{A}^T\mathbf{v}_t$
 5:     $\lambda_t = \frac{1}{\sqrt{M}}\|\mathbf{v}_t\|_2$
 6:     $\hat{\mathbf{h}}_{t+1} = \boldsymbol{\eta}_{\mathbf{st}}(\mathbf{r}_t; \alpha, \lambda_t)$
 7:     $b_{t+1} = \frac{1}{M}\sum_{i=1}^{N} \frac{\partial\eta_{\text{st}}(\mathbf{r}_t;\alpha,\lambda_t)}{\partial r_{t,i}}$
 8: end for
 9: return $\hat{\mathbf{h}} = \hat{\mathbf{h}}_T$.
```

Although the AMP algorithm can effectively deal with massive sparse signal problems, and performs well in many traditional channel-estimation algorithm schemes, for sparse beamspace channel estimation, many problems still exist, such as that the AMP algorithm has a high requirement for i.i.d. sub-Gaussian matrix A. Otherwise, the algorithm itself is prone to divergence. There are two key problems that restrict the performance of the AMP algorithm: (1) The shrinkage parameters in the AMP algorithm usually take the same value in the whole iterative process; (2) the general AMP algorithm cannot make full use of the prior distribution of beamspace channels.

4. AI Channel Estmation Algorithems

4.1. Deep Learning

Deep learning is a kind of representation learning method based on data in machine learning. Its basis is the neural network. In DL, training data $\left\{\left(y^d, x^d\right)\right\}_{d=1}^{D}$ consisting of feature and label pairs are used to train the parameters of the deep neural network, which aims to accurately predict the unknown label x associated with the newly acquired feature y. Depth networks accept y and process it in many layers, each of which usually consists of a linear transformation followed by a simple non-linear transformation. Unlike traditional shallow learning, in-depth learning emphasizes the depth of the model structure, usually with five, six or even ten layers of hidden nodes, and transforms the feature representation of samples in the original space into a new feature space through layer-by-layer feature transformation, which makes classification or prediction easier.

In general, the label space is discrete; for example, y is an audio or a picture; x is a class of cat, dog, or some other type. However, for the sparse linear problem in beam space, label x is continuous, but before that, some authors have demonstrated that a well-constructed deep neural network can predict such tags accurately.

4.2. LAMP Network

Recently, the authors of [34] proposed a LAMP network scheme based on the classical AMP algorithm. The results showed that the contraction parameters of the AMP algorithm can only take the same value during the iterative process. Therefore, it essence is to map each iteration of the AMP algorithm to each layer of the LAMP network and optimize the non-linear parameters α in each iteration.

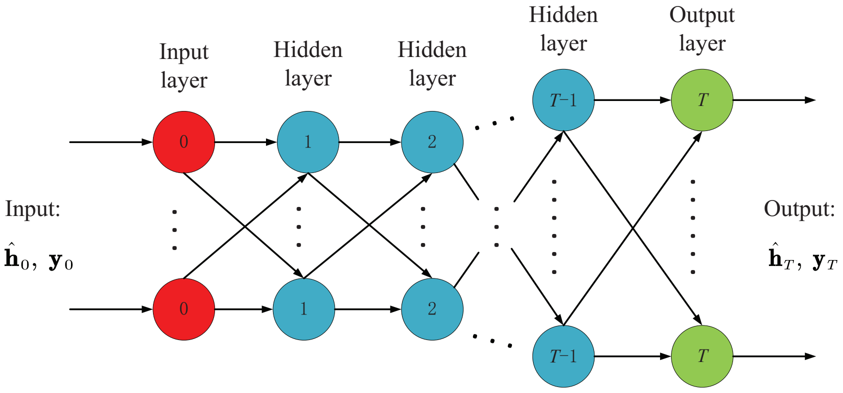

Figure 1. LAMP network structure with a total of T layers.

Figure 1 shows the network structure of the LAMP scheme with a total of T layers. Specifically, the LAMP network processes signals in the same way as the AMP algorithm, where the input of tth layer is $\mathbf{y}$, $\hat{\mathbf{h}}_t$ and $\mathbf{v}_t$. It is worth mentioning that $\mathbf{y}$ is the measurement signal, and both $\hat{\mathbf{h}}_t$ and $\mathbf{v}_t$ are $(t-1)$th layer outputs, so the processing of signals in tth layer can be summarized as follows:

$$\hat{\mathbf{h}}_{t+1} = \boldsymbol{\eta}_{\mathbf{st}}(\mathbf{r}_t; \alpha_t, \lambda_t), \tag{16}$$

$$\mathbf{v}_{t+1} = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_{t+1} + b_{t+1}\mathbf{v}_t, \tag{17}$$

where

$$\lambda_t = \frac{1}{\sqrt{M}}\|\mathbf{v}_t\|_2, \tag{18}$$

$$\mathbf{r}_t = \hat{\mathbf{h}}_t + \mathbf{B}^T\mathbf{v}_t, \tag{19}$$

$$b_{t+1} = \frac{1}{M}\sum_{i=1}^{N}\frac{\partial\eta_{\mathrm{st}}(\mathbf{r}_t; \alpha, \lambda_t)}{\partial r_{t,i}}. \tag{20}$$

From the above Formulas (16) and (19), we see the LAMP network involved in the learnable parameters compared with Algorithm 2. In the tth iteration, operations involving $(\alpha,\ \mathbf{A}^T)$ are replaced by $(\alpha_t,\ \mathbf{B}^T)$, and the shrinkage function $\boldsymbol{\eta}_{\mathbf{st}}$ of the AMP algorithm plays a role in the non-linear activation function in the conventional DNN. It is worth noting that selecting $\mathbf{A}^T$ in the AMP algorithm is only for the convenience of formula derivation. If enough training data are given, the LAMP network can use DNN's powerful learning ability to find better shrinking parameters. Thus, the performance of the original AMP algorithm can be further improved by optimizing the linear transformation coefficient $\mathbf{B}^T$ and non-linear shrinking parameter α_t.

