CNN-based Channel Estimation using NOMA for mmWave Massive MIMO System

T. S. Anu†, Tara Raveendran†
† Department of ECE, College of Engineering, Trivandrum, Thiruvananthapuram, India
‡ Affiliated to APJ Abdul Kalam Technological University, Kerala, India.

Abstract—Channel estimation is exceptionally challenging in scenarios where NOMA schemes are integrated with millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems. An accurate estimation of the channel is essential in exploiting the benefits of the pairing of the duo-NOMA and mmWave. This paper proposes a convolutional neural network (CNN) based approach to estimate the channel for NOMA-based mmWave massive MIMO systems built on a hybrid architecture. To begin with, users are grouped into different clusters based on their distances from the base station. A coarse estimate of the channel is initially devised from the received signal and is given as input to the CNN to refine the channel coefficients. Numerical illustrations show that the proposed method outperforms the least square estimate and approaches the Cramer-Rao Bound. The proposed method is also found to be robust to different channel conditions, such as varying numbers of paths and array structures.

Index Terms—Channel estimation, CNN, mmWave massive MIMO, NOMA

I. INTRODUCTION

Millimeter wave (mmWave) communication systems have been extensively researched in recent years and are considered to be a promising technology for the next generation of wireless communications [1]. The use of mmWave frequencies for wireless communications exploits the high-frequency mmWave band (6–300GHz) wherein the spectrum is less crowded and under-utilized [2].

Non-Orthogonal multiple access (NOMA) is one of the key principles for radio access in 5G and future networks [3]. NOMA is based on the idea that multiple users are multiplexed over the same orthogonal resource block [4]. The use of NOMA in the mmWave band is significant due to: (i) NOMA with mmWave network can enable mass connectivity [3], [4]. (ii) The spectral efficiency of NOMA is higher compared to Orthogonal Multiple Access (OMA) [5]. (iii) NOMA can also lead to very low latency via grant-free transmission [6].

Accurate channel state information (CSI) is a key factor in capitalizing the benefits arising from the integration of mmWave and NOMA technology. However, obtaining accurate CSI is envisaged as one of the major challenges in such scenarios. Deep learning (DL) has been gaining increasing popularity in recent years. A DL-based algorithm, namely ChannelNet, has been proposed in [7], which considers the channel’s frequency response as an image and is applied to an image restoration network via a super-resolution network to estimate the channel. Fast and flexible denoising CNN (FFDnet) [8] has been applied to estimate the channel coefficients in [9]. A spatial-frequency CNN (SF-CNN) based estimation technique has been proposed in [10], where the channels associated with the consecutive subcarriers are fed into the CNN. A deep neural network (DNN) followed by bidirectional long short-term memory (Bi-LSTM) has been proposed in [11] to estimate and track the mmWave channel.

A conventional channel estimation method is typically associated with a tractable mathematical model. However, these models often perform poorly in practical scenarios due to reasons of channel imperfections and channel complexities. CNN, on the contrary, can learn features associated with the underlying channel from a large number of training data available to it and provide a better estimate of the channel coefficients. In this work, a novel strategy for channel estimation employing a CNN-based deep learning framework is proposed in the NOMA-based mmWave massive MIMO scenario. This framework overcomes the limitations of the existing cellular network towards achieving enhanced machine-type communication as envisaged in 5G. Numerical illustrations show that the proposed framework outperforms the conventional least square estimator and approaches the Cramer-Rao Bound (CRB). The proposed framework exhibits robustness against varying channel characteristics, say, mismatches that arise due to varying numbers of main paths and array structures.

The paper is structured as follows: Section II presents the system model. The proposed CNN-based channel estimation algorithm is detailed in Section III, followed by numerical illustrations and results in Section IV. The conclusions of this study are given in Section V.

II. SYSTEM MODEL

In what follows, the statement of the problem considered evolves via a brief description of the channel model and the clustering technique.

