Received December 19, 2019, accepted January 13, 2020, date of publication January 17, 2020, date of current version January 30, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2967402*

# Deep CNN and Equivalent Channel Based Hybrid Precoding for mmWave Massive MIMO Systems

**XIAOLONG BAO, WENJIANG FENG, JIALI ZHENG, AND JINGFU LI**
College of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

Corresponding author: Xiaolong Bao (bxlong@ cqu.edu.cn)

This work was supported in part by the Chongqing Basic Science and Frontier Technology Research Project under Grant cstc2017jcyjBX0047.

**ABSTRACT** Millimeter wave (mmWave) system tends to have a large number of antenna elements to compensate for the high channel path loss. The immense number of BS antennas incurs high system costs, power, and interconnect bandwidth. To circumvent these obstacles, two-step hybrid precoding algorithms that enable the use of fewer RF chains have been proposed. However, the precoding schemes already in place are either too complex or not performing well enough. In this study, an equivalent channel hybrid precoding was proposed. The part from the transmitter RF chain to the receiver RF chain is regarded as equivalent channel. By reducing the dimension of channel matrix to the level of RF link number, baseband pre-coder is simply calculated from decomposing the equivalent channel matrix $\mathbf{H}_{equ}$, which greatly reduces the complexity. Based on this novel precoding approach and convolutional neural network (CNN), a novel combiner neural network architecture was also proposed, which can be trained to learn how to optimize the combiner for maximizing the spectral efficiency with hardware limitation and imperfect CSI. Simulation results show that the proposed approaches achieve significant performance improvement.

**INDEX TERMS** mmWave, MIMO, hybrid precoding, deep learning, CNN.

## I. INTRODUCTION

With the growing popularity of mobile internet and the explosive growth of smartphones, users' usage habits have dramatically changed. The exploration of mobile broadband access capacity significantly stimulates the demand for mobile internet and boosts the growth in mobile data services, which in turn increases the demand for mobile network capacity. There are three main ways to improve the area throughput of cellular networks: 1) Allocating more bandwidth; 2) Densifying the network by deploying more BSs; 3) Improving the spectral efficiency (SE) per cell. Higher cell density and larger bandwidth have historically dominated the evolution of the coverage tier, which explains the reason why we are approaching a saturation point where further improvements are increasingly complicated and expensive. Increasing the SE corresponds to using the BSs and bandwidth that are already in place more efficiently by virtue of new modulation and multiplexing techniques. This is why major improvements in SE are needed. Therefore, studying the scheme of improving spectral efficiency becomes hot.

By adopting massive antennas, this technology significantly improves the spectrum efficiency. Especially amid large capacity demand and extensive coverage, enabling the future networks to satisfy growing network demand. Due to the short wavelength of Millimeter-wave, it is easy to realize the distribution of large-scale antenna array, so Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) has already been considered as a promising solution to meet the requirement of the higher data rate for the future network. On account of the limited physical space with closely placed antennas and prohibitive power consumption in mmWave massive MIMO systems, it is difficult to equip a dedicated radio frequency (RF) chain for each antenna. To reduce complexity and cost, phase shifter based two-stage structure, which is usually called hybrid precoding, is widely used at both the transmitter and the receiver to connect a large number of antennas with much fewer RF chains. The hybrid precoding structure is shown in Figure 1.

Several methods are proposed to design the hybrid precoder. In the existing work, using this hybrid structure based

The associate editor coordinating the review of this manuscript and approving it for publication was Soon Xin Ng.

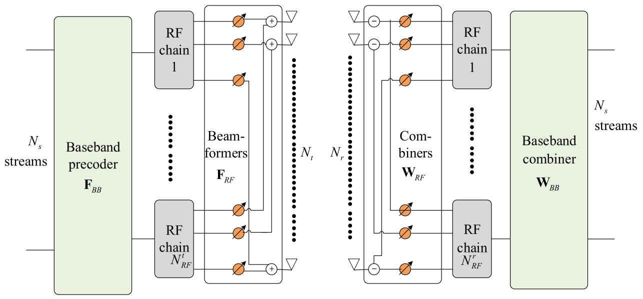



**FIGURE 1.** massive MIMO hybrid pre-coding system.

approach to handle this difficulty, a spatially sparse algorithm via orthogonal matching pursuit (SOMP) was proposed in [1]. However, the analog beam-former is limited to a pre-defined codebook.

To enhance the performance of SOMP, the manifold optimization method was applied in [3], [4] for the analog BF optimization. Using the connection between the optimum and the hybrid precoder, [5] proposed an alternating minimization approach via manifold optimization (MO-AltMin) to estimate the analog beam-former and baseband precoder. The above works provide optimization-based and greedy-based solutions for hybrid precoding problem. The complicated computations are the main drawbacks of the above techniques. In order to circumvent this issue, DNN (deep neural network) based technique was proposed to deal with the beam-forming problem [8]. Although the DNN-based technique has a low computational complexity, since its training strategy is based on geometric mean decomposition(GMD) algorithm, the system performance is far away to the performance provided by singular value decomposition(SVD)-based all-digital precoding. All of these algorithms require some approximations to simplify the original objective function, a lot of serial time consuming iterations are needed. There are still a lot of problems remaining and two major challenges are the extraordinarily high computational complexity and poor system performance. Moreover, perfect channel state information (CSI) is assumed in all of these algorithms. Recently, many researchers pay attention to use compressive channel sensing vectors to process hybrid precoding. A compressive channel sensing vectors based neural network architecture was proposed in [9], which uses compressive channel sensing vectors and DNN to deal with hybrid precoding problem.

In this study, firstly, a hybrid precoding algorithm based on equivalent channel was proposed. It's obvious that svd-based full digital precoding scheme can provide very superior performance. In hybrid precoding structure, this decomposition method can also be used to obtain baseband precoder which is designed to eliminate interference between multiple data streams. When the beam-former and combiner are obtained, the beamforming matrix and combining matrix are obtained by alternating search for maximum spectral efficiency and the part from the transmitter RF chain to the receiver RF chain can be equivalent to channel. By reducing the dimension of channel matrix to the level of RF link number, baseband precoder is simply calculated from decomposing the equivalent channel matrix $\mathbf{H}_{equ}$, which greatly reduces the complexity. Simulation results show that the proposed scheme performs excellent and has low complexity. Based on this novel pecoding approach, a novel combiner convolutional neural network architecture was proposed, which can be trained to learn how to optimize the combiner for maximizing the spectral efficiency with hardware limitation and imperfect CSI. The input of convolutional neural network (CNN) is channel matrix and the output is the combiner weights. In order to train the network, different channel realizations with synthetic noise added to each input data was generated. Compared with the conventional work that assuming perfect CSI [1], [5], [8] is known, the CNN-based approach is expected to process imperfect CSI with strong robustness.

