DEEP CNN FOR WIDEBAND MMWAVE MASSIVE MIMO CHANNEL ESTIMATION USING FREQUENCY CORRELATION

Peihao Dong[⋆]*, Hua Zhang*[⋆]*, Geoffrey Ye Li*[†]*, Navid NaderiAlizadeh*[‡]*, and Ivan Simões Gaspar*[‡]

[⋆]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
[†]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[‡]Intel Corporation, Santa Clara, CA, USA

ABSTRACT

For millimeter wave (mmWave) systems with large-scale arrays, hybrid processing structure is usually used at both transmitters and receivers to reduce the complexity and cost, which poses a very challenging issue in channel estimation, especially at the low transmit signal-to-noise ratio regime. In this paper, deep convolutional neural network (CNN) is employed to perform wideband channel estimation for mmWave massive multiple-input multiple-output (MIMO) systems. In addition to exploiting spatial correlation, our joint channel estimation approach also exploits the frequency correlation, where the tentatively estimated channel matrices at multiple adjacent subcarriers are input into the CNN simultaneously. The complexity analysis and numerical results show that the proposed CNN based joint channel estimation outperforms the non-ideal minimum mean-squared error (MMSE) estimator with reduced complexity and achieves the performance close to the ideal MMSE estimator. It is also quite robust to different propagation scenarios.

Index Terms**—** mmWave massive MIMO, CNN, channel estimation, frequency correlation

1. INTRODUCTION

Millimeter wave (mmWave) communications can meet the high data rate demand due to its broad frequency band. Its shortcoming of high propagation loss can be well compensated by using massive multiple-input multiple-output (MIMO) [1], [2]. To reduce complexity and cost, phase shifter based two-stage structure is widely used at both the transmitter and the receiver to connect a large number of antennas with much fewer radio frequency (RF) chains [3], [4].

For mmWave massive MIMO systems with two-stage transceiver structure, channel estimation is a challenging problem. In [5], a hierarchical multi-resolution codebook has been designed, based on which an adaptive channel estimation algorithm has been developed by exploiting the channel sparsity. In [6], the structured sparsity in angle domain has been utilized to estimate the wideband channel for multi-user mmWave massive MIMO uplink. In [7], a channel estimation approach has been developed for mmWave massive MIMO orthogonal frequency division multiplexing (OFDM) systems with hybrid architecture and low-precision analog-to-digital converters (ADCs) under different channel models.

Compared to the conventional methods, machine learning (ML) is more powerful to uncover the inherent characteristics inside data/signals collected in an end-to-end manner and thus can achieve better performance when addressing various problems in wireless communications [8]. In [9], deep learning (DL) has been successfully used in joint channel estimation and signal detection in OFDM systems with interference and non-linear distortion. In [10], an iterative channel estimation has been proposed for the 3D lens mmWave massive MIMO systems, where denoising neural network (NN) is used in each iteration to update the estimated channel. To reduce the channel state information (CSI) feedback overhead of the frequency duplex division (FDD) massive MIMO system, deep autoencoder has been employed in [11] to compress the channel into a low dimensional codeword with the high recovery accuracy. Exploiting time correlation of the channel, long short-term memory (LSTM) based deep NN (DNN) has been introduced in [12] to develop a more efficient channel compression method for the CSI feedback. In [13], supervised learning algorithms have been used to acquire the downlink CSI for FDD massive MIMO systems with reduced overheads for pilot and CSI feedback.

In this paper, we use a deep convolutional NN (CNN) to address channel estimation for mmWave massive MIMO-OFDM systems so that both spatial and frequency correlation can be utilized. In the proposed joint channel estimation approach, the tentatively estimated channel matrices of adjacent subcarriers are input into the CNN simultaneously. Except the spatial correlation hidden in each tentatively estimated channel matrix, the frequency correlation among adjacent subcarriers provides additional information to improve the channel estimation accuracy. The complexity analysis and numerical

The work of P. Dong and H. Zhang was supported by National Natural Science Foundation of China under Grants 61571118 and 61601115. The work of P. Dong was also supported by the China Scholarship Council (CSC) under Grants 201706090059. The work of G. Y. Li was supported in part by a research gift from Intel Corporation and in part by the National Science Foundation under Grants 1731017 and 1815637.

