

Deep Learning for Compressed Sensing Based Channel Estimation in Millimeter Wave Massive MIMO

Wenyan Ma*, Chenhao Qi*, Zaichen Zhang*[†] and Julian Cheng[‡]

*School of Information Science and Engineering, Southeast University, Nanjing, China

[†]Purple Mountain Laboratories, Nanjing 211111, China

[‡]School of Engineering, University of British Columbia, Canada

Email: {220170772,qch,zczhang}@seu.edu.cn, julian.cheng@ubc.ca

Abstract—Channel estimation is considered for multi-user millimeter wave (mmWave) massive multi-input multi-output system. A deep learning compressed sensing (DLCS) channel estimation scheme is proposed, and it consists of beamspace channel amplitude estimation and channel reconstruction. The neural network (NN) for the DLCS scheme is trained offline using simulated environments according to the mmWave channel model. Then the correlation between the received signal vectors and the measurement matrix is input into the trained NN to predict the beamspace channel amplitude. Afterwards, the channel is reconstructed based on the obtained indices of dominant beamspace channel entries. Simulation results demonstrate that the proposed DLCS channel estimation scheme outperforms the existing schemes including the orthogonal matching pursuit and the distributed grid matching pursuit in terms of the normalized mean-squared error and the spectral efficiency.

Index Terms—Channel estimation, mmWave communications, deep learning, massive MIMO.

I. INTRODUCTION

Due to the rich bandwidth resources of the millimeter wave (mmWave), mmWave communication has attracted broad attention and become an important technology in future wireless communication systems [1], [2]. The mmWave signal experiences high path loss considering high frequency. Fortunately, this challenge can be overcome by utilizing a massive multi-input multi-output (MIMO) antenna array to achieve directional beam alignment and data transmission. Since mmWave bands have short wavelengths, large antenna arrays can be packed into small form factors.

Due to the large antenna arrays of mmWave communications, channel estimation requires large time slots overhead. Note that the mmWave channels have sparsity feature in the beamspace domain, where the beamspace channels can be formed by either the lens antenna arrays or phase shifter network [3], [4]. Several channel estimation schemes have been proposed to explore the beamspace channel sparsity. For examples, a distributed grid matching pursuit (DGMP) channel estimation scheme was proposed in [3], where the channel support was obtained and updated iteratively; an orthogonal matching pursuit (OMP) channel estimation scheme was proposed in [5]; a simultaneous weighted orthogonal matching

pursuit (SWOMP) channel estimation scheme was proposed in [4], where the frequency-selective mmWave channels were considered based on the OMP method. However, these compressed sensing (CS) channel estimation schemes estimate the dominant beamspace channel entries in a sequential and greedy manner, which cannot guarantee the global optimality [6].

In this paper, we propose a deep learning compressed sensing (DLCS) channel estimation scheme for the multi-user mmWave massive MIMO systems. The DLCS scheme consists of beamspace channel amplitude estimation and channel reconstruction. In the offline training stage, we train the neural network (NN) using the simulated environment based on the mmWave channel model, and then in the online deployment stage, the correlation between the received signal vectors and the measurement matrix is input into the trained NN to predict the beamspace channel amplitude. Afterwards, the indices of dominant entries of beamspace channel are obtained, based on which the channel can be reconstructed. Note that unlike the existing work that estimates the dominant beamspace channel entries sequentially [3]–[5], we estimate dominant entries simultaneously, which will be shown to have better channel estimation performance.

We use the following notations in our paper. Symbols for vectors (lower case) and matrices (upper case) are in boldface. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the transpose, conjugate, conjugate transpose (Hermitian), and inverse, respectively. We use \mathbf{I}_K to represent identity matrix of order K . The set of $P \times Q$ complex-valued matrices and real-valued matrices are denoted by $\mathbb{C}^{P \times Q}$ and $\mathbb{R}^{P \times Q}$, respectively. We use $\mathbb{E}\{\cdot\}$ to represent expectation. The l_2 -norm of a vector and Frobenius norm of a matrix are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. We use $\mathbf{a}[p]$ to denote the p th entry of \mathbf{a} . Complex Gaussian distribution is denoted by \mathcal{CN} . We use $|\cdot|$ to denote the absolute value.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We first introduce the system model of multi-user mmWave massive MIMO. Then the channel estimation problem is