4.3. GM-LAMP Network

The LAMP network solves the problem of taking fixed shrinkage parameters in the AMP algorithm, but it does not make good use of prior information on the beamspace channel. The GM-LAMP network proposed by [36] solves this problem. The GM-LAMP network derives a new shrinking function by considering the Gaussian mixing distribution of the elements of the beamspace channel. Before that, we first introduce the expression of the probability density function of the element $\tilde{h}$ in the beamspace channel:

$$p(\tilde{h};\boldsymbol{\theta}) = \sum_{k=0}^{N_c-1} p_k \mathcal{CN}\left(\tilde{h};\mu_k,\sigma_k^2\right) \tag{21}$$

where $\boldsymbol{\theta} = \left\{p_0,\cdots,p_{N_c-1},\mu_0,\cdots,\mu_{N_c-1},\sigma_0^2,\cdots,\sigma_{N_c-1}^2\right\}$ is a set of all the parameters, p_k is the probability of the kth Gaussian component and N_c is the number of Gaussian components in the Gaussian mixture distribution. μ_k and σ_k^2 represent the mean and variance of the kth Gaussian component, respectively. $\mathcal{CN}\left(\tilde{h};\mu_k,\sigma_k^2\right) = \frac{1}{\pi\sigma_k^2}\mathrm{e}^{-\frac{(\tilde{h}-\mu_k)^*(\tilde{h}-\mu_k)}{\sigma_k^2}}$ denotes the probability density function of the k th Gaussian component.

It is interesting that when the mean and variance of Gaussian components are zero, the probability density function of Gaussian distribution can be rewritten as

$$\mathcal{CN}(\tilde{h};0,0) = \delta(\tilde{h}),$$

where the $\delta(\tilde{h})$ is the Dirac delta function, which means that the variable $\tilde{h}$ will be exactly zero. Therefore, the sparsity of beam space channel can be described as a special case by using Gaussian mixture distribution. Finally, the Gaussian Mixed Shrinkage Function considering the prior distribution of the beamspace channel can be written as

$$\eta_{\mathrm{gm}}\left(r;\boldsymbol{\theta},\sigma^2\right) = \frac{\sum_{k=0}^{N_c-1} p_k\tilde{\mu}_k(r)\mathcal{CN}\left(r;\mu_k,\sigma^2+\sigma_k^2\right)}{\sum_{k=0}^{N_c-1} p_k\mathcal{CN}\left(r;\mu_k,\sigma^2+\sigma_k^2\right)}, \tag{22}$$

where a set of all distribution parameters $\boldsymbol{\theta}$ can also be called as the shrinkage parameters.

GM-LAMP network is still based on AMP algorithm, and it has T uniform layers, in which the input and output of each layer are the same as that of LAMP network, and the difference from LAMP network is that its soft threshold shrinkage function is replaced by the Gaussian mixture shrinkage function. The channel estimation for the tth layer can be written as

$$\mathbf{r}_t = \hat{\mathbf{h}}_t + \mathbf{B}^T\mathbf{v}_t, \tag{23}$$

$$\hat{\mathbf{h}}_{t+1} = \boldsymbol{\eta}_{\mathrm{gm}}\left(\mathbf{r}_t;\boldsymbol{\theta}_t,\sigma^2\right), \tag{24}$$

where the linear transformation coefficient is $\mathbf{B}^T$ and the non-linear shrinkage parameter $\boldsymbol{\theta}_t$ is the variable that can be optimized in the training stage.

The GM-LAMP network is mainly divided into two stages: offline training and online estimation. During the offline training phase, a large amount of known training data are provided to optimize the overall trainable variable by minimizing the loss function. In the online estimation phase, the new measurement data can be input into the trained GM-LAMP network and the corresponding channel estimates can be obtained directly.

In the offline training stage, supervised learning is adopted to train the GM-LAMP network, and the training dataset can be expressed as $\left\{\mathbf{y}^d, \tilde{\mathbf{h}}^d\right\}_{d=1}^{D}$, where $\mathbf{y}^d$ is the input of the GM-LAMP network, $\tilde{\mathbf{h}}^d$ is the corresponding label and D represents the number of the training data.

