Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 1, October 2024, pp. 570~582
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i1.pp570-582

Deep learning based hybrid precoder for optimal power allocation to improve the performance of massive MIMO

Shilpa Bhairanatti, Rubini P.
Computer Science Engineering, School of Engineering and Technology, CMR University, Bangalore, India

Article Info

Article history:

Received Oct 16, 2023
Revised Jun 5, 2024
Accepted Jun 24, 2024

Keywords:

Deep learning
Hybrid precoder
Massive MIMO
Power allocation
Spectral efficiency
Sum rate

ABSTRACT

Hybrid precoding is a significant procedure for decreasing the hardware complexity and power usage in massive multiple-input multiple-output (MIMO) systems. However, the effectiveness of hybrid precoding is highly dependent on precise channel state info and designing of the beamforming matrix. In recent years, deep learning-based approaches have emerged as a promising solution to address these challenges. This research focuses on improving the performance of massive MIMO systems. However, several methods have been introduced to develop the hybrid precoding model, but these models suffer from several issues such as complexity, interference and quantization error. Currently, deep learning-based methods have gained huge attention in this domain where these methods learn from the data and try to overcome the challenges. Here, a deep learning-based model is presented where our main aim is to develop a hybrid precoder along with the deep learning-based optimal power allocation model. Therefore, the proposed model overcomes the issue of hybrid precoding and power distribution resulting in improving the overall performance of massive MIMO systems on the parameters such as spectral efficiency (SE) and the sum rate.

Corresponding Author:

Shilpa Bhairanatti
Computer Science Engineering, School of Engineering and Technology, CMR University
Bangalore, India
Email: shilpasky2121@gmail.com

1. INTRODUCTION

Recently, we have noticed a tremendous growth in wireless communication technology. This technological growth in wireless communication has increased the demand for high-speed communication and immediate access to several multimedia communications [1], [2]. A few examples of such applications are smart cities, autonomous vehicles, and smart healthcare. Moreover, a study has reported that by the end of year 2025, there will be over 39 billion active wireless devices worldwide [3]. Mobile networking play important role in wireless communication technology because it facilitates the voice and data networking facilities simultaneously with the help of radio transmission. These devices are widely used and the demand of these devices is still increasing. The voice communication over wireless networks started with the circuit switching methods and later it was upgraded to packet-switching for voice and data communication. Currently, the mobile communication systems are based on the packet switching. The communication spectrum has expanded from 1G to 4G and 5G [4]. This increased demand for wireless communication traffic would significantly affect forth coming mobile network system designs. The architectural designs must focus on several aspects [5] of communication such as:

– High traffic volume: the increased demand leads to an increase the data traffic. In the future, data traffic will be increased by 1000x of current data traffic [6].

– Increased small cell and indoor traffic: as mentioned before, that the number of active devices is increasing and the major portion of mobile data traffic is generated from indoor communication. However, it is expected to exceed this figure.
– Seamless connectivity: as the number of devices and communication demand increases, it requires better connectivity and it also requires machine-to-machine communication support.
– Optimized energy consumption: these networks must support the green network paradigm to reduce power consumption.

In the traditional cellular infrastructures e.g., GSM, the baseband and digital units are collocated at the bottom of the antenna tower with analog radio units. These units are connected through the thick low-loss coaxial cables to the antennas and positioned at the maximum height of the tower comprising amplifier units [7]. It helps to reduce the power loss by making use of coaxial cable. Distributed networks, which were previously only employed in 1G and 2G systems, are now used in 3G and 4G systems. In decentralized networks, the radio unit is separated from the baseband unit and installed above the aerial tower together with all of the transmitter and receiver modules accompanied by amplifiers. Fibre front hauls delivering signal data to the tower replace long coaxial cables that tend to drop data packets while having high frequency communication. The radio head positioned close to the antennas, allows this architecture to achieve a significant link gain. Moreover, in 4.5G and 5G systems, the baseband units are centralized to improve communication performance [8].

All these advancements are achieved with the help of advanced communication technologies. multiple-input multiple-output (MIMO) systems are recognized as a global solution to improve the efficacy of cellular communication. In MIMO systems, user terminals are equipped with multiple antennas [9]. These systems take advantage of diversity to provide improved communication link reliability. Moreover, these systems also encompass multiplexing gains by permitting simultaneous processing of data streaming by multiple users using the shared resource which ultimately leads to increased spectral efficiency (SE) [10]. Therefore, these systems are widely adopted in 3G and 4G communication systems due to their several applications. For example, in LTE-A, the spatial multiplexing is allowed up to eight layers for both frequency division multiplexing (FDD) and time division multiplexing (TDD) [11].