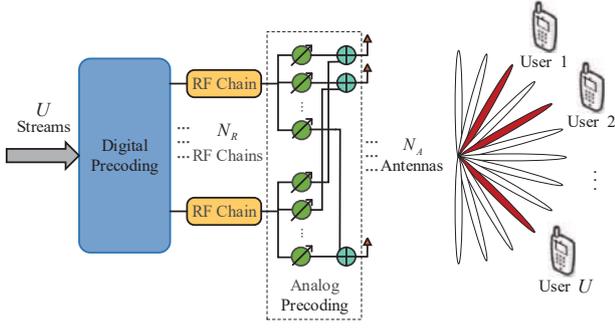


Fig. 1. Block diagram of downlink transmission in the multi-user mmWave massive MIMO system.

formulated as a CS problem to estimate the sparse channel in the beamspace.

A. System Model

We consider a downlink multi-user mmWave massive MIMO communication system that comprises a base station (BS) and U users with single antenna, as shown in Fig. 1. The BS is equipped with a uniform linear array (ULA) [1]. Let N_A and N_R denote the numbers of antennas and RF chains at the BS, respectively. Hybrid precoding is typically adopted, where the number of antennas is much larger than that of RF chains, i.e., $N_A \gg N_R$ [2]. We consider the orthogonal multiple access, where the number of active users simultaneously connected with the BS is no larger than that of the RF chains, i.e., $U \leq N_R$ [7]. If $U < N_R$, the BS will only turn on U RF chains to serve the U users simultaneously and turn off $(N_R - U)$ RF chains, which will save the power consumed at the BS.

For downlink transmission, the BS performs hybrid precoding, which consists of baseband digital precoding and RF analog precoding [8]. The received signal of all U users, denoted by $\mathbf{y}^{\text{dl}} \in \mathbb{C}^U$ can be represented as

$$\mathbf{y}^{\text{dl}} = \mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{s} + \mathbf{n} \quad (1)$$

where $\mathbf{F}_R \in \mathbb{C}^{N_A \times U}$ and $\mathbf{F}_B \in \mathbb{C}^{U \times U}$ denote the analog precoder and digital precoder, respectively. To normalize the power of the hybrid precoder, we set $\|\mathbf{F}_R\mathbf{F}_B\|_F^2 = U$. We denote the signal vector by $\mathbf{s} \in \mathbb{C}^U$ satisfying $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_U$ and additive white Gaussian noise (AWGN) vector by $\mathbf{n} \in \mathbb{C}^U$ satisfying $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_U)$. The channel matrix for the BS and all users is denoted by

$$\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_U]^T \in \mathbb{C}^{U \times N_A}. \quad (2)$$

Assuming the widely adopted Saleh-Valenzuela mmWave channel model [2], the channel vector $\mathbf{h}_u \in \mathbb{C}^{N_A}$ for the BS and the u th user is represented as

$$\mathbf{h}_u = \sqrt{\frac{N_A}{L_u}} \sum_{i=1}^{L_u} \mathbf{h}_{u,i} = \sqrt{\frac{N_A}{L_u}} \sum_{i=1}^{L_u} g_{u,i} \boldsymbol{\alpha}(N_A, \theta_{u,i}) \quad (3)$$

where the channel vector, number of multi channel paths, and complex gain of the i th path are denoted by $\mathbf{h}_{u,i}$, L_u , and

$g_{u,i}$, respectively. Typically \mathbf{h}_u consists of one line-of-sight (LOS) path (the 1st channel path), and $L_u - 1$ non-line-of-sight (NLOS) paths (the i th channel path for $2 \leq i \leq L_u$). The steering vector $\boldsymbol{\alpha}(N, \theta)$ can be expressed as

$$\boldsymbol{\alpha}(N, \theta) = \frac{1}{\sqrt{N}} [1, e^{j\pi\theta}, \dots, e^{j\pi\theta(N-1)}]^T. \quad (4)$$

Denote the angle of arrival (AoA) for the i th path of the u th user by $\vartheta_{u,i}$, which is uniformly distributed over $[-\pi, \pi]$ [9]. Then we have $\theta_{u,i} \triangleq \sin \vartheta_{u,i}$ if the distance between adjacent two antennas at the BS is half-wave length [3].