A. Channel Model

A mmWave massive MIMO system is considered wherein the base station (BS) has \( N_t \) antennas, which transmit signals to a receiver with \( N_r \) antennas. Let \( N_t^{RF} \) and \( N_r^{RF} \) be the number of RF chains at the transmitter and receiver, respectively. To connect more antennas to a fewer number of RF chains at both the transmitter and receiver, phase shifters are usually employed and hence assumed, \( N_t \gg N_t^{RF} \) and \( N_r \gg N_r^{RF} \).
where $L$ represents the total number of main paths. Each path $l$, has a complex path gain $\beta_l$, time delay $\tau_l$, angle of arrival $\theta_l^r \in [0,2\pi]$ at the receiver and angle of departure $\theta_l^t \in [0,2\pi]$ at the transmitter. The array response vectors $a_R(\theta_l^t) \in \mathbb{C}^{N_r \times 1}$ and $a_T(\theta_l^t) \in \mathbb{C}^{N_t \times 1}$ for uniform linear array (ULA) are given by:

$$a_R(\theta_l^t) = \frac{1}{\sqrt{N_r}} e^{-j2\pi \frac{d}{\lambda} n_r \sin \theta_l^t} ; n_r \in [0,1,...,(N_r-1)]^T$$

$$a_T(\theta_l^t) = \frac{1}{\sqrt{N_t}} e^{-j2\pi \frac{d}{\lambda} n_t \sin \theta_l^t} ; n_t \in [0,1,...,(N_t-1)]^T$$

where $d$ represents the antenna spacing and $\lambda$ represents the carrier wavelength.

### B. Dynamic Clustering

The user grouping scheme adopted from [13] exploits the distance from BS to group users into one or many clusters, thereby enhancing the sum throughput of the system. Clustering in downlink NOMA is accomplished by pairing the near-user and far-user into the same cluster based on the assumption that the near-user has high channel gain and the far-user has low channel gain. The strategy to pair the users in this manner results in an increased throughput since the near-user can achieve a reasonably higher rate with a smaller fraction of the power allocated while leaving a substantial fraction of power to the far-user. $N_c$ users are initially grouped into $N_c$ clusters, each cluster having two users (one near-user and a far-user). The channel associated with users in each cluster are grouped to form a channel matrix $H_c \in \mathbb{C}^{N_c \times N_t}$ given by:

$$H_c = [H_1 \ H_2 \ ... \ H_{N_c}]^T$$

where each $H_k = [h_i \ h_j]^T \in \mathbb{C}^{2 \times N_t}; k = 1,2,...N_c$ and $i,j \in N_r$ consists of channel associated with near-user ($h_i$) and far-user ($h_j$).

### III. CNN for Channel Estimation

In view of the present focus on CNN-based channel estimation, we formulate the algorithm for channel estimation in the NOMA-based mmWave massive MIMO scenario as described below.

#### A. Algorithm for Channel Estimation

Based on the channel model detailed in Section 2, the pilot signal matrix at the receiver for cluster $k$ is given by:

$$Y_k = W^H H_k F S + \tilde{N}$$

where $W = [w_1,..,w_{M_r}]$ and $F = [f_1,..,f_{M_t}]$ are combining and precoding matrices respectively. $M_r$ and $M_t$ are the numbers of combining and precoding vectors. $H_k$ is the channel matrix of cluster $k$. $\tilde{N} = W^H N$ is the resultant noise after combining and $N$ is AWGN with $CN(0,1)$ before combining. $S$ is the transmitted pilot symbols which is a $N_t \times N_t$ diagonal matrix given by:

$$S = (\sqrt{\alpha P} + \sqrt{(1-\alpha) P}) I$$

where $\alpha$ is the power scaling coefficient and $P$ is the total transmit power. The received signal $Y_k$ goes through a tentative estimation module as adopted from [10] and a coarse estimate of the channel matrix, namely $\hat{Y}_k; k = 1,2,...N_c$ is obtained as follows.