Therefore, we have noticed a noteworthy growth in the use of MIMO technology as it improves wireless link transmission capacity and reliability. In MIMO systems, several antennas are deployed at the transmitter and receiver side. MIMO antennas are crucial for boosting connection stability and channel capacity. Generally, two or four antennas are typically used in a single physical package in standard MIMO networks. Massive MIMO, conversely, is a MIMO system with a very large number of antennas. For the fourth generation (4G) mobile communications, a 2×2 MIMO system has been successfully implemented, and it is anticipated that the huge MIMO system having a high number of MIMO antennas will be extremely promising for 5G wireless communications [12]. Moreover, the massive MIMO systems are further resistant to jamming and signal interference. Thus, the MIMO and massive MIMO technologies play an important role in this domain of wireless cellular communication. Table 1 shows a brief comparison between these two techniques.

Due to these advancements in massive MIMO technology, these systems are widely adopted in the fifth generation (5G) and beyond 5G communication standards. Currently, various kinds of research have been conducted to achieve better SE and link reliability in multi-antenna systems. For instance, ultra-dense networks, mmWave, and spectrum sharing. The mmWave technique uses 30-300 GHz band frequency and it supports larger bandwidth when compared with the conventional 6 GHz massive MIMO system [13]. However, it suffers from excessive path loss. Similarly, massive M2M systems have also deployed recently. These systems use different types of detection schemes in which the antennas tend to generate interfering signals which lead to dilution of the original signal by noise and interference.

Table 1. Brief comparison between MIMO and massive MIMO techniques

	MIMO	Massive MIMO
Antenna count	≤8	≤100
Channel matrix	Low demand	High demand
Channel capacity	Low	High
Diversity gain	Low	High
Link stability	Low	High
Resistant to noise	Low	High
Antenna correlation	Low	High
Bit error rate	High	Low
Throughput	Low	High
Cost	Low	High

As discussed before, the 5G networks boost the SE which can help to meet the increasing spectrum demand [14]. Moreover, the limited bandwidth and increased demand of system capacity forces network service providers to adopt the physical layer related solutions whereas the performance of these systems can be improved by using massive MIMO systems. In massive MIMO systems, base station (BS) serves single or large no. of antennas with matching frequency bands. In massive MIMO systems, the user equipment (UE) has own processing unit to sense the data. However, the smaller processing ability of UEs is not suitable for large antennas which degrade the system performance. Therefore, researchers have introduced a precoding methodology to offer a large processing ability for detection. Despite the BS's powerful and high processing capacity, finding low complexity precoding techniques is still remains an indispensable task in advanced cellular communication systems. Precoding is a process to combine the input signals in a predetermined manner and delivering them to the various antenna components in the proper ratio. These precoding methods are essential for the development of the enormous MIMO system in the next 5G technology. Moreover, the precoding methods are helpful in achieving the increased gain by incorporating mmWave, massive MIMO and precoding techniques. In third domain, codebook design [15] and compressed sensing [16] have been considered as promising techniques to design precoder.

Recent studies have reported that fully digital precoding methods can help to achieve the hypothetical channel capacity by adjusting the precoding coefficients. However, increased no. of RF chains as well as number of the active transmitter, modules lead to increase in implementation cost. On the other hand, the fully analog beamforming model which minimizes the implementation complexity with a minimum number of RF chains, but it also reduces the gain of beamforming due to the limited resolution of phase shifters. To mitigate these issues of analog and digital precoding, researchers have introduced hybrid precoding as an efficient solution [17], [18]. Similarly, the deep learning based methods also have gained huge attention in this domain [19].

Hybrid precoding is a technique used in modern wireless communication systems, such as millimetre-wave (mmWave) and massive MIMO systems, to overcome the high cost and power ingestion linked with usage of huge number of radio frequency (RF) chains. In hybrid precoding, the transmit signal is divided into two stages: a baseband precoder and a RF precoder. The baseband precoder performs digital signal processing on the transmit signal in the baseband, such as channel equalization, spatial multiplexing, and data modulation. The RF precoder then maps the processed baseband signal onto the RF domain using a small number of RF chains. The advantage of hybrid precoding is that it enables efficient usage of the available RF chains while still achieving decent efficiency in terms of data rate and power consumption. This is particularly important in mmWave and massive MIMO systems, where a large number of antennas are used to increase the capacity of the system.

Hybrid precoding is a complex technique that involves careful design of the baseband and RF precoders, as well as channel estimation and optimization algorithms. It is an active area of research in the field of wireless communication, and many advanced hybrid precoding techniques have been proposed in recent years to further improve the performance of modern wireless communication systems. Despite of several advancements, precoding methods suffer from various challenges and problems which are as follows:

– Computational complexity: hybrid precoding uses combination of digital and analog precoding with complex algorithms and sophisticated hardware architectures which leads to increase the complexity. Moreover, varying channel conditions also increase the additional complexity.
– The hybrid precoding methods require analog components such as phase shifters and RF chains and implementation of these devices at high frequency bands can be challenging due to hardware constraints such as cost, power consumption and limited phase resolution.
– The performance of hybrid precoding model relies on channel state information (CSI) however, achieving perfect CSI remains a challenging task.