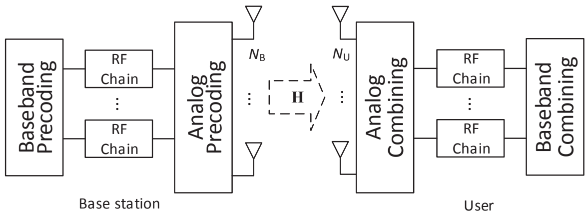

Fig. 1. System model of a mmWave massive MIMO-OFDM downlink.

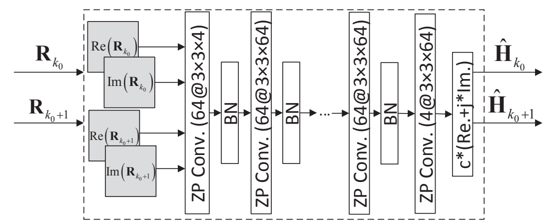

Fig. 2. Proposed CNN for joint channel estimation.

results show that the proposed CNN based joint channel estimation outperforms the existing method with reduced complexity. Furthermore, the proposed approach exhibits good robustness when facing new propagation scenarios without any knowledge of channel statistics.

Notations: In this paper, we use upper and lower case boldface letters to denote matrices and vectors, respectively. $\|\cdot\|_F$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, and $\mathbb{E}\{\cdot\}$ represent the Frobenius norm, transpose, conjugate transpose, pseudo inverse, and expectation, respectively. $\mathcal{CN}(\mu, \sigma^2)$ represents circular symmetric complex Gaussian distribution with mean μ and variance σ^2. $\delta(\cdot)$ denotes the delta function. $\mathbf{X} \otimes \mathbf{Y}$ denotes the Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$. $\mathrm{vec}(\mathbf{X})$ returns a column vector $\mathbf{x}$ whose elements are taken columnwise from $\mathbf{X}$ and $\mathrm{vec}^{-1}(\mathbf{x})$ is the inverse process.

2. SYSTEM MODEL

As shown in Fig. 1, we consider a mmWave massive MIMO-OFDM system, where the base station (BS) with N_{B} antennas and $N_{\mathrm{B}}^{\mathrm{RF}}$ RF chains transmits signals to a single user with N_{U} antennas and $N_{\mathrm{U}}^{\mathrm{RF}}$ RF chains. Phase shifters are employed to connect a large number of antennas with a much fewer number of RF chains at both the BS and the user sides. We therefore assume $N_{\mathrm{B}} \gg N_{\mathrm{B}}^{\mathrm{RF}}$ and $N_{\mathrm{U}} \gg N_{\mathrm{U}}^{\mathrm{RF}}$.

According to [6], the $N_{\mathrm{U}} \times N_{\mathrm{B}}$ channel matrix between the BS and the user in the delay domain is given by

$$\mathbf{H}(\tau) = \sqrt{\frac{N_{\mathrm{B}} N_{\mathrm{U}}}{L}} \sum_{l=1}^{L} \alpha_l \delta(\tau - \tau_l) \mathbf{a}_{\mathrm{U}}(\varphi_l) \mathbf{a}_{\mathrm{B}}^H(\phi_l), \quad (1)$$

where L is the number of paths and $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$ is the propagation gain of the lth path with σ_α^2 denoting the average power gain. For the lth path, φ_l and $\phi_l \in [0, 2\pi]$ are the azimuth angles of arrival or departure (AoA/AoD) at the user and the BS, respectively. For uniform linear array (ULA), the corresponding response vectors at the user and the BS can be expressed as

$$\mathbf{a}_{\mathrm{U}}(\varphi_l) = \frac{1}{\sqrt{N_{\mathrm{U}}}} \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\varphi_l)}, \ldots, e^{-j2\pi \frac{d}{\lambda}(N_{\mathrm{U}}-1)\sin(\varphi_l)}\right]^T, \quad (2)$$

$$\mathbf{a}_{\mathrm{B}}(\phi_l) = \frac{1}{\sqrt{N_{\mathrm{B}}}} \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\phi_l)}, \ldots, e^{-j2\pi \frac{d}{\lambda}(N_{\mathrm{B}}-1)\sin(\phi_l)}\right]^T, \quad (3)$$

where d and λ denote the distance between the adjacent antennas and carrier wavelength, respectively.

According to the channel model in (1), the channel frequency response at the kth subcarrier in OFDM is given by

$$\mathbf{H}_k = \sqrt{\frac{N_{\mathrm{B}} N_{\mathrm{U}}}{L}} \sum_{l=1}^{L} \alpha_l e^{-j2\pi\tau_l f_s \frac{k}{K}} \mathbf{a}_{\mathrm{U}}(\varphi_l) \mathbf{a}_{\mathrm{B}}^H(\phi_l), \quad (4)$$

where f_s denotes the sampling rate and K is the number of OFDM subcarriers.