$$\hat{Y}_k = G_l Y_k G_r$$

where

$\begin{align*}
G_l &= \begin{cases} W, & M_t < N_r, \\ (WW^H)^{-1} W, & M_t \geq N_r \end{cases} \\
G_r &= \begin{cases} \Phi^H, & M_t < N_t, \\ (\Phi^H (\Phi^H)^{-1}, & M_t \geq N_t \end{cases}
\end{align*}$

$\hat{Y}_k$ of each cluster are then grouped to form a $N_r \times N_t$ matrix given by:

$$\hat{Y} = [\hat{Y}_1, \ ... \ \hat{Y}_{N_c}]^T$$

The input fed to CNN is the received signal matrices ($\hat{Y}$), and the output of CNN is the estimated channel matrices ($\hat{H}_c$). The mapping function between input and output is given by:

$$\{\hat{H}_c\} = f_\psi(\hat{Y}; \psi)$$

where $\psi$ is the CNN’s parameter set.

#### B. CNN Offline Training

Towards implementing the CNN training, the training data is generated according to the channel model described in (1), which subsequently undergoes the clustering strategy mentioned earlier to select users in a NOMA cluster with ($Y_q, H_q$) denoting the $q^{th}$ sample. Specifically, $Y_q$ and $H_q$ are the input and the target data corresponding to the $q^{th}$ sample. $Y_q \in \mathbb{C}^{N_r \times N_t}$ are the received signal matrices derived via (13). $H_q \in \mathbb{C}^{N_c \times N_t}$ are the corresponding true channel matrices derived via (7). $Y_q$ is then given as input to CNN to obtain the corresponding true channel matrices $H_q$.

The implementation of a NOMA-based mmWave massive MIMO system is addressed by choosing typical values of $N_r$ and $N_c$ as 32 and 16, respectively. $Y_q \in \mathbb{C}^{16 \times 32}$ is further separated into its real and imaginary counterparts resulting in two $16 \times 32$ real-valued matrices, which are being fed to CNN.

The proposed model consists of six convolution layers ($L_c$), i.e., an input layer, four hidden layers, and an output layer. In the input convolutional layer, the two 16 x 32 matrices are processed by 64, 3 x 3 x 2 convolutional filters. Rectified Linear Unit (ReLU) activation function is used as it is fast to compute and avoids vanishing gradients. The output of this layer is 64, 16 x 32 feature matrices. Zero Padding (ZP) is used to preserve the dimensions of feature matrices after every convolution. Each of the four hidden layers consists of 64, 3 x 3 x 64 filters to perform ZP convolution with the...
feature matrices from previous layers, followed by a batch normalization (BN) layer and ReLU activation function. The BN layer is used to avoid overfitting and vanishing gradients. The output layer has two 3 x 3 x 64 convolutional filters to process the 64, 16 x 32 feature matrices and outputs two 16 x 32 real-valued matrices. The two matrices at the output convolutional layer are the real and imaginary parts of each channel matrix which are then combined to obtain the estimated channel matrix \( \hat{H}_{eq} \).

The objective for the CNN training is to minimize the Mean Squared Error (MSE) loss function.

\[
MSE_{loss} = \frac{1}{N_{tr} \sum_{q=1}^{N_{r}}} \| H_{eq} - \hat{H}_{eq} \|_F^2 \tag{12}
\]

where the total number of samples in the training set is represented by \( N_{tr} \) and \( \| . \|_F \) represents the Frobenius norm.

C. Online Deployment

The tentative estimation module and the offline trained CNN will be employed in the prior receiver to the online deployment. During the online deployment phase, the received signal undergoes coarse estimation (Y) via the tentative estimation module, which is then passed through the offline trained CNN to estimate the channel coefficients (\( \hat{H}_{.} \)). The actual channel model may differ from the channels in the training set. Fine-tuning the offline trained CNN is a typical solution. However, as illustrated in Fig. 4 and Fig. 5, the offline model is found to be robust to varying channel characteristics that have never been confronted earlier. Therefore, fine-tuning provides only a minimal improvement in performance and hence is not deemed to be necessary.