Therefore, this work focuses mainly on improvising the overall performance of the massive MIMO communication by incorporating improved hybrid precoding. The proposed hybrid precoding model is established on the notion of a deep learning scheme where we use deep learning based model for beamformer and combiner prediction along with that optimal power allocation task is also accomplished by using DL based model. The complete process is divided into several stages, where first of all this model employs RF chain and produced outcome is processed through the two consecutive layers of 1D-convolution. This convoluted output then processed through the vectorization module then deep learning operations are performed and finally the output is processed through the RF chains and fed to the receiving baseband. The main contributions of this work are listed below:

– A novel hybrid precoding approach is presented where both analog and digital precoding schemes are introduced.

- The complete precoding model is divided into two phases as BB stage and RF stage for downlink massive MIMO systems.
- Further, a method is presented to compute the phase vector which is used to compute the received signal vector and later achievable rate computation is used to derive the SINR model.
- Finally, a deep learning based model is introduced, which helps to obtain the optimal power allocation solution.

Rest of the article has the following sections: section 2 presents the literature review on massive MIMO systems where we explored the current trends and advancements in the field of massive MIMO. Section 3 presents the proposed deep learning based solution for hybrid precoding, where the system architecture is briefly explained, section 4 presents the outcome of proposed approach and comparative analysis. Finally, section 5 presents the concluding remarks of the work.

2. LITERATURE SURVEY

This section presents a brief discussion about existing precoding schemes, which includes analog, digital and hybrid precoding techniques have gained huge attention in mmWave massive MIMO, and 5G communication standards because of its nature to reduce the RF chains while achieving the desired performance of fully digital architecture. Several precoding algorithms have been presented such as codebook-based hybrid precoding, simultaneous orthogonal matching pursuit, and manifold optimization based alternate minimization. In these algorithms, it becomes an important step for BS to acquire the appropriate downlink CSI.

On the other hand, it becomes a challenging task for BS to obtain the downlink CSI due to uplink-downlink reciprocity. At this stage, the user UEs estimates the downlink CSI and it is reported to the BS via feedback links. Several feedback mechanisms are presented in literature such as codebook design [15] and compressive sensing [16]. Similarly, some researchers have focused on jointly optimizing the CSI feedback and hybrid precoding [17], [18]. Li *et al.* [17] presented a two-stage beamformer for downlink massive MIMO system in FDD mode where both links are furnished with hybrid beamforming structures. Further, users are grouped based on channel statistics and analog beamforming where UEs feedback its intra-group effective channel. It aids in reducing the overall cost of the CSI.

Wang *et al.* [18] reported that the hybrid precoding methods play important role in improving the system capacity, reducing the hardware cost and minimizing the power consumption, therefore, it has become a key technology in the 5G and 6G millimetre-wave communication systems. Though, designing a hybrid precoder with minimal computational complexity is considered as a challenging task. To overcome this issue, authors introduced SVD based hybrid precoding system for multi-user massive MIMO system. This model focused on developing the joint analog precoder and combiner while considering the single-user massive MIMO systems. The main aim of this model was to obtain the maximized square of sum eigenvalues for an equivalent channel. Later, a corresponding equivalent channel is developed. The digital precoding method is realized with the help of SVD technique. However, these methods lead to increase the implementation cost and computational overhead to the system for large no. of antennas and users.

Recently, deep learning-based schemes are widely adopted to develop the CSI feedback and hybrid precoding methods. Wen *et al.* [19] reported in their work that, for increasing the gains of massive MIMO systems, the downlink CSI must get transmitted to BS through the feedback links. However, this transmission is hindered due to excessive feedback overhead. For overcoming this problem, researchers adopted deep learning approach and developed CsiNet which is a channel sensing and recovery mechanism which learns channel structure effectively. During learning phase, it learns transformation from CSI to codewords and inverse transformation from codewords to CSI.

Wang *et al.* [20] also reported the importance of CSI feedback in massive MIMO to achieve the desired gain performance in FDD mode. However, increased no. of antennas leads to an increase in feedback overhead. Therefore, authors presented a deep learning-based CSI feedback scheme which is based on the combination of CsiNet and long short-term memory (LSTM). This model helps to improve the recovery quality and enhances the trade-off between compression ratio and complexity. It is obtained because the CsiNet model directly learns from the spatial structures and time–varying massive MIMO channels. Sun *et al.* [21] introduced a deep learning model to realize CSI and hybrid precoding for mmWave communication systems in FDD mode. Authors reported that the traditional algorithms consider the CSI and precoding as two distinct problems, therefore, the deep learning model is designed which bypasses channel reconstruction phase and integrated with the hybrid precoder and combiner by taking the feedback codewords into consideration.