In order to avoid over-fitting, a layer-by-layer training method is adopted. Specifically, the whole training process can be divided into T training subroutines according to $0, 1, \cdots, T-1$ sequence. For the tth training subprocess, its objective is to optimize trainable variable $\boldsymbol{\Omega}_t = \{\mathbf{B}_i, \boldsymbol{\theta}_i\}_{i=0}^{t}$ of the $i = 0, \cdots, i = t^{th}$ layer. In this simulation, we refer to the GM-LAMP algorithm that was first proposed in the literature [36]. In the model training stage, the huber loss is used to define two loss functions of linear transformation coefficient $\mathbf{B}_t$ and non-linear shrinkage parameter $\boldsymbol{\theta}_t$:

$$L_t^{\text{linear}}(\boldsymbol{\Omega}_t) = \begin{cases} \frac{1}{2}(\mathbf{r}_t^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d)^2 & V_1 \le \delta \\ \delta|\mathbf{r}_t^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}, \tag{25}$$

$$L_t^{\text{non-linear}}(\boldsymbol{\Omega}_t) = \begin{cases} \frac{1}{2}(\hat{\mathbf{h}}_{t+1}^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d)^2 & V_2 \le \delta \\ \delta|\hat{\mathbf{h}}_{t+1}^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}, \tag{26}$$

where $V_1 = |\mathbf{r}_t^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d|$, $V_2 = |\hat{\mathbf{h}}_{t+1}^d\left(\mathbf{y}^d, \boldsymbol{\Omega}_t\right) - \tilde{\mathbf{h}}^d|$, δ is a hyperparameter, $\mathbf{r}_t^d$ is the output of linear transformation operation in Formula (23), and $\hat{\mathbf{h}}_{t+1}^d$ is the output of non-linear contraction operation in Formula (24). Based on these two loss functions, the training sub-process at the tth layer can be divided into linear training to minimize L_t^{linear} and non-linear training to minimize $L_t^{\text{non-linear}}$.

Algorithm 3 represents the specific layer-by-layer training method of the GM-LAMP network that was provided in [36]. To avoid the trapped-in local optimization caused by over-fitting, we firstly adopt the separate optimization method in step 2, and then jointly optimize the method in step 3 and 4, which are set to $\mathbf{B}_0$ and $\boldsymbol{\theta}_0$, respectively. Then the training is carried out in sequence from the 1th layer to the $(T-1)$th layer. For the training sub-process of the tth layer, the training variable is set to the value of the training variable of the $(t-1)$th layer before training. Steps 7–8 indicate that the linear transformation coefficient $\mathbf{B}_t$ is optimized separately, and then $\mathbf{B}_t$ and $\boldsymbol{\Omega}_t$ are optimized jointly. Similarly, steps 9–10 indicate that the non-linear shrinkage parameter $\boldsymbol{\theta}_t$ is optimized separately first, and then $\mathbf{B}_t$, $\boldsymbol{\Omega}_t$ and $\boldsymbol{\theta}_t$ are optimized jointly.

Algorithm 3 Layer-by-Layer Training Method

```
1:  Initialize $\mathbf{B}_0 = \mathbf{A}^T$ and $\boldsymbol{\theta}_0$
2:  Learn $\mathbf{B}_0$ by minimizing $L_t^{\text{linear}}$
3:  Learn $\boldsymbol{\theta}_0$ with fixed $\mathbf{B}_0$ by minimizing $L_t^{\text{non-linear}}$
4:  Relearn $\boldsymbol{\Omega}_0 = \{\mathbf{B}_0, \boldsymbol{\theta}_0\}$ to minimize $L_0^{\text{non-linear}}$
5:  for $t = 1, 2, \cdots, T-1$ do
6:  $\mathbf{B}_t = \mathbf{B}_{t-1}, \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$
7:  Learn $\mathbf{B}_t$ with fixed $\boldsymbol{\Omega}_{t-1}$ by minimizing $L_t^{\text{linear}}$
8:  Relearn $\{\mathbf{B}_t, \boldsymbol{\Omega}_{t-1}\}$ by minimizing $L_t^{\text{linear}}$
9:  Learn $\boldsymbol{\theta}_t$ with fixed $\{\mathbf{B}_t, \boldsymbol{\Omega}_{t-1}\}$ by minimizing $L_t^{\text{non-linear}}$
10: Relearn $\{\mathbf{B}_t, \boldsymbol{\Omega}_{t-1}, \boldsymbol{\theta}_t\}$ by minimizing $L_t^{\text{non-linear}}$
11: end for
12: return $\boldsymbol{\Omega}_{T-1}$
```

A trained GM-LAMP network can be obtained after optimizing the overall trainable variables $\boldsymbol{\Omega}_{T-1}$ of T layers. In the online estimation phase, the corresponding estimates can be directly generated by inputting new measurement signals into the trained GM-LAMP network.