B. Problem Formulation

In order to design \mathbf{F}_B and \mathbf{F}_R for downlink data transmission, \mathbf{H} should be well estimated. Based on channel reciprocity, the estimate of downlink channel can be obtained by employing uplink channel estimation to estimate \mathbf{H} . Note that the proposed DLCS channel estimation scheme can be used for the downlink channel estimation as well. In this work, we focus on the uplink channel estimation, since the BS usually having more computing power than each user is better to perform the NN training and prediction than the users. For uplink channel estimation, mutually orthogonal pilot sequences are transmitted by all users to distinguish different signals from different users for K times. Each pilot sequence is with length of U . Denote the pilot matrix consisted of the U mutually orthogonal pilot sequences from U users by $\mathbf{P} \in \mathbb{C}^{U \times U}$. For the uplink pilot transmission, we use K different analog precoding matrices and digital precoding matrices, denoted by $\mathbf{F}_R^k \in \mathbb{C}^{N_A \times N_R}$ and $\mathbf{F}_B^k \in \mathbb{C}^{N_R \times N_R}$, respectively, for $k = 1, 2, \dots, K$. The pilot sequences received at the BS for the k th sending are given by

$$\mathbf{Y}_k^{\text{ul}} = (\mathbf{F}_R^k \mathbf{F}_B^k)^T \mathbf{H}^T \mathbf{P} + (\mathbf{F}_R^k \mathbf{F}_B^k)^T \mathbf{N}_k \quad (5)$$

where the AWGN matrix for the k th transmission is denoted by \mathbf{N}_k . Each entry of \mathbf{N}_k obeys $\mathcal{CN}(0, \sigma^2)$. Based on the orthogonality of U mutually orthogonal pilot sequences, i.e., $\mathbf{P}\mathbf{P}^H = \mathbf{I}_U$, we multiply \mathbf{Y}_k^{ul} by \mathbf{P}^H and obtain

$$\mathbf{R}_k \triangleq \mathbf{Y}_k^{\text{ul}} \mathbf{P}^H = (\mathbf{F}^k)^T \mathbf{H}^T + \widetilde{\mathbf{N}}_k \quad (6)$$

where

$$\begin{aligned} \mathbf{F}^k &\triangleq \mathbf{F}_R^k \mathbf{F}_B^k \in \mathbb{C}^{N_A \times N_R}, \\ \widetilde{\mathbf{N}}_k &\triangleq (\mathbf{F}_R^k \mathbf{F}_B^k)^T \mathbf{N}_k \mathbf{P}^H \in \mathbb{C}^{N_R \times U}. \end{aligned} \quad (7)$$

After each user repeatedly transmits orthogonal pilot sequences for K times, \mathbf{R}_k for $k = 1, 2, \dots, K$ can be stacked as

$$\mathbf{R} = [\mathbf{R}_1^T, \dots, \mathbf{R}_K^T]^T = \mathbf{F}^T \mathbf{H}^T + \widetilde{\mathbf{N}} \quad (8)$$

where

$$\begin{aligned} \mathbf{F} &\triangleq [\mathbf{F}^1, \dots, \mathbf{F}^K] \in \mathbb{C}^{N_A \times N_R K}, \\ \widetilde{\mathbf{N}} &\triangleq [\widetilde{\mathbf{N}}_1^T, \dots, \widetilde{\mathbf{N}}_K^T]^T \in \mathbb{C}^{N_R K \times U}. \end{aligned} \quad (9)$$

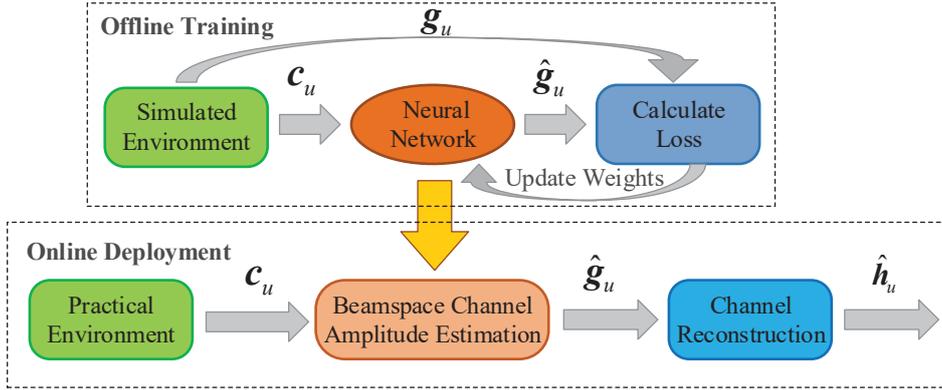


Fig. 2. Block diagram of the DLCS channel estimation scheme: offline training and online deployment.