To estimate $\mathbf{H}_k$, the BS transmits pilot signal $x_{k,u}$ on the beamforming vector $\mathbf{f}_{k,u} \in \mathbb{C}^{N_{\mathrm{B}} \times 1}$, $u = 1, \ldots, M_{\mathrm{B}}$, during M_{B} successive instants and the user employs M_{U} combining vectors $\mathbf{w}_{k,v} \in \mathbb{C}^{N_{\mathrm{U}} \times 1}$, $v = 1, \ldots, M_{\mathrm{U}}$, to process each beamforming vector. The pilot signal matrix associated with the kth subcarrier at the baseband of the user is written as

$$\mathbf{Y}_k = \mathbf{W}_k^H \mathbf{H}_k \mathbf{F}_k \mathbf{X}_k + \tilde{\mathbf{N}}_k, \quad (5)$$

where $\mathbf{W}_k = [\mathbf{w}_{k,1}, \ldots, \mathbf{w}_{k,M_{\mathrm{U}}}]$ and $\mathbf{F}_k = [\mathbf{f}_{k,1}, \ldots, \mathbf{f}_{k,M_{\mathrm{B}}}]$ are combining matrix and beamforming matrix, respectively, $\mathbf{X}_k$ is an $M_{\mathrm{B}} \times M_{\mathrm{B}}$ diagonal matrix with its uth diagonal element being $x_{k,u}$, $\tilde{\mathbf{N}}_k = \mathbf{W}_k^H \mathbf{N}_k$ denotes the effective noise after combining at the user, and $\mathbf{N}_k$ is additive white Gaussian noise (AWGN) with $\mathcal{CN}(0, 1)$ elements before combining.

3. CNN-BASED CHANNEL ESTIMATION

In this section, we first develop the CNN based channel estimation approach, followed by the complexity analysis for the online estimation.

3.1. Algorithm Description

1) Signal Preprocessing: Without loss of generality, we assume the worst case that $\mathbf{W}_k = \mathbf{W}$, $\mathbf{F}_k = \mathbf{F}$, and $\mathbf{X}_k = \sqrt{P}\mathbf{I}$ for all subcarriers with pilots. The pilot signal matrix, $\mathbf{Y}_k$, is vectorized as

$$\begin{aligned}\bar{\mathbf{y}}_k &= \mathrm{vec}(\mathbf{Y}_k) \overset{(a)}{=} \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H)\mathrm{vec}(\mathbf{H}_k) + \mathrm{vec}(\tilde{\mathbf{N}}_k) \\ &= \mathbf{Q}\bar{\mathbf{h}}_k + \bar{\mathbf{n}}_k,\end{aligned} \quad (6)$$

where $\mathbf{Q} = \sqrt{P}(\mathbf{F}^T \otimes \mathbf{W}^H)$, $\bar{\mathbf{h}}_k = \mathrm{vec}(\mathbf{H}_k)$, and $\bar{\mathbf{n}}_k = \mathrm{vec}(\tilde{\mathbf{N}}_k)$. In (6), step (a) is based on $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$.

The vectorized pilot signal, $\bar{\mathbf{y}}_k$, is further processed to obtain the tentatively estimated channel matrix at subcarrier k by

$$\mathbf{R}_k = \mathrm{vec}^{-1}(\bar{\mathbf{r}}_k) = \mathrm{vec}^{-1}(\mathbf{Q}^\dagger \bar{\mathbf{y}}_k). \quad (7)$$

The tentatively estimated channel matrices at S successive subcarriers, $\mathbf{R}_{k_0}, \mathbf{R}_{k_0+1}, \ldots, \mathbf{R}_{k_0+S-1}$, within the same coherence bandwidth will be input into the CNN simultaneously.