Massive MIMO systems benefit greatly from hybrid precoding's ability to reduce hardware complexity and power consumption, but it also comes with a number of issues that need to be resolved. Some of the issues are listed below:

- Quantization error: quantization errors may be introduced by the analog-to-digital converter (ADC) used in RF chains to digitise the signal. The effectiveness of hybrid precoding systems can be greatly lowered by these problems.
- Limited feedback: accurate CSI is necessary for hybrid precoding to work well. Yet, because to the constrained feedback resources in actual systems, acquiring sufficient user UE feedback is necessary to produce correct CSI.
- Interference: hybrid precoding is extremely susceptible to interference, particularly when there are several antennae. Performance of the system may suffer significantly as a result, particularly if the interference is co-channel.
- Hardware constraints: the efficiency of hybrid precoding may be constrained by hardware limitations related to RF chains and subarrays. The quantity of RF chains available, incidentally, may restrict the number of subarrays, or the RF chains' power consumption may make their deployment impractical.

As a summary of review the recent methods of hybrid precoding in this domain of massive MIMO. However, despite of achieving perfect CSI, designing the precoder still remains a challenging task because developing the analog precoder via phase shifter array leads to impose the non-convex constraint on the designing parameters. Due to these constraints, the computational complexity of these systems increases. Moreover, this type of hybrid architectures affects the uplink channel sensing capability. Some recent hybrid precoding methods are still relying on the traditional communication methodology where the precoding process is decomposed into two separate components as (a) channel estimation and channel sensing and (b) designing the downlink precoding. During the channel estimation phase, these methods exploit the spatial correlation of mmWave channel. In recent years, deep learning-based hybrid precoding has been proposed as a technique to cut down the complexity and enhance the efficacy of massive MIMO systems. The proposed algorithms use various types of neural networks, including deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), to learn the association link between the CSI and the hybrid precoding matrix. The simulation results demonstrate that the deep learning-based algorithms achieve better outcomes than conventional algorithms, including OMP, CGD, ADMM, ZF, MMSE, sparse precoding, and random matrix-based algorithms. However, the design and training of the deep neural networks require significant computational resources and may require a large amount of training data. Therefore, further research is required to develop efficient and practical deep learning-based hybrid precoding algorithms for massive MIMO systems.

3. PROPOSED DEEP LEARNING BASED HYBRID PRECODING FOR MASSIVE MIMO SYSTEMS

Previous sections have described the significance of hybrid precoding in massive MIMO systems. The current technological advancements have identified deep learning based models as promising solutions to exploit the spatial multiplexing gain and beamforming gain. Moreover, it requires less number of RF chains and this reduced RF chain requirement is beneficial in mitigating the complexities. Therefore, we adopt the deep learning based model and present a hybrid precoding model to solve the challenges faced in the massive MIMO systems.

Hybrid precoding has been proposed as a solution to decrease the cost and complexity of mmWave communication systems. Hybrid precoding uses fewer RF chains than full-digital precoding, which reduces the hardware complexity and power depletion. Hybrid precoding also provides the flexibility to manipulate the beamforming and channel characteristics in real time. Hybrid precoding is composed of two stages: analog precoding and digital precoding. Analog precoding is performed in the RF front-end and involves steering the signal in the desired direction. Analog precoding is performed using a phase shifter network, which manipulates the amplitude and phase of the RF signal. The phase shifter network is designed to have a low loss and high resolution.

Digital precoding is performed in the baseband processor and involves processing the signal using DSP techniques. Digital precoding is used to further manipulate the signal to compensate for the channel distortion and improve the signal-to-noise ratio (SNR). Digital precoding is performed using a matrix that maps the signal to the desired directions. Hybrid precoding has several challenges, including the design of the phase shifter network, the optimization of the analog and digital precoding matrices, and the trade-off between performance and complexity.