*Notation:* The following notation are used throughout this paper: **A** is a matrix, ***a*** is a vector, a is a scalar. |A| is the determinant of A, whereas $\mathrm{A}^{\mathrm{T}}$ and $\mathrm{A}^{*}$ are its transpose and Hermitian (conjugate transpose). $NC(\mathrm{m}, \mathrm{R})$ is a complex Gaussian random vector with mean m and covariance R. $E[\cdot]$ is used to denote expectation.

## II. SIGNAL MODEL AND PROBLEM FORMULATION

### A. SYSTEM AND CHANNEL MODEL

Considering the fully-connected hybrid architecture depicted in Figure 1, where a transmitter employs $N_t$ antennas and $N_{RF}^t$ RF (Radio Frequency) chains is communicating via $N_s$ streams with a receiver which has $N_r$ antennas and $N_{RF}^r$ RF chains. The transmitter precodes the transmitted signal using a $N_{RF}^t \times N_s$ baseband pre-coder $\mathbf{F}_{BB}$ and a $N_t \times N_{RF}^t$ RF beam-former $\mathbf{F}_{RF}$, while the receiver combines the received signal using the $N_r \times N_{RF}^r$ RF combiner $\mathbf{W}_{RF}$ and the $N_{RF}^r \times N_s$ baseband combiner $\mathbf{W}_{BB}$, which are subject to constraints $N_s \leq N_{RF}^t \leq N_t$ and $N_s \leq N_{RF}^r \leq N_r$.

Suppose the transmit power of $\mathbf{s}$ is normalized, i.e. $\mathrm{E}\left[SS^*\right] = \frac{1}{N_s}\mathbf{I}_{\mathbf{N_s}}$. Since the analog RF beam-formers and combiners, $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$, are implemented in the analog domain using RF circuits, every entry of the RF beam-former/combiner is assumed to have a constant modulus, i.e. $[\mathrm{F}_{RF}]_{m,n} = \frac{1}{\sqrt{N_t}}e^{-j\theta_{m,n}t}$ (and similarly for elements of $\mathbf{W}_{RF}$). Further, the total power constraint is enforced by normalizing the baseband precoder $\mathbf{F}_{BB}$ to satisfy $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_s$. After precoding in the baseband domain, the resultant vector given as $\mathbf{F}_{BB}\mathbf{s}$ is the input of $N_{RF}^t$ chains for the up conversion. Next, an analog beam-former $\mathbf{F}_{RF}$ is applied for adjusting phase/angle to maximize the system capacity or minimize the interference according to the system requirements. Therefore, the discrete-time transmitted signals is finally represented as $\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s}$.

For narrowband block-fading propagation scenarios, the received signals at the receiver can be shown as (1), where $\mathbf{y}$ is the $N_r \times 1$ received signal vector, $\mathbf{H}$ is the channel matrix with $N_r \times N_t$ dimensions, $\rho$ is the average received power, and n is the additive white Gaussian noise (AWGN) vector following i.i.d. distribution $CN(0, \sigma_n^2)$.

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}. \tag{1}$$

The transmitted signal is received and processed by analog and baseband combiners as (2)

$$\widetilde{\mathbf{y}} = \sqrt{\rho}\mathbf{W}_{BB}^{*}\mathbf{W}_{RF}^{*}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{W}_{BB}^{*}\mathbf{W}_{RF}^{*}\mathbf{n}. \tag{2}$$

Using the clustered channel model, the matrix channel $\mathbf{H}$ is assumed to be the sum of the contributions of $N_{cl}$ scattering clusters, and each of them contributes $N_{ray}$ propagation paths to the channel matrix $\mathbf{H}$. Therefore, the discrete-time narrowband channel $\mathbf{H}$ can be written as

$$\mathbf{H} = \gamma \sum_{i,l} \alpha_{i,l}\Lambda_r\left(\varphi_{i,l}^r, \theta_{i,l}^r\right)\Lambda_t\left(\varphi_{i,l}^t, \theta_{i,l}^t\right) \times a_r\left(\varphi_{i,l}^r, \theta_{i,l}^r\right) a_t\left(\varphi_{i,l}^t, \theta_{i,l}^t\right)^{*}. \tag{3}$$

where $\gamma$ is a normalization factor such that $\gamma = \sqrt{N_t N_r / N_{cl} N_{ray}}$ and $\alpha_{i,l}$ is the complex gain of the $l^{th}$ ray in the $i^{th}$ scattering cluster, whereas $\varphi_{i,l}^r(\theta_{i,l}^r)$ and $\varphi_{i,l}^t(\theta_{i,l}^t)$ are its azimuth (elevation) angles of arrival and departure respectively. The functions $\Lambda_r(\varphi_{i,l}^r, \theta_{i,l}^r)$ and $\Lambda_t(\varphi_{i,l}^t, \theta_{i,l}^t)$ represent the transmitter's and receiver's antenna element gain at the corresponding angles of departure and arrival. Finally, the vectors $a_r(\varphi_{i,l}^r, \theta_{i,l}^r)$ and $a_t(\varphi_{i,l}^t, \theta_{i,l}^t)$ represent the normalized receive and transmit array response vectors at the azimuth (elevation) angle of $\varphi_{i,l}^r(\theta_{i,l}^r)$ and $\varphi_{i,l}^t(\theta_{i,l}^t)$ respectively.

Note that when uniform planar array structure is considered, both $\boldsymbol{a}_r(\bullet)$ and $\boldsymbol{a}_t(\bullet)$ can be expressed as the format given as (5), where W and H are the numbers of antenna elements, $k = 2\pi/\lambda$ ($\lambda$ is the corresponding wavelength) and d is the inter-element spacing between two neighbor antennas, $N = WH$ is the size of the antenna array. For simplicity, we use $\mathbf{A}_t$ and $\mathbf{A}_r$ to represent the matrices consisting of all array response vectors $\boldsymbol{a}_r(\bullet)$ and $\boldsymbol{a}_t(\bullet)$ respectively. These matrix representations will be used in the sequel.

$$a_{UPA}(\varphi, \theta) = \frac{1}{\sqrt{N}}[1, \ldots, e^{jkd(m\sin(\varphi)\sin(\theta)+n\cos(\theta))}, \ldots, e^{jkd((W-1)\sin(\varphi)\sin(\theta)+(H-1)\cos(\theta))}] \tag{4}$$

### B. PROBLEM FORMULATION

Given the system model (2), its spectral efficiency is described as (5) [1], where $R_n = \sigma_n^2\mathbf{W}_{BB}^{*}\mathbf{W}_{RF}^{*}\mathbf{W}_{RF}\mathbf{W}_{BB}$ is the noise covariance matrix, the received signal is combined at the receiver. Therefore, for mm-Wave massive MIMO systems, the task of hybrid pre-coding is to maximize $R$ through designing $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ jointly.

$$R = \log_2(|\mathbf{I}_{N_s} + \frac{\rho}{N_s}R_n^{-1}\mathbf{W}_{BB}^{*}\mathbf{W}_{RF}^{*}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB} \times \mathbf{F}_{BB}^{*}\mathbf{F}_{RF}^{*}\mathbf{H}^{*}\mathbf{W}_{RF}\mathbf{W}_{BB}|) \tag{5}$$

To maximize R in (5), $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$ should be optimal. Nevertheless, due to the non-convex constraints of $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$, it is extremely difficult to tackle the optimization problem maximizing $R$ via $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$ and $\mathbf{W}_{RF}$. Consequently, the problem of designing analog/digital precoders and combiners is usually decoupled into two independent sub-problems. One is to design precoder for the transmitter, another is to design combiner for the receiver, and both of them follow the similar approach to obtain solutions.