D. Computational Complexity Analysis

For complexity analysis, the metric used is the number of floating point operations (FLOPs). The coarse estimation in (10) requires FLOPs of:

\[
C_{CE} \sim O(N_uN_t(N_r + N_t) + N_rN_t^2) \tag{13}
\]

where \( N_u = N_r/N_c \), is the number of users in each cluster. The computational complexity of CNN processing, according to [14], is:

\[
C_{CNN} \sim O \left( \sum_{d=1}^{L_c} m_{1,d}m_{2,d}s_d^n_{d-1}n_d \right) \tag{14}
\]

where, \( m_{1,d}, m_{2,d} \) represents the number of rows and columns of each \( d^{th} \) layer feature output (\( m_{1,d} = N_r, m_{2,d} = N_t \) for all the layers), \( s_d \) is the spatial size of the filter, \( n_{d-1}, n_d \) represents the number of \( d^{th} \) layer feature inputs and outputs. The computational complexity of proposed CNN-based channel estimation \( C_{P} \) is deduced from equations (13) and (14) as:

\[
C_{P} \sim O \left( N_uN_t(N_r + N_t) + N_rN_t^2 + \sum_{d=1}^{L_c} s_d^n_{d-1}n_d \right) \tag{15}
\]

IV. NUMERICAL ILLUSTRATIONS AND RESULTS

To explore the suitability of the proposed system and its performance, the simulation results are presented by way of implementation details, metrics used for evaluation and performance comparisons.

Implementation details: The number of antennas at the transmitter (\( N_t \)) and receiver (\( N_r \)) is chosen to be 32 and 16, respectively. Users are divided into clusters in such a way that each cluster has two users. Hence, \( N_c = N_r/2 = 8 \). Numbers of RF chains at the transmitter (\( N_{tr}^{RF} \)) and the receiver (\( N_{tr}^{RF} \)) are 2 each. \( W \) and \( F \) are set as the first \( M_r \) and \( M_t \) columns of \( N_r \times N_r \) and \( N_t \times N_t \) discrete Fourier transform (DFT) matrices respectively. \( M_r \) and \( M_t \) are chosen as 2 and 32, respectively. The power scaling coefficient \( \alpha \) is set as 0.2. The carrier frequency (\( f_c \)) and the number of main paths (\( L \)) are set as 28 GHz and 3, respectively.

The training set, validation set, and test set for the CNN contain 8100, 900, and 1900 samples, respectively. The model is optimized using Adam optimizer with a constant learning rate of \( 3 \times 10^{-4} \), batch size of 128 for 100 epochs.

Metrics used for evaluation: The performance is evaluated using Normalized Mean Square Error (NMSE). NMSE is used to measure the performance of the channel estimation and is defined as:

\[
NMSE = \mathbb{E} \left\{ \frac{\| H_c - \hat{H}_c \|_F^2}{\| H_c \|_F^2} \right\} \tag{16}
\]

Performance Comparison: The performance of the proposed technique is compared with LS estimator, SF-CNN, 1 subcarrier (an OMA technique) [10], adaptive channel estimation technique [15] and CRB. CRB represents the lower bound on the variance achievable by the estimator [16].

Fig. 1 shows the Normalized MSE performance versus signal-to-noise ratio (SNR) of CNN-based channel estimation. From the figure, it can be seen that the proposed technique is close to the CRB and outperforms the LS estimator, adaptive channel estimation technique and SF-CNN, 1 subcarrier. The improved performance of the proposed system is attributed to the duo, NOMA and DL. The proposed technique follows CRB but with a performance degradation of 3.4dB at an NMSE value of \( 10^{-1} \). However, compared to SF-CNN, the adaptive estimation technique and Least Squares, the proposed method exhibits performance advantages of 4.2dB, 10.9dB, and 11.6dB at an NMSE value of \( 10^{-1} \).