3.1. Network model and problem formulation

This work mainly considers a downlink massive MIMO system where BS has $\mathcal{N}_t$ transmit antennas and, similarly, $\mathcal{N}_r$ receiver antennas are serving total K users. This user UEs are arranged in a clustered group. For this configuration, the received signal at k^{th} user $y_k \in \mathbb{C}^{N_r \times 1}$can be estimated according to (1):

$$y_k = H_k V_k s_k + \sum_{m=1, m\neq k}^{K} H_k V_m s_m + n_k, \forall_k \in \mathcal{K} \quad (1)$$

Where $H_k \in \mathbb{C}^{N_r \times N_t}$ represents the channel matrix between user k and BS, $V_k \in \mathbb{C}^{N_t \times d_k}$ represents the precoding matrix for the user k, d is the data stream. During this transmission, there is a probability that nose may get added to the original signal therefore we consider additive white Gaussian noise. The Gaussian noise distribution is $\mathcal{CN}(0, \sigma^2 I)$, the data transmitted by user k is denoted as $s_k \in \mathbb{C}^{d_k \times 1}$ which also satisfied the condition $\mathbb{E}[s_k s_k^H] = 1$. We assume that the data transmission between two users in independent and $\mathcal{K}$ represents the set of users as $\mathcal{K} = \{1,2,,3 \dots, K\}$. Figure 1 depicts the massive MIMO architecture with hybrid precoding.

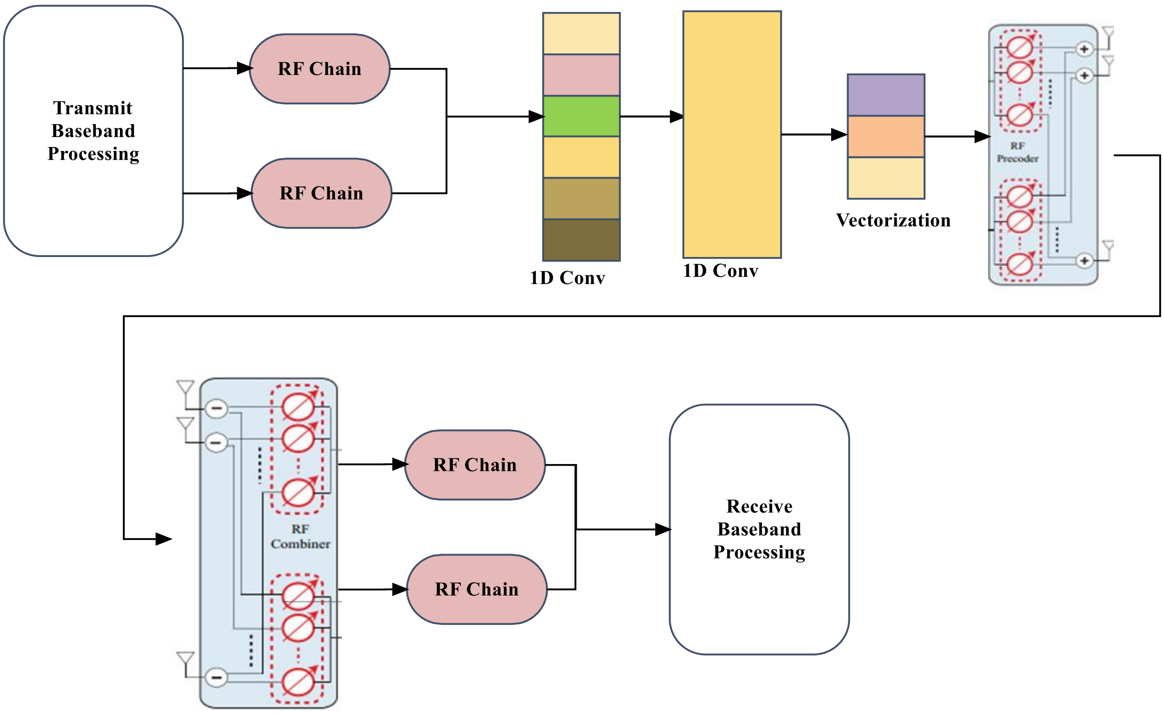

Figure 1. Massive MIMO with hybrid precoding

The hybrid precoding architecture consists of two stages: BB stage and RF stage which are connected through the available N_{RF} number of RF chains. This interconnectivity helps to reduce the hardware complexities. In this work, the first phase includes the development of RF beamformer as $F \in \mathbb{C}^{M \times N_{RF}}$ by employing the low cost phase shifters. Similarly, the second stage includes the development of digital BB precoder as $B = [b_1, b_2, \dots, b_K] \in \mathbb{C}^{N_{RF} \times K}$ where $b_k \in \mathbb{C}^{N_{RF}}$ represents the precoding vector of BB precoder, and it also considers the power allocation matrix $P = diag(\sqrt{p1}, \dots, \sqrt{pK}) \in \mathbb{R}^{K \times K}$ where $p_k$denotes the allocated power for k^{th} UE. Considering all these components, below given (2) can be used to the downlink vector can be expressed as:

$$s = FBPd \in \mathbb{C}^M \quad (2)$$

Where M is the antenna array and d is the data vector given as $d = [d_1, \dots, d_K] \in \mathbb{C}^K$. Similarly, the channel vector for the k^{th} user can be computed based on (3):

$$h_k^T = \sum_{l=1}^{Q} \mathcal{T}_{k_l}^{\eta} z_{k_l} \phi^T \left(\gamma_{x,k_l}, \gamma_{y,k_l}\right) = z_k^T \phi_k \in \mathbb{C}^M \quad (3)$$