2) CNN Training: For the proposed CNN, the training data set consists of labeled pairs $(\underline{\mathbf{R}}_n, \underline{\mathbf{H}}_n)$, where $\underline{\mathbf{R}}_n \in \mathbb{C}^{N_\mathrm{U} \times N_\mathrm{B} \times S}$ and $\underline{\mathbf{H}}_n \in \mathbb{C}^{N_\mathrm{U} \times N_\mathrm{B} \times S}$ are three-dimensional matrices and the sth two-dimensional matrices of them are $\mathbf{R}^n_{k_0+s-1} \in \mathbb{C}^{N_\mathrm{U} \times N_\mathrm{B}}$ and $\frac{\mathbf{H}^n_{k_0+s-1}}{c} \in \mathbb{C}^{N_\mathrm{U} \times N_\mathrm{B}}$, respectively. In the above, $\mathbf{R}^n_{k_0+s-1}$ is the tentatively estimated channel matrix at subcarrier k_0+s-1 given by (7) and $\mathbf{H}^n_{k_0+s-1}$ is the corresponding true channel matrix. $c>0$ is a scaling constant to make the value range of the real and imaginary parts of all the target data match the activation function and to easily recover the channel. The tentatively estimated channel matrices of S subcarriers, $\underline{\mathbf{R}}_n$, are fed into the CNN to approximate the corresponding scaled channels $\underline{\mathbf{H}}_n$.

For the mmWave massive MIMO systems, we assume $N_\mathrm{B} = M_\mathrm{B} = 32$, $N_\mathrm{U} = M_\mathrm{U} = 16$ and $S = 2$ in Fig. 2. The CNN receives the processed complex pilots, $\mathbf{R}^n_{k_0} \in \mathbb{C}^{16\times 32}$ and $\mathbf{R}^n_{k_0+1} \in \mathbb{C}^{16\times 32}$ as the input and separates their real and imaginary parts so that 4 16×32 real-valued matrices are obtained. In the subsequent zero padding (ZP) convolutional layer, the 4 matrices are processed by 64 $3\times 3\times 4$ convolutional filters with the rectified linear unit (ReLU) activation function to generate 64 16×32 real-valued matrices, which are then processed by a batch normalization (BN) layer. For the next 8 ZP convolutional layers, each layer uses 64 $3\times 3\times 64$ convolutional filters and outputs 64 16×32 real-valued matrices with the ReLU activation function, followed by a BN layer. The output layer uses 4 $3\times 3\times 64$ convolutional filters to process the 64 16×32 real-valued matrices and obtains the estimated real and imaginary parts of the scaled channel matrices of the k_0th and (k_0+1)th subcarriers. Hyperbolic tangent activation function is used in the output layer to map the output into interval $[-1,1]$. After scaling up and combining the corresponding real and imaginary parts, the 16×32 complex-valued estimated channel matrices, $\hat{\mathbf{H}}^n_{k_0}$ and $\hat{\mathbf{H}}^n_{k_0+1}$, are obtained.

The objective of the offline training for the CNN is to minimize the MSE loss function

$$\mathrm{MSE}_\mathrm{Loss} = \frac{1}{N_\mathrm{tr} c^2} \sum_{n=1}^{N_\mathrm{tr}} \sum_{s=1}^{S} \left\| \mathbf{H}^n_{k_0+s-1} - \hat{\mathbf{H}}^n_{k_0+s-1} \right\|_F^2, \quad (8)$$

where $\hat{\mathbf{H}}^n_{k_0+s-1}$ is the sth output matrix of the CNN when the input is $\underline{\mathbf{R}}_n$.

3) CNN Testing: After the centralized training, the CNN will be deployed at the receiver to obtain the estimated channel matrices, $\hat{\mathbf{H}}_{k_0}, \hat{\mathbf{H}}_{k_0+1}, \ldots, \hat{\mathbf{H}}_{k_0+S-1}$, by jointly processing the pilot matrices, $\mathbf{R}_{k_0}, \mathbf{R}_{k_0+1}, \ldots, \mathbf{R}_{k_0+S-1}$.

If the actual channel model differs from that in the training stage, a straightforward solution is fine-tuning but it is hindered by the difficulty to collect the true channel. Fortunately, as shown by Fig. 4 in Section 4, the offline trained CNN is quite robust to the new channel statistics that are not observed before. This implies that further online fine-tuning might only provide marginal performance improvement and hence is not necessary.

Table 1. CNN Parameter Settings

l	$M_{1,l}$	$M_{2,l}$	F_l	N_{l-1}	N_l
1	16	32	3	4	64
$2\sim 9$	16	32	3	64	64
10	16	32	3	64	4

3.2. Complexity Analysis

In this subsection, we analyze the computational complexity of the proposed CNN based channel estimation approach in testing stage and compare with minimum mean-squared error (MMSE) approach. The required number of floating point operations (FLOPs) is used as the metric.