Generally, digital precoding is an upper bound for hybrid precoding. For full digital precoding, the unconstrained optimal precoder ($\mathbf{F}_{opt}$) is simply calculated from decomposing the channel matrix $\mathbf{H}$, which is known at the transmitter. Singular Value Decomposition (SVD) is applied to the channel matrix $\mathbf{H}$, such that $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{H}}$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices with dimensions ($N_r \times$ rank($\mathbf{H}$) and ($N_t \times$ rank($\mathbf{H}$)), respectively, and $\Sigma$ is a diagonal matrix with dimensions (rank($\mathbf{H}$) $\times$ rank($\mathbf{H}$)). Taking advantage of the above decomposition, $\mathbf{F}_{opt}$ equals the first $N_s$ columns of $\mathbf{V}$, which are orthonormal basis for the channel's row space.

In this study, transmitter is assumed to use optimal precoders ($\mathbf{F}_{opt}$) to transmit the streams, this makes optimize receiver combiners possible, the achievable mutual information of the mm-Wave massive MIMO system can be shown as (6).

$$\begin{aligned} &MAX(R) \\ &\quad = \log_2(|\mathbf{I}_{N_s} + \frac{\rho}{N_s}R_n^{-1}\mathbf{W}_{BB}^{*}\mathbf{W}_{RF}^{*}\mathbf{H}\mathbf{F}_{\text{opt}}\mathbf{F}_{opt}^{*}\mathbf{H}^{*}\mathbf{W}_{RF}\mathbf{W}_{BB}|) \\ &s.t: \mathbf{W}_{RF} \in \mathcal{W}_{RF}, \quad [\mathcal{W}_{RF}]_{n,m} = \frac{1}{\sqrt{N_r}}e^{i\theta_{n,mt}}. \end{aligned} \tag{6}$$

where $\mathcal{W}_{RF}$ is the set that contains all the possible analog combiners considering the hardware constraints, which consists of the sending phase of each path of each cluster. The aim is to optimize $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ to make $R$ as big as possible. Once the analog combiners $\mathbf{W}_{RF}$, $\mathbf{W}_{BB}$ have been designed, the precoder $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$ at the transmitter can be obtained via novel equivalent channel approach.

## III. HYBRID PRECODING STRATEGIES

This part shows a novel approach to optimize hybrid pre-coder and using the novel pre-coding approach to build a neural network model which can be adopted in the mm-Wave massive MIMO for achieving end-to-end highly efficient hybrid pre-coding. Finding the transmission paths that make the spectrum most efficient and adjusting the transmission phase of the signal through the beam-former so that the system spectral efficiency would be improved. Then regard the beam-former and combiner as a part of channel. By reducing the dimension of channel matrix to the level of RF link number, baseband pre-coder is simply calculated from decomposing the equivalent channel matrix $\mathbf{H}_{equ}$, which greatly reduces the complexity. As for the neural network, the splendid learning ability of deep learning enables the spatial features to be exploited of the mm-Wave massive MIMO system and regard the entire system as a black box to capture useful features for hybrid precoding. Developing the proposed CNN framework and describing how can the nonlinear operation be mapped to the hybrid pre-coder, and then we provide a novel training strategy for facilitating the performance of the CNN.

### A. NOVEL EQUIVALENT CHANNEL HYBRID PRECODING

To solve the optimization problem in (5), its essence is to select out the $N_{RF}^r$ most relevant antenna response vectors from $\{a_r\left(\varphi_{i,l}^r, \theta_{i,l}^r\right), \forall_{i,l}\}$, which is denoted as $\mathbf{A}_r$ for notational simplicity. Assuming the precoders are optimal, we seek to design hybrid combiners $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ that minimize the mean-squared-error (MSE) between the transmitted and processed received signals. The combiner design problem can therefore be stated as (7)

$$(\mathbf{W}_{RF}, \mathbf{W}_{BB}) = \arg\min \mathbb{E}\left[\left\|\mathbf{s} - \mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{y}\right\|_2^2\right]$$
$$s.t: \mathbf{W}_{RF} \in \mathcal{W}_{RF}, \quad [\mathcal{W}_{RF}]_{n,m} = \frac{1}{\sqrt{N_r}}e^{i\theta_{n,mt}}. \tag{7}$$

In the absence of any hardware limitations that restrict the set of feasible linear receivers, the exact solution to (7) is well known [19] to be

$$\begin{aligned}\mathbf{W}_{MMSE} &= \mathbb{E}\left[\mathbf{s}\mathbf{y}^*\right]\mathbb{E}\left[\mathbf{y}\mathbf{y}^*\right]^{-1}\\ &= \frac{1}{\sqrt{\rho}}(\mathbf{F}_{opt}^*(\mathbf{H}^*\mathbf{H})\mathbf{F}_{opt} + \frac{\sigma_n^2}{\rho}N_s\mathbf{I}_{Ns})^{-1}\mathbf{F}_{opt}^*\mathbf{H}^* \end{aligned}\tag{8}$$

As a result of [1], the MMSE estimation problem is equivalent to finding hybrid combiners that solve

$$(\mathbf{W}_{RF}, \mathbf{W}_{BB}) = \arg\min \left\|\mathbb{E}[\mathbf{y}\mathbf{y}^*]^{1/2}(\mathbf{W}_{MMSE} - \mathbf{W}_{RF}\mathbf{W}_{BB})\right\|_F$$
$$s.t.\ \mathbf{W}_{RF} \in \mathcal{W}_{RF}. \tag{9}$$

which amounts to finding the projection of the unconstrained MMSE combiner $\mathrm{W_{MMSE}}$ onto the set of hybrid combiners of the form $\mathbf{W}_{RF}\mathbf{W}_{BB}$ with $\mathbf{W}_{RF} \in \mathcal{W}_{\mathrm{RF}}$