Where Q denotes the total number of paths, $\mathcal{T}_{k_l}$ distance of l^{th} path, z_{k_l} is the complex path gain of path l which is distributed as $z_{k_l} \sim \mathcal{CN}(0, 1/Q)$, η represents the path loss exponent, $\phi(.,.) \in \mathbb{C}^M$ represents the phase response vector, γ_{x,k_l} represents the coefficients which reflects elevation AoD denoted as $\gamma_{x,k_l} = \sin(\theta_{k_l})\cos(\psi_{k_l})$, similarly γ_{y,k_l} represents the azimuth AoD as $\gamma_{x,k_l} = \sin(\theta_{k_l})\cos(\psi_{k_l})$ for the corresponding path. Based on these parameters, below given (4) can be used to compute the phase vector as follows:

$$\phi(\gamma_x, \gamma_y) = \left[1, e^{-j2\pi d\gamma_x}, \dots, e^{-j2\pi d(M_x-1)\gamma_x}\right]^T \otimes \left[1, e^{-j2\pi d\gamma_y}, \dots, e^{-j2\pi d(M_y-1)\gamma_y}\right]^T \in \mathbb{C}^M \quad (4)$$

Where d represents the antenna spacing. Finally, we can use (5) to compute the received signal vector at k^{th}UEwhich can be expressed as:

$$\begin{aligned} r_k &= h_k^T s + n_k = h_k^T FBPd + n_k \\ &\sqrt{p_k} h_k^T F b_k d_k + \textstyle\sum_{t \neq k}^{K} \sqrt{p_t}\, h_k^T F b_t d_t + n_k \end{aligned} \quad (5)$$

Based on these assumptions, our main aim in this work is to maximize the sum rate while considering the transmit power constraint, thus, the overall sum rate problem can be computed by using (6) as follows:

$$\begin{aligned} &P_1: \max_{\{V_k\}} \textstyle\sum_{k=1}^{K} \alpha_k R_k \\ &s.t. \textstyle\sum_{k=1}^{K} Tr(V_k V_k^H) \leq P_T \quad (6) \end{aligned}$$

Where P_T denotes the constraints for transmit power,α_k is the scalar value which is used to represent the priority of user k. Based on aforementioned parameters, the achievable rate can be computed by following the (7). Thus the achievable rate for k^{th} user is denoted by R_k and expressed as:

$$R_k = \log\det\left(I + H_k V_k V_k^H H_k^h \left(\textstyle\sum_{m \neq k} H_k V_m V_m^H H_k^H + \sigma^2 I\right)^{-1}\right) \quad (7)$$

Further, we also incorporate SINR of k^{th}UE. In (8) uses power, beanforemer and precoder parameters to estimate the SINR value which can be represented as follows:

$$SINR_k(F, B, P) = \frac{p_k \left|h_k^T F b_k\right|^2}{\sum_{t \neq k}^{K} p_t \left|h_k^T F b_t\right|^2 + \sigma_n^2} \quad (8)$$

The sum rate capacity corresponding to the SINR can be computed as $R_{sum\,rate} = \mathbb{E}\{\sum_{k=1}^{K} \log_2[1 + SINR_k(F, B, P)]\}$. Relying on these expressions, in (9) presents the sum rate problem corresponding to SINR, which is given as:

$$\begin{aligned} &\max_{F,B,P} \sum_{k=1}^{K} \log_2\left(1 + \frac{p_k |h_k^T F b_k|^2}{\sum_{t \neq k}^{K} p_t |h_k^T F b_t|^2 + \sigma_n^2}\right) \\ &\quad s.t. \quad C_1: \mathbb{E}\{\|s\|_2^2\} = \textstyle\sum_{k=1}^{K} p_k b_k^h F^H F b_k \leq P_T \\ &\quad\quad\quad C_2: p_k \geq 0, \forall k \\ &\quad\quad\quad C_3: \left|[F]_{i,j}\right| = \frac{1}{\sqrt{M}}, \forall_{i,j} \end{aligned} \quad (9)$$

This problem is refereed as a non-convex optimization problem because the distributed power is associated with each other and constrains to RF beamformer. Therefore, it leads to increase the high computational costs and delay which raises the difficulty to implement these modules for real-time systems.

3.2. Deep learning model

The traditional models accomplish the hybrid precoding design in three stages which include leveraging the sparsity of mmWave channels, channel reconstruction, and using this constructed channel to design hybrid coding matrices. However, these methods do not use the information from prior channel observation to reduce the training complexity. In this work, we present a deep learning based model adopted from [22] which consists of two main components, a channel encoder and a precoder. The channel encoder module considers the channel vectors as input to the model and passes it through the 1D convolution operation which performs the Kronecker product operation. The output of this module is passed to the precoder module where fully connected layer and output layers are employed for prediction.