For the proposed approach, the FLOPs come from the received pilots processing in (7) and the CNN processing. Since $\mathbf{Q}^\dagger$ in (7) is independent of channel realizations and only needs to be calculated once, the computational complexity of the received pilots processing is mainly caused by matrix product and is $C_\mathrm{CNN,1} \sim \mathcal{O}(SN_\mathrm{B}^2N_\mathrm{U}^2)$. According to [14], the computational complexity of CNN processing is $C_\mathrm{CNN,2} \sim \mathcal{O}\left(\sum_{l=1}^{L_\mathrm{c}} M_{1,l}M_{2,l}F_l^2N_{l-1}N_l\right)$, where L_c is the number of convolutional layers, $M_{1,l}$ and $M_{2,l}$ denote the numbers of rows and columns of each feature map output by the lth layer, F_l is the side length of the filters used by the lth layer, N_{l-1} and N_l denote the numbers of input and output feature maps of the lth layer. These parameters are listed in Table 1. The computational complexity of the proposed approach is

$$C_\mathrm{CNN} \sim \mathcal{O}\left(SN_\mathrm{B}^2N_\mathrm{U}^2 + N_\mathrm{B}N_\mathrm{U}\sum_{l=1}^{L_\mathrm{c}} F_l^2 N_{l-1}N_l\right), \quad (9)$$

For the MMSE channel estimation, the channel covariance matrix needs to be calculated based on least square (LS) channel estimation once per channel realization, which incurs the computational complexity of $C_\mathrm{MMSE,1} \sim \mathcal{O}(SN_\mathrm{B}^2N_\mathrm{U}^2)$ if considering both spatial and frequency channel statistics. Then the channel estimation is refined by the covariance matrix and the corresponding computational complexity is $C_\mathrm{MMSE,2} \sim \mathcal{O}(S^3N_\mathrm{B}^3N_\mathrm{U}^3)$. Therefore, the overall computational complexity of MMSE is

$$C_\mathrm{MMSE} \sim \mathcal{O}(S^3N_\mathrm{B}^3N_\mathrm{U}^3). \quad (10)$$

It is hard to compare C_CNN with C_MMSE straightforwardly in general since the former depends on L_c, F_l, N_{l-1}, and N_l besides S, N_B and N_U. If $N_\mathrm{B}=32$, $N_\mathrm{U}=16$, $S=2$ and the other parameters for the CNN are listed in Table 1, the proposed CNN based approach is with computational complexity in the order of magnitude of 10^8 while MMSE needs a higher complexity in the order of magnitude of 10^9.

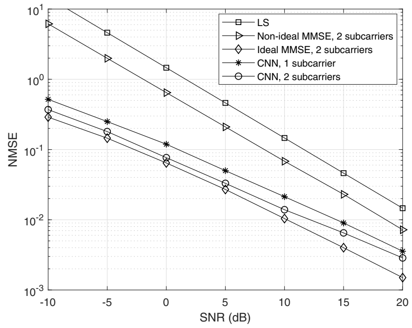

Fig. 3. NMSE versus transmit SNR for the proposed CNN based channel estimation and the existing methods.

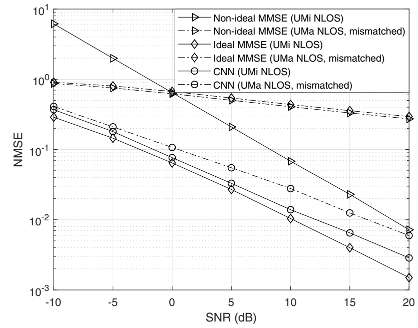

Fig. 4. Robustness for different scenarios.

4. NUMERICAL RESULTS

In this section, we present the simulation results of the proposed CNN based channel estimation approach and compare it with LS, non-ideal MMSE with estimated covariance matrix and ideal MMSE with ture covariance matrix. $\mathbf{F}$ and $\mathbf{W}$ in Section 3.1 are set as the first M_{B} columns of an $N_{\mathrm{B}} \times N_{\mathrm{B}}$ discrete Fourier transform (DFT) matrix and the first M_{U} columns of an $N_{\mathrm{U}} \times N_{\mathrm{U}}$ DFT matrix.

The channel data are generated according to the 3rd Generation Partnership Project (3GPP) TR 38.901 Release 15 channel model [15]. Specifically, we use the clustered delay line (CDL) models, the carrier frequency, $f_c = 28$ GHz, the sampling rate, $f_s = 100$ MHz, the number of antennas at the BS, $N_{\mathrm{B}} = 32$, the number of antennas at the user, $N_{\mathrm{U}} = 16$, the number of main paths, $L = 3$, the number of subcarriers, $K = 64$, the number of beamfroming vectors, $M_{\mathrm{B}} = 32$, and the number of combining vectors, $M_{\mathrm{U}} = 16$.