4.4. Complexity Analysis

In this subsection, we perform the complexity analysis of the channel estimation algorithms used in the simulation process. Considering the OMP algorithm, when the sparsity of the channel vector is set to K, the computational complexity in the atom selection step is about $\mathcal{O}(KMN)$, and the computational complexity of the k-th iteration is at least $\mathcal{O}(K^3)$. Therefore, we ignore trivial operations; the computational complexity can be calculated as $\mathcal{O}(KMN) + \mathcal{O}\left(K^3 M\right)$. In addition, since the LAMP network and GM-LAMP network are constructed that both are based on the AMP algorithm, the computational complexity of the AMP algorithm, LAMP network and GM-LAMP network is roughly the same values that equals to $\mathcal{O}(TMN)$.

5. Numerical Results

In this section, we compare the performance of traditional beamspace channel estimation algorithms and DL-based schemes in mixed-ADC architecture. Specifically, DL-based schemes include the LAMP network and the GM-LAMP network. This experiment provides a widely used Saleh–Valenzuela channel model.

5.1. Parameter Setting

In our simulations, we consider that the BS is deployed a lens antenna array, where the numbers of antennas and RF chains are set to $N = 256$ and $N_{RF} = 16$, individually. The number of single-antenna users is set to $K = 16$, and the number of measurements is set to $M = 128$. For the the Saleh–Valenzuela channel model, we set the same channel parameters for each user k, where the number of channel paths $L_k = 3$, the complex gain on the lth path satisfies $\beta_{k,l} \sim \mathcal{CN}(0,1)$ for $l = 1,2,3$, and the range of azimuth $\phi_{k,l}$ and elevation $\theta_{k,l}$ on each path is between $-\pi/2$ and $\pi/2$. In order to train and test the LAMP network and the GM-LAMP network, we generate 80,000, 2000 and 2000 samples as the training, the validation and the testing set based on the above setup, respectively. Then the number of training layers is set as $T = 8$ for both network schemes, where the number of nodes in each layer depends on the number of measurements M and the number of dimensions N of the beamspace channel. Finally, we use the normalized mean square error (NMSE) to quantify the accuracy of channel estimation for each user, which is mathematically defined as

$$\mathrm{NMSE} = 10\log_{10}\left(\mathbb{E}\left[\frac{\|\hat{\mathbf{h}} - \tilde{\mathbf{h}}\|^2}{\|\tilde{\mathbf{h}}\|^2}\right]\right), \tag{27}$$

where $\hat{\mathbf{h}}$ is the recovered channel matrix using the channel estimation algorithms.

5.2. Simulation Results on the Saleh–Valenzuela Channel Model

Figure 2 shows the NMSE performance comparison of various algorithms under the considered UPA and mixed-resolution ADC quantization architecture. It can be seen from the figure that the performance of the traditional two algorithms is poor, and in general, the performance of the AMP algorithm is better than the OMP algorithm. For the DL-based LAMP network and GM-LAMP network, the performance of the scheme has been greatly improved. In addition, considering the prior distribution of beamspace channels, the GM-LAMP network has better channel estimation accuracy than the LAMP network.

Figure 3 shows a comparison of the complexity of the above four kinds of channel estimation schemes. We can find that the traditional OMP algorithm requires more complex multipliers than other schemes, because the OMP algorithm is a kind of greed-tracking algorithm, the number of complex multiplications increases exponentially with the number of antennas, so the OMP algorithm has an excessive demand on the number of antennas. In addition, thanks to the powerful learning of DNNs, the LAMP network and GM-LAMP network can converge faster than the AMP algorithm. Therefore, the complex multiplier required by the LAMP network and GM-LAMP network is smaller than that required by the AMP algorithm.

Next, we discuss the effects of resolution of ADC and channel sparsity on NMSE performance for two schemes based on DL. Figure 4 shows the performance changes of LAMP network schemes with different resolution ADC quantization. It can be seen that lower estimation error can be obtained by using higher-resolution ADC quantization, and the estimation performance after using mixed resolution ADC is better than four-bit ADC. With the increase in SNR, the performance difference between mixed resolution ADC and other low-resolution ADCs increases.

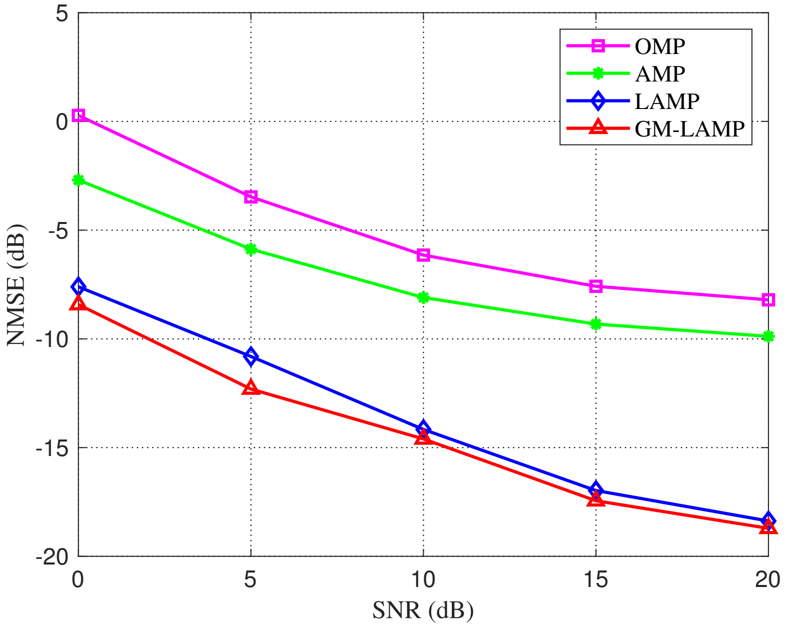

Figure 2. NMSE performance comparison of each channel estimation scheme for UPAs based on the Saleh-Valenzuela channel model in the mixed resolution ADC architecture.