In the encoder module, we adopt the neural network model to employ the compressive sensing model. Let P is given as $N_t \times M_t$ and Q in denoted as $N_r \times M_r$, represents the channel measurement matrices. These matrices are used in channel sensing H. In a condition where pilot symbols are equal to 1 then (10) can be used to model the received datawhichcan be expressed as:

$$Y = \sqrt{P_T} Q^{\mathrm{H}} H P + Q^{\mathrm{H}} V \tag{10}$$

Where $[V]_{m,n} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2)$ represents the noise parameter. Further, this measurement matrix is vectorised and y can be written as mentioned in (11):

$$y = \sqrt{P_T}(P^T \otimes Q^{\mathrm{H}})h + v_q \tag{11}$$

Where $y = vec\ (Y)$, $v = vec(Q^{\mathrm{H}} V\ )$and $h = vec(H)$

In this encoder architecture, the convolution layer uses M_r kernels as filter where each kernel size and stride are given as N_r. The kernel weights are used for representing the entries of receive measurement vector in Q. The output of this layer is obtained as feature map and transformed into a vector form. Further, this vectorised data is fed to the second convolution layer. The second layer also consists of M_t number of kernels to realize the measurement matrix P.

The final received channel measurement vector is y where we use the deep neural network model to directly map the received vector and beamforming or combining vector. Here, main aim of deep learning module is to predict the beamforming vector F_{RF} and combining vector W_{RF}. once the RF beamformer and combiner vectors are generated then the low-dimensional effective channel can be constructed as $W_{RF}^H H F_{RF}$ which is used to construct the baseband precoder and combiner. However, these beamforming and combining vectors are selected from the codebooks denoted as $\mathcal{F}$ and $\mathcal{W}$ and the prediction of beamformer and combiner is formulated as multi-label classification problem.

In prediction phase, the precoder network plays important role. This network consists of fully connected layer and two output layers and the input is received from the channel encoder module. Further, the fully-connected layer is connected to the rectified linear unit (ReLu) activation function which is further connected to the batch normalization layer. For ReLU function, the activation function is defined as $ReLU(a) = \max(0, a)$for any given argument$a$. Further, the output of the network can be denoted as characterized in (12) given:

$$z = f(v, w) = f^{(n-1)}\left(f^{(n-2)}\left(\dots f^{1}(v)\right)\right) \tag{12}$$

Where v denotes the input data, n denotes the number of layers and w denotes the weights of the neural network. Finally, this layer has two output layers which are used for predicting the beamforming vector and combiner vectors, respectively. The dimension of the obtained vector is equal to the combining vector.

Along with this, we also incorporate the power allocation module for efficient power distribution. As discussed before, we have a channel matrix H and baseband precoder B, which are processed through the feature scaling and vectorization operations. These parameters can be used to represent the feature vector and the final vector can be obtained according to (13). Thus, the final vector is expressed as:

$$x_0 = \begin{bmatrix} \alpha_1 x_{\overline{h_1}} \\ \vdots \\ \alpha_1 x_{\overline{h_K}} \\ \alpha_2 x_{b_1} \\ \vdots \\ \alpha_2 x_{b_K} \\ \alpha_3 x_{BB} \end{bmatrix} \tag{13}$$

With the help of this feature vector, we compute the output of the i^{th} hidden layer which is computed as $x_i = f_r(W_{i-1} x_{i-1} + b_{i-1}) \in \mathbb{R}^{L_i}$ where W is the weight matrix given as $W_{i-1} \in \mathbb{R}^{L_i \times L_{i-1}}$ and $b$is the bias vector. In (8) uses the ratio of power for varied number of users to obtain the maximum absolute optimal power which can be expressed as (14).

$$\bar{p}_k = \frac{p_k^{opt}}{\max\left(p_1^{opt}, \dots, p_K^{opt}\right)} \in [0,1] \tag{14}$$

This power related information is used in (15) which is used to predict the overall power with the help of DNN which can be expressed as:

$$[\hat{p}_1, \hat{p}_2, \dots, \hat{p}_K] = f_\sigma(W_3 x_3 + b_3)$$
$$= f_\sigma(W_3 f_r(W_2 f_r(W_1 f_r(W_0 x_0 + b_0) + b_1) + b_2) + b_3) \tag{15}$$

With the help of this expression and baseband precoder which is expressed as $B = \left[\widetilde{H}^H \widetilde{H} + K \frac{\sigma_n^2}{P_T} I_{N_{RF}}\right]^{-1} H^H \in \mathbb{C}^{N_{RF} \times K}$, we derive the power value which satisfy the power constraint. This is used in (16) to obtain the final power allocation which is characterized as (16).