$\mathbf{A}_r = [a_r\left(\varphi_{1,1}^r, \theta_{1,1}^r\right), \cdots, a_r\left(\varphi_{N_{cl},N_{ray}}^r, \theta_{N_{cl},N_{ray}}^r\right)$ is a $N_r \times N_{cl}N_{ray}$ receive matrix of array response vectors. Firstly, we calculate all the combinations $\mathrm{Qw} = \begin{pmatrix} N_{cl}N_{ray} \\ N_{RF}^r \end{pmatrix}$ of $N_{RF}^r$ paths that select from all the transmission paths and get all possible antenna response vectors to form a combiner matrix. All combinations constitute candidate sets of combiner $\mathcal{W}_{RF} = [W_{RF}^1\ W_{RF}^2\ \cdots\ W_{RF}^{Qw}]$. That's mean we need to choose $N_{RF}^r$ columns from $\mathrm{A_r}$ to get $\mathbf{W}_{\mathrm{RF}}$. Namely, because of the structure of clustered mm-Wave channels, near-optimal receivers can be found by further constraining $\mathbf{W}_{RF}$ to select elements from corresponding column of $\mathbf{A}_r$. Consequently, Solve problem (9) turns into solving

$$\mathbf{W}_{BB} = arg\ min \left\|\mathbb{E}[\mathbf{y}y^*]^{1/2}\mathbf{W}_{MMSE} - \mathbb{E}[\mathbf{y}\mathbf{y}^*]^{1/2}\mathbf{W}_{RF}\tilde{\mathbf{W}}_{BB}\right\|_F$$
$$s.t.\ \left\|diag(\mathbf{W}_{BB}\mathbf{W}_{BB}^*)\right\|_0 = N_{RF}^r \tag{10}$$

Suppose the transmit power of $\mathbf{s}$ and the receive power of $\mathbf{y}$ are normalized

$$\mathbf{P}_t = \frac{1}{N_S}\mathbf{I}_{Ns}, \quad \mathbf{P}_r = \rho\mathbf{H}\mathbf{F}_{opt}\mathbf{P}_t\mathbf{F}_{opt}^*\mathbf{H}^* + \sigma_n^2\mathbf{I}_{N_r} \tag{11}$$

The effective robust digital combiner are given by [1]

$$\mathbf{W}_{BB} = (\mathbf{W}_{RF}^*\mathbf{P}_r\mathbf{W}_{RF})^{-1}\mathbf{W}_{RF}^*\mathbf{P}_r\mathbf{W}_{MMSE}. \tag{12}$$

When we get $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ candidate sets, put them into the spectrum efficiency formula (6), and the candidate set for maximum spectrum efficiency is the optimal combiner. Owing to the result of $\mathbf{W}_{RF}$, we can know the downlink equivalent receiving channel $\mathbf{H}_{eqd} = \mathbf{W}_{RF}^*\mathbf{H}$. After the downlink receiving channel matrix is obtained, we know the phase of the receiving signal with the maximum gain of the downlink channel, and the maximum spectrum efficiency can be obtained by letting the transmitter use the phase information to transmit. We focus on the large array gain design through the RF combining and leave the interference canceling to the baseband processing in the proposed scheme. Using the RF domain processing matrices $\mathbf{W}_{RF}$ and $\mathbf{F}_{RF}$ construct the equivalent channel $\mathbf{H}_{equ} = \mathbf{W}_{RF}^*\mathbf{H}\mathbf{F}_{RF}$ and $\mathbf{H}_{equ}$ should sufficiently harvest the array gain so that it can provide as large gain for each stream transmission as possible, which will lead to a good performance and significantly low complexity under the hybrid precoding structure. Based on the obtained baseband equivalent channel $\mathbf{H}_{equ}$ performing the singular value decomposition (SVD) processing with low dimension $\mathbf{H}_{equ}$, the baseband precoder $\mathbf{F}_{BB}$ and combiner $\mathbf{W}_{BB}$, which are used to cancel the inter-streams interference, are available.

Finally, Algorithm 1 describes the framework to obtain the feasible hybrid precoder based on the principle of equivalent channel.

### B. PROPOSED CNN-BASED NEURAL NETWORK ARCHITECTURE

Based on the scheme mentioned above, the novel combiner neural network architecture was proposed show in Figure 2, which learns how to predict the RF combining vectors of the hybrid architecture directly from the receive channel state information. Supervised learning is tasked with learning a function from labeled training data in order to predict the value of any valid input.

The neural network has a total of eight layers, including an input layer, two convolution layers, a pooling layer, three full connection layers and a regression output layer. Many units are deployed in each layer, and the output can be generated based on the output of these units with the aids of activation functions. In most cases, the rectified linear unit (ReLU) function and the sigmoid function are used in the nonlinear operation. Assuming $\alpha$ as the argument, they are defined as $\mathrm{ReLU}(\alpha) = \max(0, \alpha)$ and $\mathrm{sigmoid}(\alpha) = \frac{1}{1+e^{-\alpha}}$, respectively.

The input layer inputs the 3d matrix generated by the channel matrix. The first channel of the input is the element-wise absolute value of the channel matrix as $[\mathbf{X}^{(l,n)}]_{i,j,1} = |[\mathbf{H}^{(l,n)}]_{i,j}|$. The second and the third channels are defined as the real and the imaginary parts of the channel matrix as $[\mathbf{X}^{(l,n)}]_{i,j,2} = \mathrm{Re}\{[\mathbf{H}^{(l,n)}]_{i,j}\}$ and $[\mathbf{X}^{(l,n)}]_{i,j,3} = \mathrm{Im}\{[\mathbf{H}^{(l,n)}]_{i,j}\}\ \forall_{ij}$. The second and third layers are the

**Algorithm 1** Equivalent Channel Hybrid Precoding

inputs: $\mathbf{H}$, $\mathbf{A}_t$, $\mathbf{A}_r$, $N_{RF}^t$, $N_s$, $N_{RF}^r$

ouputs: $\mathbf{W}_{BB}$, $\mathbf{W}_{RF}$, $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$

calculate Combination $\mathrm{Qw} = \begin{pmatrix} N_{cl}N_{ray} \\ N_{RF}^r \end{pmatrix}$, subsetWRF(Qw)

$[\sim, \sim, \mathbf{F}_{opt}] = SVD(\mathbf{H})$,

$\mathbf{W}_{MMSE} = \frac{1}{\sqrt{\rho}}(\mathbf{F}_{opt}^*(\mathbf{H}^*\mathbf{H})\mathbf{F}_{opt} + \frac{\sigma_n^2}{\rho}N_s\mathbf{I}_{Ns})^{-1}\mathbf{F}_{opt}^*\mathbf{H}^*$.