Denote the u th column of \mathbf{R} by \mathbf{r}_u for $u = 1, 2, \dots, U$. Then \mathbf{r}_u can be represented as

$$\mathbf{r}_u = \mathbf{F}^T \mathbf{h}_u + \tilde{\mathbf{n}}_u \quad (10)$$

where $\tilde{\mathbf{n}}_u$ is the u th column of $\tilde{\mathbf{N}}$.

Note that the mmWave channels have sparsity feature in the beamspace domain [3], [4], [10]. We define

$$\mathbf{h}_u^b = \mathbf{A} \mathbf{h}_u \quad (11)$$

as a beamspace channel vector where $\mathbf{A} \in \mathbb{C}^{G \times N_A}$ is the dictionary matrix consisted of G column vectors $\boldsymbol{\alpha}(N_A, \phi_t)$, with $\phi_t \triangleq -1 + 2(t-1)/G$ representing the t th point of the angle grid. Note that the range of AoAs is quantified into G grids for $t = 1, 2, \dots, G$. Based on the fact that $\mathbf{A}^H \mathbf{A} = G \mathbf{I}_{N_A}/N_A$, equation (11) can be further rewritten as

$$\mathbf{r}_u = \frac{N_A}{G} \mathbf{F}^T \mathbf{A}^H \mathbf{h}_u^b + \tilde{\mathbf{n}}_u. \quad (12)$$

Due to the sparse property of \mathbf{h}_u^b , equation (12) is essentially a sparse recovery problem, which can be tackled by CS techniques. Note that the sparsity of \mathbf{h}_u^b can be impaired by channel power leakage caused by the limited beamspace resolution of \mathbf{A} [11], which indicates that \mathbf{h}_u^b is not perfectly sparse and many entries of \mathbf{h}_u^b are small but nonzero. CS channel estimation schemes such as OMP and DGMP estimate the dominant beamspace channel entries in a sequential and greedy manner. However, they cannot guarantee the global optimality. Therefore, in the following we will propose a DLCS channel estimation scheme to estimate dominant beamspace channel entries simultaneously.

III. DLCS CHANNEL ESTIMATION

The proposed DLCS channel estimation scheme consists of beamspace channel amplitude estimation and channel reconstruction. The main idea of the DLCS scheme is to first estimate the beamspace channel amplitude using an offline-trained NN, and then sort the estimated beamspace channel amplitude in descending order to select the indices of dominant entries, and finally reconstruct the channel according to the selected indices. The block diagram of the DLCS scheme

Algorithm 1 DLCS Channel Estimation

- 1: *Input:* Φ, \mathbf{r}_u, J .
 - 2: Initialization: $\hat{\mathbf{h}}_u^b \leftarrow \mathbf{0}^G$.
 - 3: (*Beamspace Channel Amplitude Estimation*)
 - 4: Obtain \mathbf{c}_u via (14).
 - 5: Input \mathbf{c}_u to the offline-trained NN to get $\hat{\mathbf{g}}_u$.
 - 6: (*Channel Reconstruction*)
 - 7: Obtain Γ based on J dominant entries of $\hat{\mathbf{g}}_u$.
 - 8: Compute $\hat{\mathbf{h}}_u^b[\Gamma]$ via (17).
 - 9: Obtain $\hat{\mathbf{h}}_u$ according to (18).
 - 10: *Output:* $\hat{\mathbf{h}}_u$.
-

is illustrated in Fig. 2. The detailed steps of the DLCS scheme is summarized in Algorithm 1.

A. Beamspace Channel Amplitude Estimation

We define

$$\Phi \triangleq \frac{N_A}{G} \mathbf{F}^T \mathbf{A}^H \in \mathbb{C}^{N_R K \times G} \quad (13)$$

as the measurement matrix in (12). As shown in Algorithm 1, we input Φ and \mathbf{r}_u to output the estimate of \mathbf{h}_u , denoted as $\hat{\mathbf{h}}_u$, for $u = 1, 2, \dots, U$. The correlation vector between Φ and \mathbf{r}_u , denoted as $\mathbf{c}_u \in \mathbb{C}^G$, can be expressed as

$$\mathbf{c}_u = \Phi^H \mathbf{r}_u. \quad (14)$$

The CS channel estimation schemes sequentially select the atoms, i.e., column vectors of Φ , which yield the greatest correlation with \mathbf{r}_u . However, such greedy algorithms cannot guarantee the global optimality, which motivates us to use the NN to estimate the atoms simultaneously instead of sequentially.