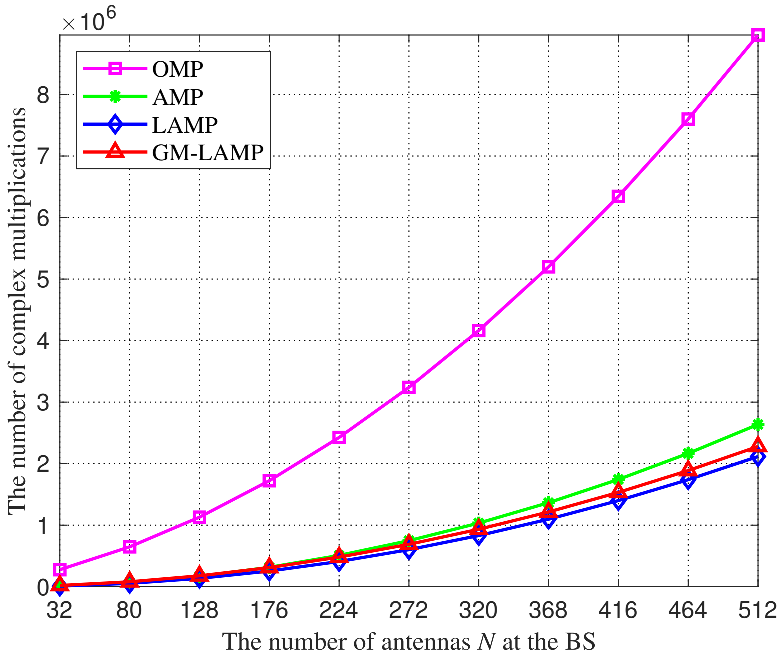

Figure 3. The number of complex multiplications against the number of antennas N.

Figure 5 plots the variation of estimation error in low-resolution ADCs from one to four bits and mixed resolution after adopting the GAMP network scheme. The sparsity in this figure is equivalent to the number of channel paths (L_k). With the increase in sparsity, the channel estimation performance of all quantization resolutions will decline. It can be seen from the four curves in the comparison that the higher the resolution, the smaller the estimation error.

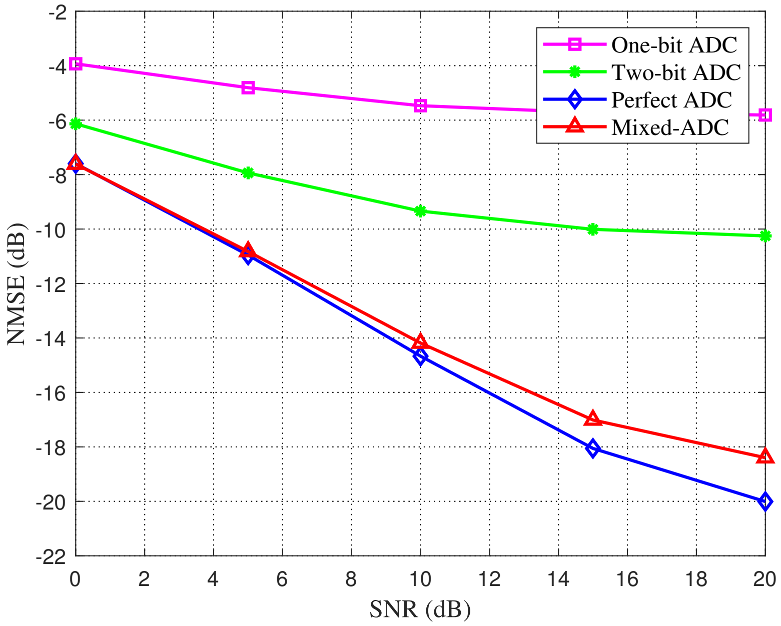

Figure 4. Channel estimation error is a function of SNR. Variations of estimation errors in different resolution ADCs quantification in LAMP network schemes are plotted.