Fig. 2 shows RMSE versus epochs for different learning rates for SNR = 15dB. From the figure, it can be seen that RMSE decreases for learning rates \( 3 \times 10^{-2}, 3 \times 10^{-3} \) and \( 3 \times 10^{-4} \). However, when the learning rate is further reduced to \( 3 \times 10^{-5} \), the model takes longer time to converge for the same 20 epochs. A learning rate of \( 3 \times 10^{-4} \) gives better performance compared to other learning rates and hence is chosen for the proposed model.

Fig. 3 shows RMSE versus epochs for different numbers of convolutional layers (\( L_c \)) for SNR=15dB. As the layer increases, the model performance improves till \( L_c = 6 \). When a hidden layer is added further, i.e., \( L_c = 7 \), the models take

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
a longer time to converge for the same number of epochs. The model achieves the best performance for $L_c = 6$ and is therefore selected for the proposed model. Increasing the hidden layers in the network does not necessarily mean better performance but increased complexity in network training and testing.

**Robustness Analysis:** The channel characteristics may vary in real-world applications, and the trained model must be robust against these mismatches. A mismatch can arise due to the varying number of main paths and array structures. Fig. 4 shows the robustness of the proposed scheme against the different number of main paths ($L$). The model is trained with $L = 3$ and tested for $L = 1, 2, 3, 4$. The channel characteristics become simpler for $L = 1$ and $L = 2$. Therefore, the proposed method performs better than $L = 3$. However, for $L = 4$, the channel characteristics become slightly complex. As a result, there is a minor degradation in the performance. The trained model shows better robustness for $L = 1$ and $L = 2$ and is less robust to $L = 4$.

**V. CONCLUSIONS**

A CNN-based channel estimation approach for the mmWave NOMA system is developed. The proposed method outperforms the conventional least square estimate and approaches CRB. The proposed technique also shows robustness to varying channel conditions.
REFERENCES

[1] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5g wireless network slicing for embb, urlc, and mmtc: A communication-theoretic view,” IEEE Access, vol. 6, pp. 55 765–55 779, 2018.

[2] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” IEEE Communications Magazine, vol. 49, no. 6, pp. 101–107, 2011.

[3] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, “Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends,” IEEE Communications Magazine, vol. 53, no. 9, pp. 74–81, 2015.

[4] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2181–2195, 2017.

[5] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (noma) for cellular future radio access,” in 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), 2013, pp. i–5.

[6] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, “Wireless access for ultra-reliable low-latency communication: Principles and building blocks,” IEEE Network, vol. 32, no. 2, pp. 16–23, 2018.

[7] M. Soltnani, V. Pourrahmadi, A. Mirzaei, and H. Sheikhzadeh, “Deep learning-based channel estimation,” IEEE Communications Letters, vol. 23, no. 4, pp. 652–655, 2019.

[8] K. Zhang, W. Zuo, and L. Zhang, “Fidnet: Toward a fast and flexible solution for cnn-based image denoising,” IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608–4622, 2018.

[9] Y. Jin, J. Zhang, S. Jin, and B. Ai, “Channel estimation for cell-free mmwave massive mimo through deep learning,” IEEE Transactions on Vehicular Technology, vol. 68, no. 10, pp. 10 325–10 329, 2019.

[10] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, “Deep cnn-based channel estimation for mmwave massive mimo systems,” IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 989–1000, 2019.

[11] S. Moon, H. Kim, and I. Hwang, “Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications,” Journal of Communications and Networks, vol. 22, no. 3, pp. 177–184, 2020.

[12] 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 Communications Letters, vol. 20, no. 6, pp. 1259–1262, 2016.

[13] M. S. Ali, H. Tabassum, and E. Hossain, “Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (noma) systems,” IEEE Access, vol. 4, pp. 6325–6343, 2016.

[14] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5353–5360.

[15] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831–846, 2014.

[16] L. Bertriche, K. Abed-Meraim, and J. Belfiore, “Cramer-rao bounds for mimo channel estimation,” in 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, 2004, pp. iv–iv.

[17] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, “The capacity optimality of beam steering in large millimeter wave mimo systems,” in 2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2012, pp. 100–104.