As shown in Fig. 2, the beamspace channel amplitude estimation has two stages, including the offline training of the NN and online deployment of it. The NN is first trained offline and then used as the kernel of the beamspace channel amplitude

estimation. The input of the NN is \mathbf{c}_u . The amplitude of \mathbf{h}_u^b can be denoted by

$$\mathbf{g}_u \triangleq \left[\left| \mathbf{h}_u^b[1] \right|, \left| \mathbf{h}_u^b[2] \right|, \dots, \left| \mathbf{h}_u^b[G] \right| \right]^T \in \mathbb{R}^G. \quad (15)$$

The output of the NN is denoted by $\hat{\mathbf{g}}_u$ and is expected to be \mathbf{g}_u .

As illustrated in Fig. 3, the adopted NN in this work consists of three hidden layers and a fully connected (FC) layer. Since the NN can only deal with the real number, the input of the NN is a real-valued vector with length of $2G$ composed by the imaginary and real parts of \mathbf{c}_u . Each hidden layer includes an FC layer and a batch normalization (BN) layer. The numbers of neurons in these three hidden layers are set as 1,024, 512, and 256. The activation function adopted in the FC layer is the ReLU function, which can be represented as $f_{\text{Re}}(x) = \max(0, x)$.

During the offline training of the NN, we generate the dataset of \mathbf{c}_u and \mathbf{g}_u based on the simulated mmWave channel environment. With the beamspace channel amplitude in (15) and the correlation of the received signals and the measurement matrix in (14), the training data of \mathbf{c}_u and \mathbf{g}_u can be obtained. In fact, the process to obtain \mathbf{c}_u and \mathbf{g}_u involves the following four steps. **i)** We randomly generate a channel vector based on the mmWave channel model in (3). **ii)** We obtain \mathbf{g}_u based on (15). **iii)** We compute the received signal vector \mathbf{r}_u based on (10). **iv)** We figure out the correlation vector \mathbf{c}_u based on (14). The output of the NN is $\hat{\mathbf{g}}_u$.

The training of the NN aims to minimize the difference between $\hat{\mathbf{g}}_u$ and \mathbf{g}_u . The difference typically named as the loss in machine learning, can be calculated in several ways. In our work, we calculate the loss by measuring the mean square error as

$$f_{\text{Loss}}(\mathbf{g}_u, \hat{\mathbf{g}}_u) = \frac{1}{G} \sum_{n=1}^G (\mathbf{g}_u[n] - \hat{\mathbf{g}}_u[n])^2. \quad (16)$$

We adopt the adaptive moment estimation (Adam) optimizer to train the NN by TensorFlow [12]. The NN is trained for 1,000 epochs, where 50 mini-batches are utilized in each epoch. The learning rate is set to be a step function, which decreases with the increasing of training epochs. The learning rate is initialized with the value of 0.01 and decreases 5-fold every 400 epochs.

During the online deployment of the NN, we obtain the real measured \mathbf{r}_u from practical mmWave channel environments. We compute \mathbf{c}_u based on (14), which is then input to the offline-trained NN. The prediction of \mathbf{g}_u by the NN is $\hat{\mathbf{g}}_u$.

B. Channel Reconstruction

Note that the sparsity of \mathbf{h}_u^b can be impaired by channel power leakage caused by the limited beamspace resolution of \mathbf{A} [11], which indicates that \mathbf{h}_u^b is not perfectly sparse and many entries of \mathbf{h}_u^b are small but nonzero. Denote the number of dominant entries of \mathbf{g}_u by J , which is the beamspace channel sparse level. In the online deployment stage, we sort $\hat{\mathbf{g}}_u$ in descending order according to the absolute value of $\hat{\mathbf{g}}_u$.

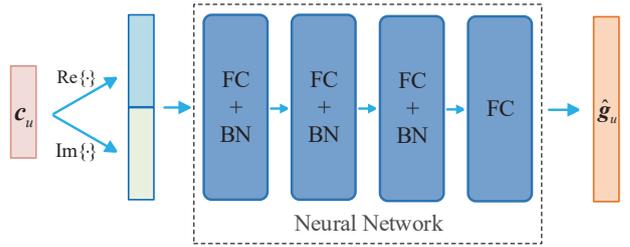


Fig. 3. Illustration of the NN.