For the CNN, the training set, validation set, and testing set contain 81,000, 9,000, and 19,000 samples, respectively. The parameters of each layer of the CNN are as Table 1. Adam is used as the optimizer. The epochs are set as 800 while the corresponding learning rates are 10^{-4} for the first 200 epochs, 5×10^{-5} for the next 400 epochs, and 10^{-5} for the last 200 epochs, respectively. The batch size is 128. The scaling constant is set as $c = 2$.

To measure the channel estimation performance, we use the normalized MSE (NMSE), defined as,

$$\mathrm{NMSE} = \mathbb{E}_{\mathbf{H}_k} \left\{ \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|_F^2 / \|\mathbf{H}_k\|_F^2 \right\}. \qquad (11)$$

Fig. 3 shows the NMSE performance versus signal-to-noise ratio (SNR) of LS, MMSE, and the CNN based channel estimation over 2 adjacent subcarriers in the urban micro (UMi) street non-line of sight (NLOS) scenario. The performance of the CNN based approach at single subcarrier is also plotted to demonstrate that frequency correlation is helpful to improve the channel estimation accuracy. Through offline training, the CNN based channel estimation outperforms the non-ideal MMSE with estimated covariance matrix significantly yet requiring lower estimation complexity according to this figure and Section 3.2. Moreover, the performance of the CNN based approach is very close to the ideal MMSE with true covariance matrix, especially at the low and medium SNRs.

The robustness of the MMSE and proposed CNN based approaches is shown in Fig. 4. The joint channel estimation over 2 subcarriers is considered. The CNN is trained in the UMi street NLOS scenario and is tested in both UMi street NLOS scenario and urban macro (UMa) NLOS scenario. For the MMSE, its covariance matrix is calculated in the UMi street NLOS scenario and then the channel matrix is estimated in both UMi street NLOS scenario and UMa NLOS scenario. The channel statistics are unknown to both CNN and MMSE when they predict the channels in the UMa NLOS scenario. From this figure, the CNN based channel estimation exhibits good robustness when facing the significantly different channel statistics. Even under the mismatched UMa NLOS scenario, the CNN based approach still outperforms the non-ideal MMSE without mismatch. In contrast, both the ideal and non-ideal MMSE fail to cope with the change of channel statistics and suffer significant performance loss.

5. CONCLUSIONS

In this paper, we have developed a deep CNN based joint channel estimation approach for mmWave massive MIMO-OFDM systems. By inputting the preprocessed pilots of multiple adjacent subcarriers into the CNN simultaneously, both spatial and frequency correlation are utilized to provide more accurate channel estimation. The proposed approach is with reduced complexity but outperforms the non-ideal MMSE and is close to the ideal MMSE. In the case with channel statistics mismatch, the proposed approach exhibits good robustness and outperforms the mismatched ideal and non-ideal MMSE significantly.

Acknowledgements

We would like to thank the support and valuable comments from Intel Corporation, which have helped us improve the quality of the paper significantly.

6. REFERENCES

[1] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Comm. Mag.*, vol. 52, no. 9, pp. 56−62, Sep. 2014.

[2] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393−3406, Jul. 2018.

[3] R. W. Heath Jr., N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics in Signal Process.*, vol. 10, no. 3, pp. 436−453, Apr. 2016.

[4] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653−656, Dec. 2014.

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

[6] 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.

[7] Y. Wang, W. Xu, H. Zhang, and X. You, "Wideband mmWave channel estimation for hybrid massive MIMO with low-precision ADCs," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 285−288, Feb. 2019.

[8] Z. Qin, H. Ye, G. Y. Li, and B.-H. Juang, "Deep learning in physical layer communications," *arxiv preprint arXiv:1807.11713*, Sep. 2018.

[9] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114−117, Feb. 2018.

[10] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852−855, Oct. 2018.

[11] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748−751, Oct. 2018.

[12] T. Wang, C. K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, to be published.

[13] P. Dong, H. Zhang, and G. Y. Li, "Machine learning prediction based CSI acquisition for FDD massive MIMO downlink," in *Proc. IEEE Global Commun. Conf. (Globecom'18)*, Abu Dhabi, UAE, Dec. 2018.

[14] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 5353−5360.

[15] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), TR 38.901 V15.0.0, June 2018.