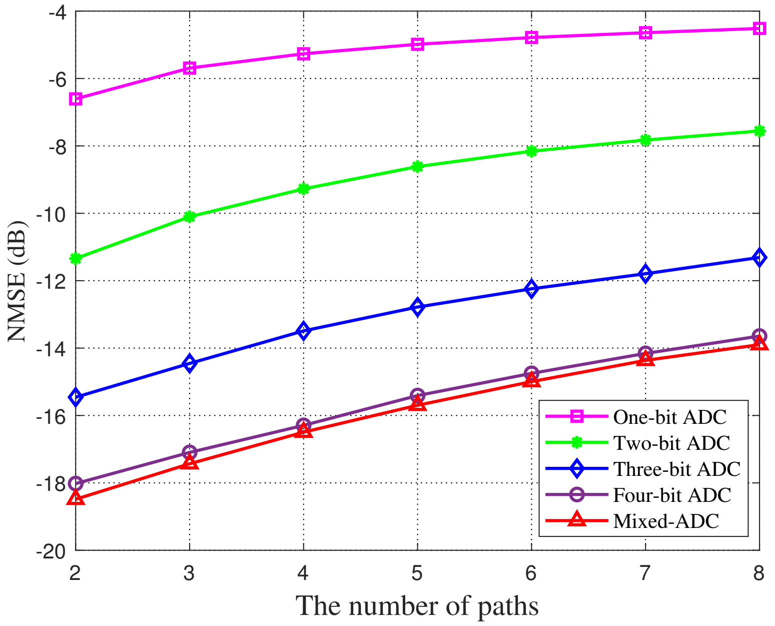

Figure 5. Channel estimation error is a function of channel sparsity. Variations of estimation errors in different resolution ADCs quantification in GAMP network schemes are plotted.

In addition, we evaluate the effects of NMSE and SNR on the total rate of beam selection in beamspace channel estimation. In this simulation, we reference the parameters of [37] to model the estimated beamspace channel $\hat{\mathbf{H}}$ in case of imperfect CSI as

$$\hat{\mathbf{H}} = \alpha\tilde{\mathbf{H}} + \sqrt{1-\alpha}\mathbf{E} \tag{28}$$

where $\alpha \in (0,1)$ is the error parameter, $\mathbf{E}$ denotes the error matrix, whose elements satisfies the independent and identical distribution of the zero means and variance error, i.e., $\mathcal{CN}(0,$

NMSE), and $\tilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_K\right]$ denotes the channel matrix with perfect CSI for K users. In order to demonstrate the effectiveness of this work, we use the results provided in [37]; that is, AI beam selection with perfect CSI as the benchmark.

In Figure 6, we consider UPA based on the Saleh-Valenzuela channel model, and compare the sum-rate between an imperfect CSI and a perfect CSI with a downlink SNR of 10 dB. As shown in Figure 6, for the imperfect CSI under the mixed-resolution ADC architecture, when the NMSE is about −23 dB, the total rate-loss due to beam selection is less than 5% compared to the perfect CSI, and the beam-selection rate quantized by a one-bit ADC is always quite different from that of the mixed-resolution ADC. Figure 7 shows the effect of the change of SNR on the total rate of beam selection for an imperfect CSI and a perfect CSI when the NMSE is set to −10 dB. We can clearly see that with the increase in SNR, the beam selection rate of all three increases, but the gap between the rate of imperfect CSI and perfect CSI gradually increases.

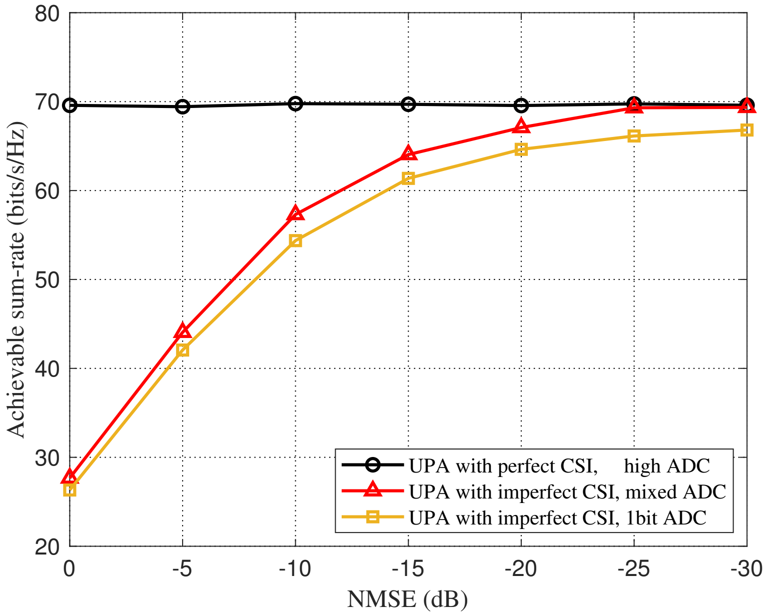

Figure 6. Total rate for beam selection against different NMSE for the beamspace channel estimation.