Then we obtain the indices of the first J entries, which are the prediction of the indices of J dominant entries in \mathbf{g}_u .

We denote the prediction of these J indices by $\mathbf{\Gamma} \in \mathbb{R}^J$. We further let $\hat{\mathbf{h}}_u^b$ denote an estimate of \mathbf{h}_u^b . We initialize $\hat{\mathbf{h}}_u^b$ to be zero. Then the J dominant entries of $\hat{\mathbf{h}}_u^b$ can be computed via the least squares (LS) estimation as

$$\hat{\mathbf{h}}_u^b[\mathbf{\Gamma}] = (\mathbf{\Phi}_{\mathbf{\Gamma}}^H \mathbf{\Phi}_{\mathbf{\Gamma}})^{-1} \mathbf{\Phi}_{\mathbf{\Gamma}}^H \mathbf{r}_u \quad (17)$$

where $\mathbf{\Phi}_{\mathbf{\Gamma}}$ consists of J columns of $\mathbf{\Phi}$ and the column indices are denoted by $\mathbf{\Gamma}$. Then using the result $\mathbf{A}^H \mathbf{A} = G \mathbf{I}_{N_A} / N_A$, the estimated channel vector for the u th user can be obtained based on (11) as

$$\hat{\mathbf{h}}_u = \frac{N_A}{G} \mathbf{A}^H \hat{\mathbf{h}}_u^b. \quad (18)$$

It is shown that the proposed DLCS channel estimation scheme can avoid greedy manner that is commonly adopted by the existing channel estimation schemes based on CS, since the DLCS scheme estimates dominant entries simultaneously instead of sequentially.

IV. SIMULATION RESULTS

In the following we will present the performance evaluation for the proposed DLCS channel estimation scheme. Considering a multi-user mmWave massive MIMO communication system, the BS equipped with $N_R = 4$ RF chains and $N_A = 64$ antennas serves $U = 3$ users with single antenna. We set $G = 128$ according to [4], and we set the number of multi paths in mmWave channel as $L_u = 2$, where $g_{u,1} \sim \mathcal{CN}(0, 1)$ and $g_{u,2} \sim \mathcal{CN}(0, 0.5)$ [4], [9]. For the uplink pilot transmission, we set $\mathbf{F}_B^k = \mathbf{I}_{N_R}$. Each entry of \mathbf{F}_R^k is randomly drawn from the set $\{e^{j2\pi n/2^Q}, n = 1, 2, \dots, 2^Q\}$, where the quantization bit number of the phase shifters used at the BS is $Q = 4$ [4]. Since \mathbf{h}_u^b is not ideally sparse due to the power leakage, the beamspace channel sparse level should be larger than L_u , i.e., $J > 2$. We set $J = 6, 7$ in performance simulating. Note that the NN is trained to predict the beamspace channel amplitude, where the training process of the NN is independent of J . The proposed DLCS channel estimation scheme is compared with the existing OMP [5] and DGMP [3] channel estimation schemes.

As shown in Fig. 4, the channel estimation performance for the proposed DLCS scheme together with the existing schemes is compared in terms of SNR. The channel estimation

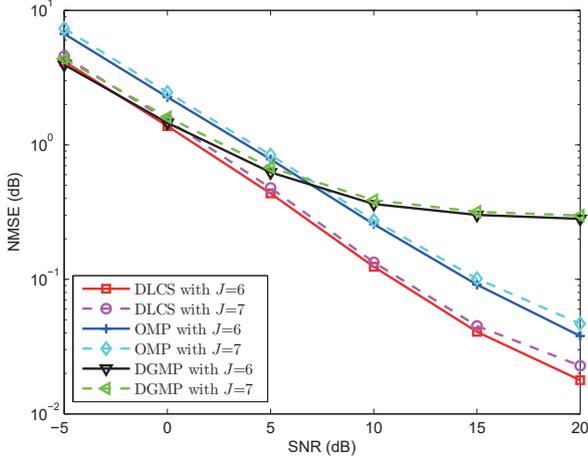


Fig. 4. Comparisons of channel estimation performance in terms of SNR for different schemes.