**for** $qw = 1$: Qw **do**

$\mathbf{W}_{\mathrm{F}}(qw) = \mathbf{A}_{\mathrm{r}}(:,\mathrm{subsetWRF}(qw)))$;

$\mathbf{W}_{\mathbf{B}}(qw) = (\mathbf{W}_F^*\mathbf{P}_r\mathbf{W}_F)^{-1}\mathbf{W}_F^*\mathbf{P}_r\mathbf{W}_{MMSE}$

$R(qw) = \log_2(|\mathbf{I}_{N_s} + \frac{\rho}{N_s}R_n^{-1}\mathbf{W}_B^*(qw)\times \mathbf{W}_F^*(qw)\mathbf{H}\mathbf{F}_{opt}\mathbf{F}_{opt}^*\mathbf{H}^*\mathbf{W}_F(qw)\mathbf{W}_B(qw)|)$

**end**

$[\sim, qwm] = \max(R(:, 1))$

$\mathbf{W}_{RF} = \mathbf{W}_F\,(qwm)$, $\mathbf{W}_{BB} = \mathbf{W}_B\,(qwm)$

$\mathbf{H}eqd = \mathbf{W}_{RF}^*\mathbf{H}$

$\mathbf{F}_{RF} = \frac{e^{j(angle(\mathbf{H}\mathrm{eqd}^*))}}{\sqrt{N_t}}$

$\mathbf{H}equ = \mathbf{W}_{RF}^*\mathbf{H}\mathbf{F}_{RF}$

$[\sim, \sim, \mathbf{F}_{BB}] = SVD(\mathbf{H}equ)$

$\mathbf{F}_{BB} = \frac{\sqrt{N_s}\mathbf{F}_{BB}}{\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F}$

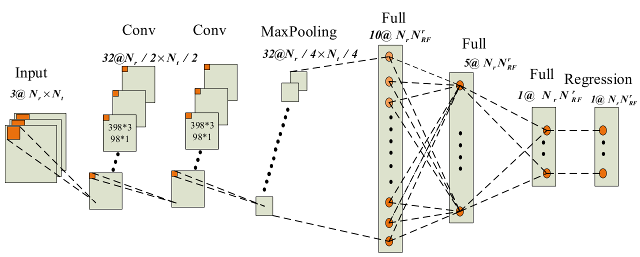



**FIGURE 2.** The proposed CNN framework for combiner design.

convolutional layers with 32 filters and the convolution kernel is $3 \times 3$. The activation functions in the convolution layer are all ReLU functions. The main task of convolutional layers filters are to extract and select data feature vectors. CNN can automatically extract and select feature vectors from data to obtain representative vectors from input data. After that, the parameters to be estimated are compressed by a 2D maximum pooling layer, and the data dimension after pooling is 1/4 or 1/8 of the original dimension. Then, the data after pooling is rearranged into a one-dimensional vector. The next full connection layers make a weighted sum of the features of the previous layer and map the feature space to the sample marker space by linear transformation. In order to prevent overfitting, dropout Layer is added, so that the neural network is sampled in the training process, while dropout is not used in the prediction, and the random sampling probability is usually set at 0.5. And the next two layers are full connection layer. The activation functions in the full connection layers are also all ReLU functions. Finally, the regression layer is used as the output layer to output the results with computing the mean-squared-error loss.

In the training phase, the combiner is trained end-to-end in a supervised manner. More specifically, a dataset of the mm-Wave channels and the corresponding RF combining matrices are constructed and the combiner is trained to be able to predict the indices of the RF combining vectors of the hybrid architecture given the input channel vector. The training labels, RF combining matrices, are calculated based on the scheme mentioned above, equivalent channel hybrid precoding algorithm. For the loss function, we use the mean square error (MSE) for the multi-label regression that corresponds to each task, so the prediction loss of the model is

$$J_{loss}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(f_\theta(x_i) - W_{RF}(i,:))^2 \tag{13}$$

Then, based on the proposed method, the stochastic gradient descent (SGD) algorithm with momentum is employed to update the parameter sets ($\theta$) of the network, which is given by

$$\begin{aligned} \boldsymbol{\theta}^{j+1} &= \boldsymbol{\theta}^j + v \\ v &= \alpha v - \varepsilon g \\ g &= \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = \frac{1}{n}\sum_{i=1}^{n}\nabla(f_{\theta_j}(x_i) - W_{RF}(i,:))^2 \end{aligned} \tag{14}$$

where $v$ *is* denoted the velocity for facilitating the gradient element, the iteration number is denoted as j, $\alpha$ denotes the momentum parameter and $\varepsilon$ denotes the learning rate. Synchronously, g represents the gradient element. The momentum term increases for dimensions whose gradients point in the same directions and reduces updates for dimensions whose gradients change directions. As a result, the convergence of the training process is accelerated and reduced oscillation.

In order to improve the processing capacity of imperfect channel state information, such collocation is carried out when generating training data sets. For each perfect channel state information matrix was generated, nine imperfect channel state information matrices were generated, which were generated by using the same perfect channel state information matrix to add noise. However, all training tags are produced by using perfect channel state information. Nine noisy channel matrices are obtained, where the added element-wise synthetic noise is defined by

$$\mathbf{H}_{noise} = \mathbf{H} + \mathbf{n} \tag{15}$$

where $\mathbf{H}$ is the channel matrix with $N_r \times N_t$ dimensions, and n is the additive white Gaussian noise (AWGN) vector following i.i.d distribution $CN(0, \sigma_n^2)$.

During the training process, the CNNs are fed with the training data generated for N = 1000, L = 10. In the training process, 70% and 30% of all generated data are selected as the training and validation datasets, respectively. The parameters of the model include ordinary parameters and hyperparameters (some parameters related to model design and training). Using the BP (Error back-propagation) algorithm can only train ordinary parameters, but cannot "train" the hyperparameters of the model.

Therefore, set the verification data set and feedback through the effect of the verification data set.

According to the effect, it is necessary to terminate the current model training or retrain after changing the hyperparameters, and finally get the best model. The initial learning rate is set to 0.01, the number of training iterations is set to 5000, and the size of each batch is set to 200.We summarize the algorithmic steps of the training data generation in Algorithm 2.