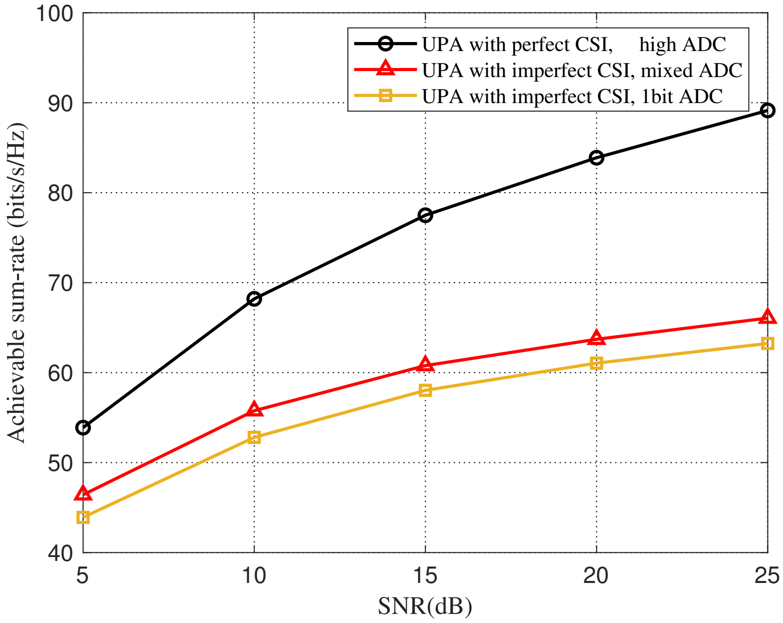

Figure 7. Total rate for beam selection against different SNR for the beamspace channel estimation.

6. Conclusions

In this paper, the channel estimation for a mmWave massive MIMO system with a mixed-resolution ADC architecture was investigated, where the BS deployed the lens antenna arrays. By using the sparsity of the mmWave channel, the beamspace channel estimation can be expressed as a sparse signal-recovery problem, and the channel can be recovered by the algorithm based on compressed sensing. We compare the traditional channel-estimation scheme with the channel-estimation scheme using DL. Simulation results showed that the performance of the DNN-based estimation scheme is significantly better than that of the traditional estimation scheme in the same configuration. Furthermore, the performance gap between using mixed-resolution ADC and using high-resolution ADC is slowly closing as the SNR increases. In future work, we will investigate the performance of IRS-assisted mmWave massive MIMO systems with mixed-ADC architectures, which aims to study the channel estimation problems in cascaded channels.

Author Contributions: W.T. conceived and designed the idea. R.Z. and X.W. performed the experiments and analyzed the data. W.N. and T.L. gave valuable suggestions on the structuring of the paper and assisted in the revising and proofreading. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Guangdong Province of China under Grant 2020A1515010484, the Guangzhou Science and Technology Planning Project under Grant 202102020879, the 2023 Guangzhou Basic Research Program Municipal School (College) Joint Funding Project, the 2022 Joint Research Fund for Guangzhou University and Hong Kong University of science and technology, the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2021D03, the National Nature Science Foundation of China under Grant 62101274, the Nature Science Foundation of Jiangsu Province under Grant BK20210640, and Scientific Research Cultivation Project for Young Scholars of South China Normal University under Grant 21KJ07.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to extend their gratitude to the anonymous reviewers and the editors for their valuable and constructive comments, which have greatly improved the quality of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Giordani, M.; Polese, M.; Mezzavilla, M.; Rangan, S.; Zorzi, M. Toward 6G networks: Use cases and technologies. *IEEE Commun. Mag.* **2020**, *58*, 55–61. [CrossRef]
2. Fozi, M.; Sharafat, A.R.; Bennis, M. Fast MIMO Beamforming via Deep Reinforcement Learning for High Mobility mmWave Connectivity. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 127–142. [CrossRef]
3. Zhang, J.; Liu, J.; Ma, S.; Wen, C.K.; Jin, S. Large System Achievable Rate Analysis of RIS-Assisted MIMO Wireless Communication With Statistical CSIT. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 5572–5585. [CrossRef]
4. Chen, S.; Zhang, J.; Jin, Y.; Ai, B. Wireless powered IoE for 6G: Massive access meets scalable cell-free massive MIMO. *China Commun.* **2020**, *17*, 92–109. [CrossRef]
5. Zeng, Y.; Zhang, R. Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm. *IEEE Trans. Commun.* **2016**, *64*, 1557–1571. [CrossRef]
6. Zeng, Y.; Zhang, R.; Chen, Z.N. Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1194–1206. [CrossRef]
7. Brady, J.; Behdad, N.; Sayeed, A.M. Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements. *IEEE Trans. Antennas Propag.* **2013**, *61*, 3814–3827. [CrossRef]
8. Gao, X.; Dai, L.; Chen, Z.; Wang, Z.; Zhang, Z. Near-optimal beam selection for beamspace mmWave massive MIMO systems. *IEEE Commun. Lett.* **2016**, *20*, 1054–1057. [CrossRef]
9. Amadori, P.V.; Masouros, C. Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection. *IEEE Trans. Commun.* **2015**, *63*, 2212–2223. [CrossRef]

10. Mo, J.; Schniter, P.; Heath, R.W. Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs. *IEEE Trans. Signal Process.* **2017**, *66*, 1141–1154. [CrossRef]
11. Zhang, J.; Dai, L.; Sun, S.; Wang, Z. On the spectral efficiency of massive MIMO systems with low-resolution ADCs. *IEEE Commun. Lett.* **2016**, *20*, 842–845. [CrossRef]
12. Tan, W.; Li, S.; Zhou, M. Spectral and energy efficiency for uplink massive MIMO systems with mixed-ADC architecture. *Phys. Commun.* **2022**, *50*, 101516. [CrossRef]
13. Zhou, M.; Zhang, Y.; Qiao, X.; Tan, W.; Yang, L. Analysis and optimization for downlink cell-free massive MIMO system with mixed DACs. *Sensors* **2021**, *21*, 2624. [CrossRef] [PubMed]
14. Liang, N.; Zhang, W. Mixed-ADC massive MIMO uplink in frequency-selective channels. *IEEE Trans. Commun.* **2016**, *64*, 4652–4666. [CrossRef]
15. Zhang, J.; Dai, L.; He, Z.; Jin, S.; Li, X. Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1327–1338. [CrossRef]
16. Liu, J.; Xu, J.; Xu, W.; Jin, S.; Dong, X. Multiuser massive MIMO relaying with mixed-ADC receiver. *IEEE Signal Process. Lett.* **2016**, *24*, 76–80. [CrossRef]
17. Yuan, J.; He, Q.; Matthaiou, M.; Quek, T.Q.; Jin, S. Toward massive connectivity for IoT in mixed-ADC distributed massive MIMO. *IEEE Internet Things J.* **2019**, *7*, 1841–1856. [CrossRef]
18. Venugopal, K.; Alkhateeb, A.; Heath, R.W.; Prelcic, N.G. Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 6493–6497.