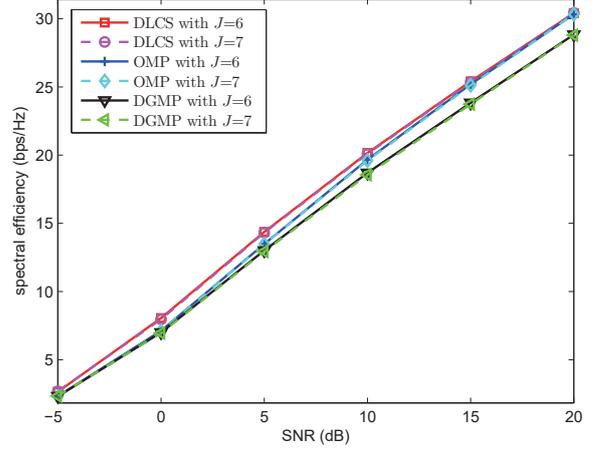


Fig. 5. Comparisons of spectral efficiency in terms of SNR for different schemes.

performance is measured by the normalized mean-squared error (NMSE), which can be denoted by

$$\text{NMSE} \triangleq \frac{\sum_{u=1}^U \|\hat{\mathbf{h}}_u - \mathbf{h}_u\|_2^2}{\sum_{u=1}^U \|\mathbf{h}_u\|_2^2}. \quad (19)$$

We use $K = 8$ time slots to transmit pilots for uplink channel estimation. To make a fair comparison, we fix the pilot training time slots to be 8 for the OMP and DGMP schemes, respectively. It is shown that the DLCS scheme has better channel estimation performance than existing schemes. When $\text{SNR} = 10$ dB, the DLCS scheme with $J = 6$ has 51.7% and 65.8% performance improvements over the OMP and DGMP schemes, respectively, while the DLCS scheme with $J = 7$ has 51.3% and 65.5% performance improvements over the OMP and DGMP schemes, respectively. We explain the reason for the performance gap as follows. The OMP scheme estimates the beamspace channel dominant entries sequentially, which cannot guarantee global optimality. The DGMP scheme only estimates the LOS path, while our proposed DLCS scheme can simultaneously estimate all the dominant beamspace channel entries.

Fig. 5 makes a comparison of the spectral efficiency for the proposed DLCS scheme together with the existing schemes in terms of SNR. Based on the estimated channel, there are various methods to design the hybrid precoding for mmWave downlink transmission. Similar to [4], in this work we wish to compare the upper bound of the downlink spectral efficiency, which can be simply measured by the fully-digital precoding. We denote the estimated channel matrix for the BS and all users by $\hat{\mathbf{H}} \triangleq [\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_U]^T \in \mathbb{C}^{U \times N_A}$. The zero-forcing (ZF) precoding matrix can be represented by $\mathbf{F}^{\text{dl}} \triangleq (\hat{\mathbf{H}}^* \hat{\mathbf{H}}^T)^{-1} \hat{\mathbf{H}}^*$. In order to meet the total power budget, the u th row of \mathbf{F}^{dl} , denoted by \mathbf{f}_u^{dl} , should be normalized, i.e., $\mathbf{f}_u^{\text{dl}} \leftarrow \mathbf{f}_u^{\text{dl}} / \|\mathbf{f}_u^{\text{dl}}\|_2$ such that $\|\mathbf{f}_u^{\text{dl}}\|_2 = 1$ for $u = 1, 2, \dots, U$. We further denote the effective channel

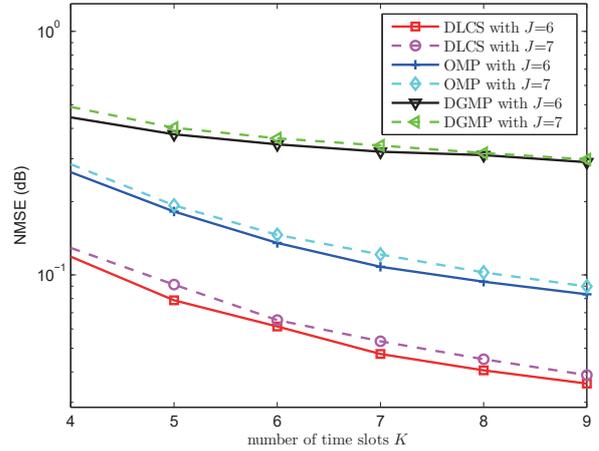


Fig. 6. Comparisons of channel estimation performance in terms of the number of time slots for channel training for different schemes.