**Algorithm 2** CNN-Based Neural Network Architecture

1: inputs: $\mathbf{H}$, $\mathbf{A}_t$, $\mathbf{A}_r$, $N_{RF}^t$, $N_{RF}^r$, $N_s$, N, L =10
2: ouputs: $\mathbf{W}_{BB}$, $\mathbf{W}_{RF}$, $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$
3: Construct the proposed CNN framework.
4: Generate train data:
  **for** $1 \le n \le N$
    **for** $1 \le l \le L$ **do**
      **if** $l = 1$, $\mathbf{H}^{(l,n)} = \mathbf{H}$; generate perfect CSI; use Algorithm 1 to get $\mathbf{W}_{RF}^{l,n}$ as label
      **else** $\mathbf{H}^{(l,n)} = \mathbf{H} + \mathbf{n}$; generate imperfect CSI; copy $\mathbf{W}_{RF}^{l,n}$ as label
      $[\mathbf{X}^{(l,n)}]_{i,j,1} = |[\mathbf{H}^{(l,n)}]_{i,j}|$.
      $[\mathbf{X}^{(l,n)}]_{i,j,2} = \mathrm{Re}\{[\mathbf{H}^{(l,n)}]_{i,j}\}$.
      $[\mathbf{X}^{(l,n)}]_{i,j,3} = \mathrm{Im}\{[\mathbf{H}^{(l,n)}]_{i,j}\}\ \forall ij$.
    **end**
  **end**
5: Feeding $\mathbf{X}$ to train the CNN by processing the SGD with momentum
6: Use the CNN to predict $\mathbf{W}_{RF}$
7: Use Algorithm 1 to get $\mathbf{F}_{RF}$ $\mathbf{W}_{BB}$ $\mathbf{F}_{BB}$

## IV. NUMERICAL SIMULATIONS

### A. COMPLEXITY ANALYSIS

One of the key advantages of the proposed EQUA hybrid precoding is that this method decreases the computational complexity. To verify the low computational complexity of the EQUA scheme intuitively, the computational complexity of the proposed method and some other typical precoding approaches are presented, which is illustrated as TABLE 1. According to [1], the main complexity of SOMP hybrid precoding comes from the following three parts:

1) The first one is derived from the channel matric $\mathbf{H}$ SVD computation of $\mathbf{F}_{opt}$ and $\mathbf{W}_{opt}$, since $\mathbf{H}$ is $N_r \times N_t$, so the main complexity is $\mathcal{O}(3N_t^2 N_r)$ [20].

2) In order to get $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$, the main complexity are the calculation of $\mathbf{\Psi}\mathbf{\Psi}^*$ ($\mathbf{\Psi} = \mathbf{A}_t * \mathbf{F}_{opt}$) and invert $\mathbf{F}_{RF}^*\mathbf{F}_{RF}$ The main complexity of this part is $\mathcal{O}(N_{RF}^t(\mathrm{L}\mathrm{N}_t\mathrm{N}_s + L^4\mathrm{N}_s + (N_{RF}^t)^5\,\mathrm{N}_t + \mathrm{N}_t N_{RF}^t \mathrm{N}_s))$. L ($\mathrm{L} = N_{cl}N_{ray}$) is the number of transmission paths.

3) The third one originates from the computation of $\mathbf{\Psi}\mathbf{\Psi}^*$ ($\mathbf{\Psi} = \boldsymbol{A}_r * \boldsymbol{E}[\mathbf{y}\mathbf{y}^*] * \mathbf{W}_{MMSE}$) and invert $\mathbf{W}_{RF}^*\mathbf{W}_{RF}$ to get $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$. The main complexity of this part is $\mathcal{O}(N_{RF}^r(\mathrm{L}\mathrm{N}_r\mathrm{N}_s + L^4\mathrm{N}_s + (N_{RF}^r)^5\,\mathrm{N}_r + \mathrm{N}_r N_{RF}^t \mathrm{N}_s))$.

As for the EQUA hybrid precoding, we can see from algorithm 1 that the main complexity of the first four steps are the same to the first part of SOMP [1]. The main complexity stem from the calculation of RF sections in step 5 and the singular value decomposition of equivalent channel matrix $\mathbf{H}_{equ}$ in step 10. The complexity of calculation for $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ is $\mathcal{O}(\binom{L}{N_{RF}^r}((N_{RF}^r)^3 + 2\mathrm{N}_s N_{RF}^r \mathrm{N}_r + 2N_s^2 N_t^2))$. The complexity of calculation for $\mathbf{F}_{BB}$ and $\mathbf{F}_{RF}$ is $\mathcal{O}(3\,(N_{RF}^r)^3)$.

The complexity of the MO-AltMin [5] algorithm is relatively high. In each iteration, the update of the analog precoder involves a linear search algorithm, so the nested loops in the MO-AltMin algorithm will slow down the whole solving procedure. Furthermore, the Kronecker products in calculation Euclidean gradient of the cost function will result in two matrices of dimension $N_{RF}^t \mathrm{N}_t \times \mathrm{N}_t \mathrm{N}_s$, which scales with the antenna size and results in an exponential increase of the computational complexity in the MO-AltMin algorithm. According the algorithm description in [5], the main complexity that calculate for $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$ is $\mathcal{O}(K^2(N_{RF}^t N_t N_S + (N_{RF}^t)^2 N_t^5 N_S + (N_{RF}^t)^2\, N_t^3))$. K is the number satisfied with stopping criterion triggers. And the main complexity that calculates for $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ is

$$\mathcal{O}(K^2(N_{RF}^r N_r N_S + (N_{RF}^r)^2 N_r^5 N_S + (N_{RF}^r)^2\, N_r^3)).$$

The complexity of the CNN-based precoding is minimal. Because convolution layers have 32 filters and the convolution kernel is $3 \times 3$, the input layer has 3 channels and the dimension of output layer is $N_r N_{RF}^r \times 1$. The main complexity $\mathcal{O}\left(96\mathrm{k}N_t + 320 N_{RF}^t N_r + 55\,(N_{RF}^t)^2 N_r^2\right)$.

From TABLE 1, we can find the highest degree of SOMP is $N_{RF}^t$ to the sixth, the highest degree of MO-AltMin is $N_t$ to the fifth and $N_r$ to the fifth. However the highest degree of EQUA is only $N_{RF}^r$ to the third and $N_{RF}^r$ is far less than $N_t$. Overall comparison, the complexity of algorithm EQUA is much less than algorithm MO-AltMin and algorithm SOMP. And the CNN-based precoding is the simplest of these algorithms.