19. Rodríguez-Fernández, J.; González-Prelcic, N.; Venugopal, K.; Heath, R.W. Frequency-Domain Compressive Channel Estimation for Frequency-Selective Hybrid Millimeter Wave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2946–2960. [CrossRef]
20. Lee, D. MIMO OFDM channel estimation via block stagewise orthogonal matching pursuit. *IEEE Commun. Lett.* **2016**, *20*, 2115–2118. [CrossRef]
21. Jayanthi, P.; Ravishankar, S. Sparse channel estimation for MIMO-OFDM systems using compressed sensing. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; pp. 1060–1064.
22. Donoho, D.L.; Maleki, A.; Montanari, A. Message passing algorithms for compressed sensing: I. motivation and construction. In Proceedings of the 2010 IEEE information theory workshop on information theory (ITW 2010, Cairo), Cairo, Egypt, 6–8 January 2010; pp. 1–5.
23. Qin, Z.; Ye, H.; Li, G.Y.; Juang, B.H.F. Deep learning in physical layer communications. *IEEE Wirel. Commun.* **2019**, *26*, 93–99. [CrossRef]
24. O'shea, T.; Hoydis, J. An introduction to deep learning for the physical layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575. [CrossRef]
25. Liang, F.; Shen, C.; Wu, F. An iterative BP-CNN architecture for channel decoding. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 144–159. [CrossRef]
26. Kim, M.; Kim, N.I.; Lee, W.; Cho, D.H. Deep learning-aided SCMA. *IEEE Commun. Lett.* **2018**, *22*, 720–723. [CrossRef]
27. He, H.; Wen, C.K.; Jin, S.; Li, G.Y. Model-driven deep learning for MIMO detection. *IEEE Trans. Signal Process.* **2020**, *68*, 1702–1715. [CrossRef]
28. Kang, J.M.; Chun, C.J.; Kim, I.M. Deep-learning-based channel estimation for wireless energy transfer. *IEEE Commun. Lett.* **2018**, *22*, 2310–2313. [CrossRef]
29. Neumann, D.; Wiese, T.; Utschick, W. Learning the MMSE channel estimator. *IEEE Trans. Signal Process.* **2018**, *66*, 2905–2917. [CrossRef]
30. Huang, H.; Yang, J.; Huang, H.; Song, Y.; Gui, G. Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8549–8560. [CrossRef]
31. He, H.; Wen, C.K.; Jin, S.; Li, G.Y. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 852–855. [CrossRef]
32. Chun, C.J.; Kang, J.M.; Kim, I.M. Deep learning-based channel estimation for massive MIMO systems. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1228–1231. [CrossRef]
33. Wang, T.; Wen, C.K.; Jin, S.; Li, G.Y. Deep learning-based CSI feedback approach for time-varying massive MIMO channels. *IEEE Wirel. Commun. Lett.* **2018**, *8*, 416–419. [CrossRef]
34. Borgerding, M.; Schniter, P.; Rangan, S. AMP-inspired deep networks for sparse linear inverse problems. *IEEE Trans. Signal Process.* **2017**, *65*, 4293–4308. [CrossRef]
35. Tan, W.; Yang, X.; Ma, S. Spectral Efficiency of Massive MIMO With LMMSE Receivers Under Finite Dimensional Channels. *IEEE Wirel. Commun. Lett.* **2021**, *11*, 18–22. [CrossRef]
36. Wei, X.; Hu, C.; Dai, L. Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems. *IEEE Trans. Commun.* **2020**, *69*, 182–193. [CrossRef]
37. Gao, X.; Dai, L.; Han, S.; Chih-Lin, I.; Heath, R.W. Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 998–1009. [CrossRef]