matrix by $\mathbf{H}_{\text{eff}} \triangleq \mathbf{F}^{\text{dl}} \hat{\mathbf{H}}^T$. Then the spectral efficiency can be denoted by [4]

$$R \triangleq \sum_{u=1}^U \log_2 \left(1 + \frac{\text{SNR}}{U} \lambda_u(\mathbf{H}_{\text{eff}})^2 \right) \quad (20)$$

where $\lambda_u(\mathbf{H}_{\text{eff}})$ denotes the u th eigenvalue of \mathbf{H}_{eff} for $u = 1, 2, \dots, U$.

It is seen from Fig. 5 that the DLCS scheme has better channel estimation performance than existing schemes. When $\text{SNR} = 10$ dB, the DLCS scheme with $J = 6$ has 2.5% and 7.8% performance improvements over the OMP and DGMP schemes, respectively, while the DLCS scheme with $J = 7$ has 2.6% and 8.3% performance improvements over the OMP and DGMP schemes, respectively.

In Fig. 6, the channel estimation performance for the DLCS, OMP, and DGMP schemes is compared in terms of the

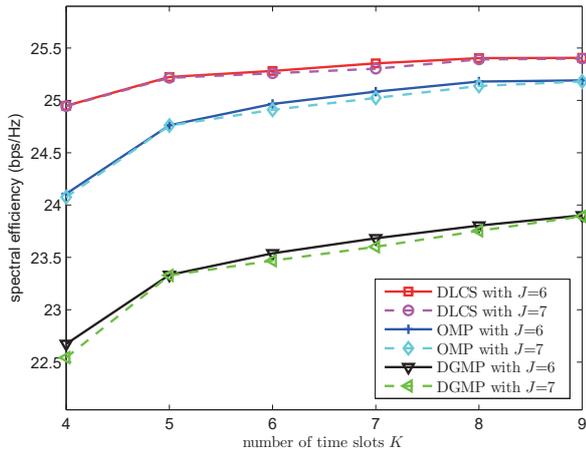


Fig. 7. Comparisons of spectral efficiency in terms of the number of time slots for channel training for different schemes.

number of time slots for channel training. We use the same number of pilot training time slots for the DLCS, OMP, and DGMP schemes. SNR is fixed as 15 dB. From Fig. 6, it is shown that the DLCS scheme has the best channel estimation performance. When fixing the number of pilot training time slots to be $K = 7$, the DLCS scheme with $J = 6$ has 56.1% and 85.2% performance improvements over the OMP and DGMP schemes, respectively, while the DLCS scheme with $J = 7$ has 56.0% and 84.3% performance improvements over the OMP and DGMP schemes, respectively.

Fig. 7 compares the spectral efficiency for different schemes in the aspect of the number of time slots for channel training. The system parameters for performance simulation are set to be the same as those for Fig. 6. It is shown that the DLCS scheme can have better channel estimation performance than the OMP and DGMP schemes. When the number of channel training time slots is more than 8, the spectral efficiency of the DLCS scheme remains constant, indicating that $K = 8$ is sufficient to obtain the full channel state information.

V. CONCLUSIONS

We proposed a DLCS channel estimation scheme for the multi-user mmWave massive MIMO communication systems. The proposed scheme was compared with the existing schemes in the aspect of NMSE and spectral efficiency. Simulation results showed that the proposed scheme has better channel

estimation performance than existing schemes. As the future work, it is worth developing the hybrid precoding design for multi-user mmWave massive MIMO transmission considering the limited quantization bits of the phase shifters.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61871119, by the Natural Science Foundation of Jiangsu Province under Grant BK20161428 and by the Fundamental Research Funds for the Central Universities.

REFERENCES

- [1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [2] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 168–178, Jan. 2015.
- [3] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, June 2016.
- [4] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [5] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. Gonzalez-Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *Proc. IEEE ICASSP*, New Orleans, USA, Mar. 2017, pp. 6493–6497.
- [6] W. Ma and C. Qi, "Beamspace channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4839–4853, Sep. 2018.
- [7] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.
- [8] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multi-user millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.
- [9] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [10] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.
- [11] J. Brady, N. Behdad, and A. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.
- [12] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, Dec. 2018.