**TABLE 1. Computational complexity of several precoding schemes of mmWave massive MIMO.**

| Methods | Computational complexity |
|---|---|
| SOMP | $\mathcal{O}(3N_t^2 N_r + N_{RF}^t(\mathrm{L}\mathrm{N}_t\mathrm{N}_s + L^4\mathrm{N}_s + (N_{RF}^t)^5\mathrm{N}_t + \mathrm{N}_t N_{RF}^t \mathrm{N}_s)$ $+N_{RF}^r(\mathrm{L}\mathrm{N}_r\mathrm{N}_s + L^4\mathrm{N}_s + (N_{RF}^r)^5\mathrm{N}_r + \mathrm{N}_r N_{RF}^t \mathrm{N}_s))$ |
| MO-AltMin | $\mathcal{O}(3N_t^2 N_r + K^2(N_{RF}^t N_t N_S + (N_{RF}^t)^2 N_t^5 N_S + (N_{RF}^t)^2 N_t^3)$ $+K^2(N_{RF}^r N_r N_S + (N_{RF}^r)^2 N_r^5 N_S + (N_{RF}^r)^2 N_r^3))$ |
| Proposed EQUA | $\mathcal{O}(3N_t^2 N_r + \binom{L}{N_{RF}^r}((N_{RF}^r)^3 + 2\mathrm{N}_s N_{RF}^r \mathrm{N}_r + 2N_s^2 N_t^2) + 3(N_{RF}^r)^3$ |
| CNN | $\mathcal{O}(96\mathrm{k}N_t + 320 N_{RF}^t N_r + 55(N_{RF}^t)^2 N_r^2)$ |

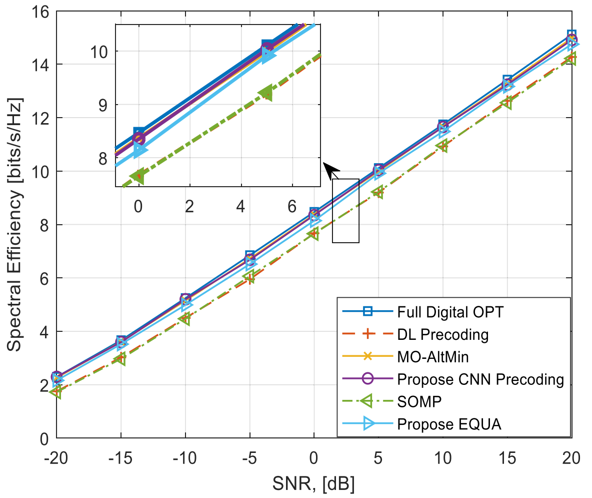



**FIGURE 3.** Achievable spectral efficiency of proposed EQUA hybrid pre-coding and CNN-based pre-coding when $N_t = 16$, $N_{RF} = 4$, $N_r = 16$, perfect CSI, $N_s = 1$ streams) compared to other pre-coding systems.

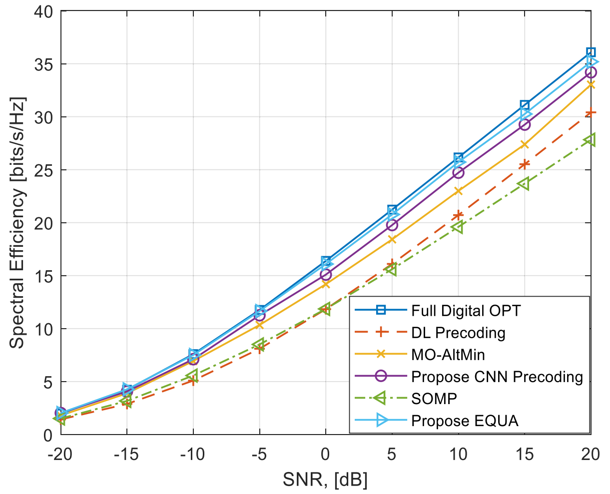



**FIGURE 4.** Achievable spectral efficiency of proposed EQUA hybrid precoding and CNN-based pre-coding(when $N_t = 16$, $N_{RF} = 4$, $N_r = 16$, perfect CSI, $N_s = 3$ streams)compared to other pre-coding systems.

### B. SIMULATION RESULT

In this section, the performance of the proposed algorithm is evaluated by measuring the spectral efficiency of the mm-Wave system with square planar array. The proposed algorithms (Algorithm 1 EQUA and Algorithm 2 CNN precoding) are compared with the fully digital pre-coding (Full Digital OPT), the prior work SOMP in [1], the prior work manifold optimization alternating minimization (MO-AltMin) in [5], and the prior work DNN-based pre-coding (DL Pre-coding)[8]. For each channel matrix realization, the propagation environment is modeled with $\mathrm{N_{cl}} = 4$ and $\mathrm{N_{ray}} = 4$ for each clusters with $\sigma_\theta^2 = 5^\circ$ for all transmit and receive azimuth and elevation angles which are uniform randomly selected from the interval $[-60°, 60°]$ and $[-20°, 20°]$. The complex gain of scattering path in cluster is a complex gaussian distribution with mean 0 and variance 1. The center carrier frequency of the transmission is 28GHz.

Figure 3 presents the spectral efficiency achieved by the proposed EQUA and the CNN-based hybrid pre-coding compared with some existing solutions for hybrid pre-coding in mm-Wave massive MIMO system. We set $\mathrm{N}_t = 16$, $\mathrm{N_{RF}} = 4$, $\mathrm{N}_r = 16$, and $N_s = 1$. As can be seen from the figure, the approach that was proposed by us has similar performance with full digital pre-coding and MO-AltMin. The prior DL pre-coding has similar performance with SOMP and there is a small gap between SOMP and full digital precoding.

In figure 4, we set $\mathrm{N}_t = 16$, $\mathrm{N_{RF}} = 4$, $\mathrm{N}_r = 16$, $N_s = 3$ streams, and use perfect CSI. It clearly shows that the proposed scheme EQUA and the CNN-based pre-coding can achieve a considerably higher spectral efficiency than the prior hybrid pre-coding and the gap between the proposed CNN-based hybrid pre-coding with prior DL pre-coding and SOMP is obvious.

The system configuration difference in figure 3 and figure 4 is the number of sending data streams. Figure 3 represents the system sends single data stream, figure 4 system sends multiple data stream. Through the comparison of the two figures, it is obvious that after adding data streams, the spectral efficiency of all pre-coding schemes increased significantly. And the performance of prior pre-coding schemes has been reduced compared with full digital optimal pre-coding, while the proposed scheme is still very close to the result of full-digital pre-coding. This indicates that the proposed scheme has excellent performance when dealing with multiple data streams.

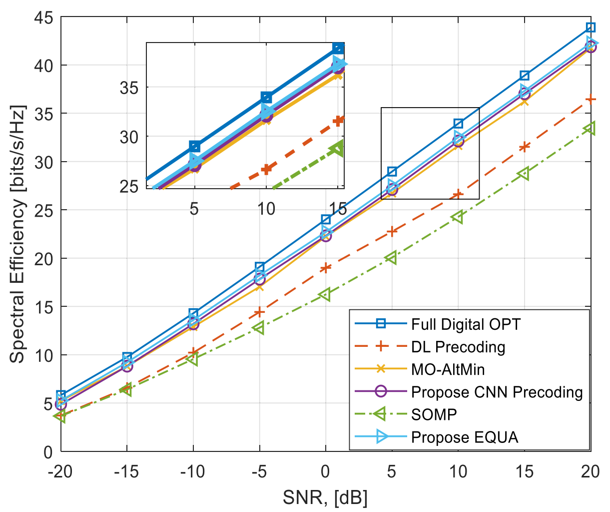



**FIGURE 5.** Achievable spectral efficiency of proposed EQUA hybrid pre-coding and CNN-based pre-coding (when $N_t = 36$, $N_{RF} = 4$, $N_r = 36$, perfect CSI, $N_s = 3$ streams) compared to other pre-coding systems.

Figure 5 shows the spectral efficiency achieved in a $36 \times 36$ system with square planar arrays at both transmitter and receiver. For the proposed pre-coding strategy, both transmitter and receiver are assumed to have four transceiver chains and transmit $N$s $= 3$ streams, perfect channel knowledge is available to the transmitter. Figure 5 show that the proposed hybrid precoding achieves spectral efficiency which close to

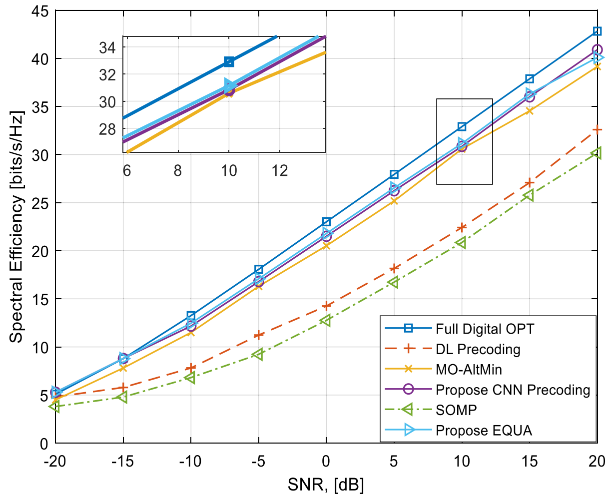



**FIGURE 6.** Achievable spectral efficiency of proposed EQUA hybrid precoding and CNN-based pre-coding (when $N_t = 36$, $N_{RF} = 4$, $N_r = 36$, imperfect CSI, $N_s = 3$ streams))compared to other pre-coding systems.

optimal full digital precoding. This implies that the proposed strategy can very approximate the scheme that decomposing channel matrix to obtain the channel's dominant singular vectors as a pre-coder. Compared with the results in figure 4, the spectral efficiency is also significantly improved when increasing the number of antenna.

Figure 6 illustrates the SE performance achieved by mm-Wave system with 36 receive antennas and 36 transmit antennas, as well as 4 RF chains at the transmitter and transmit $Ns = 3$ streams. Different from previous simulations, white gaussian noise is added to the channel state information matrix in this simulation. Compared the results in figure 5 and figure 6, we find that the spectral efficiency of all pre-coding schemes decreases when imperfect channel state information was used. Especially, the performance of the SOMP hybrid pre-coding and DL Pre-coding significantly decreased. However, the proposed algorithm still has good performance under imperfect channel state information.

## V. CONCLUSION

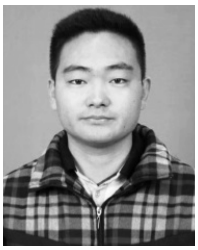
This paper presents two approaches to design hybrid pre-coding in mmWave Massive MIMO systems. Results show that the performance of the proposed equivalent channel pre-coding algorithm and the CNN-based precoding algorithm is much closer to the all-digital precoding algorithm, compared to the optimal unconstrained precoding algorithm and some existing schemes. Meanwhile, the proposed scheme achieves a superior performance as SOMP. SOMP performs poor due to the fact of not matching the set of array responses from the dictionary which maximizes spectrum efficiency. MO-AltMin performs sufficiently well, while EQUA performs better than it. Furthermore, the performance improves more apparently between the proposed CNN-based precoding and the prior DNN-based strategy. By using convolutional neural network, the complexity of calculating combiner is greatly reduced, not only the time is saved but also the system performance is improved.

To sum up, the performance of the proposed EQUA schemes has little gap compared with full digital precoding and the computational complexity decreased significantly. And it can be concluded that the proposed CNN exhibits much stronger robustness to imperfect CSI than the traditional algorithms and it also significantly decrease computational complexity. The less accurate the channel estimate is, the larger the performance gap is.

## REFERENCES

[1] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[2] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics 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] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.
[6] 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.
[7] S. Dorner, S. Cammerer, J. Hoydis, and S. T. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.
[8] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.
[9] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," 2019, *arXiv:1905.13212*. [Online]. Available: https://arxiv.org/abs/1905.13212
[10] 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.
[11] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," Apr. 2018, *arXiv:1804.10334*. [Online]. Available: https://arxiv.org/abs/1804.10334
[12] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[13] A. M. Elbir, K. V. Mishra, and Y. C. Eldar, "Cognitive radar antenna selection via deep learning," *IET Radar, Sonar Navigat.*, vol. 13, no. 6, pp. 871–880, Jun. 2019.
[14] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 340–352, May 2018.
[15] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2278–2281, Nov. 2018.
[16] 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.
[17] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, Jul. 2019.
[18] P. Dong, H. Zhang, G. Y. Li, N. Naderializadeh, and I. S. Gaspar, "Deep CNN for wideband mmwave massive Mimo channel estimation using frequency correlation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019.

[19] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.
[20] P. Comon and G. Golub, ''Tracking a few extreme singular values and vectors in signal processing,'' *Proc. IEEE*, vol. 78, no. 8, pp. 1327–1343, 1990.
[21] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, ''State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems,'' *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2432–2455, 4th Quart., 2017.
[22] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, ''Unsupervised learning-based fast beamforming design for downlink MIMO,'' *IEEE Access*, vol. 7, pp. 7599–7605, 2019.



**XIAOLONG BAO** received the B.S. degree in communication engineering from the Kunming University of Science and Technology, Kunming, China, in 2018. He is currently pursuing the master's degree in microelectronics and communication engineering with Chongqing University, China. His current research interests include massive MIMO and mmWave communications.

**WENJIANG FENG** received the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 2000. He is currently a Professor with the College of Microelectronics and Communication Engineering, Chongqing University. His research interests include all aspects of MIMO communications, including limited feedback techniques, antenna design, interference management and full-duplex communication, cognitive radio, special mobile communication system, and emergency communication. He is a Peer Review Expert of the Natural Science Foundation of China and is a Senior Member of the China Institute of Communications. He also serves as an Editorial Board Member of Data Communication, China.

**JIALI ZHENG** received the B.S. degree in communication engineering from Chongqing University, Chongqing, China, in 2018, where she is currently pursuing the master's degree in microelectronics and communication engineering. Her current research interests include massive MIMO and unmanned aerial vehicle communications.

**JINGFU LI** received the B.S. degree in communication engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016, where he is currently pursuing the Ph.D. degree in microelectronics and communication engineering. His current research interests include massive MIMO and interference alignment.