$$P = \sqrt{\frac{P_T}{\sum_{k=1}^{K} \hat{p}_k b_k^H b_k}}\, diag\left(\sqrt{\hat{p}_1}, \sqrt{\hat{p}_2}, .., \sqrt{\hat{p}_K}\right) \tag{16}$$

According to this approach, the initial signal is produced with all required parameters such as total number of users, noise consideration and varying channel conditions. Later, BB and and RF stages are defined where BB precoder considers the power allocation components for UEs. Similarly, this model computes the final channel vector for the given users. Later it considers power constraints which becomes important aspect for these communication standards because appropriate power allocation is helpful in achieving the sum rate. Algorithm 1 presents the algorithmic overview of proposed approach:

Algorithm 1. Proposed approach

Define system parameters: N_t Number of transmit antennas, N_r Number of receive antennas, N_u - Number of users, N_f - Number of frequency bands
Initialize deep learning model architecture for power allocation
Channel Estimation
Obtain channel state information (CSI):
- $H_{est} \in \mathbb{C}^{N_t \times N_r}$ Estimated channel matrix

RF and Baseband Codebook Design
RF precoders and combiners:
- $F_{RF} \in \mathbb{C}^{N_t \times N_{RF}}$ Analog precoding matrix
- $W_{RF} \in \mathbb{C}^{N_t \times N_{RF}}$ - Analog combining matrix

BB precoders and combiners:
- $F_{BB} \in \mathbb{C}^{N_t \times N_d}$ digital precoding matrix

$W_{BB} \in \mathbb{C}^{N_t \times N_d}$ - digitalcombining matrix
Hybrid Precoding Optimization:
$F_{hybrid} = F_{RF} . F_{BB}$
$W_{hybrid} = W_{RF} . W_{BB}$
Deep Learning-Based Power Allocation:
- Train a DNN model to predict power allocation:
- Input: Features including CSI H_{est}), interference levels, system parameters

Output: Power allocation matrix P_{DL}

4. RESULTS AND DISCUSSION

This section presents the complete experimental analysis of the proposed deep learning based hybrid precoding approach for a massive MIMO system. The obtained performance is compared with the existing mechanisms as mentioned in [23]. The complete simulation parameter set is described in Table 2. According to this experiment, we have considered a total of 16 transmitter and 16 receiver antennas which are part of a downlink massive MIMO communication system.

Table 2. Simulation parameters

Parameter name	Parameter value
Number of antennas	$M = 16 \times 64 = 1024$
Transmit power of BS	20 dBm
BS height	10 m
UE height	1.5m -2.5m
Path Loss exponent	3.76
Channel bandwidth	10 kHz
Number of path	20
Antenna spacing	0.5

Similarly, the deep learning module also required certain configuration during training process. Table 3 shows the configuration parameter for deep learning. Based on these parameters we generate the deep learning data as 〖10〗^5 to accomplish the supervised learning process. In each round, the channel vector is presented as mentioned in (3) which is generated for each UE by randomly varying the path gain, AoD parameters and UE locations with respect to BS. In this offline learning process, the dataset is split into a ratio of 80%-20% where 80% of data is considered in the training process and 20% of data is used for the validation process. In this experiment, we have divided the complete experiment into two cases and considered variable parameters presented in Table 4.

Table 3. Configuration parameters for deep learning architecture

Parameter name	Value
Size 1st hidden layer	1024
Size of 2nd hidden layer	512
Size of 3rd hidden layer	256
Epoch	25
Batch size	32
Learning rate	0.001
Optimizer	ADAM

Table 4. Variable parameters for two cases

Parameter name	Values for case 1	Values for case 2
N_t	16	64
N_r	2	4
K	4~8	8~16
d_k	1	2

First of all, we measure the performance of proposed approach in terms of sum rate for varied number of users for two cases of d_k (i.e. with varied d_k and same d_k) by considering weighted minimum mean square error (WMMSE), and low complexity precoding (LCP) as mentioned in [23]. This precoding matrix is designed based on the noisy channel $\widehat{H} = H + n$. The experiment shows that the existing WMMSE, eigen-based zero-forcing (EZF), and deep learning based LCP approach face difficulty due to their dependency of the accurate channel estimation whereas the proposed model uses deep learning based optimal power allocation model to improve the overall performance of the system. Moreover, the extensive training process of the proposed model helps to adapt these noisy parameters. Figure 2 shows the obtained sum rate performance for varied number of users. As per the experiment, the average performance is obtained as 47.8, 42.4, 44.5, and 43.2 bits/s/Hz by using the proposed approach, deep LCP with zero filling, deep LCP without zero filling, and WMMSE, respectively.

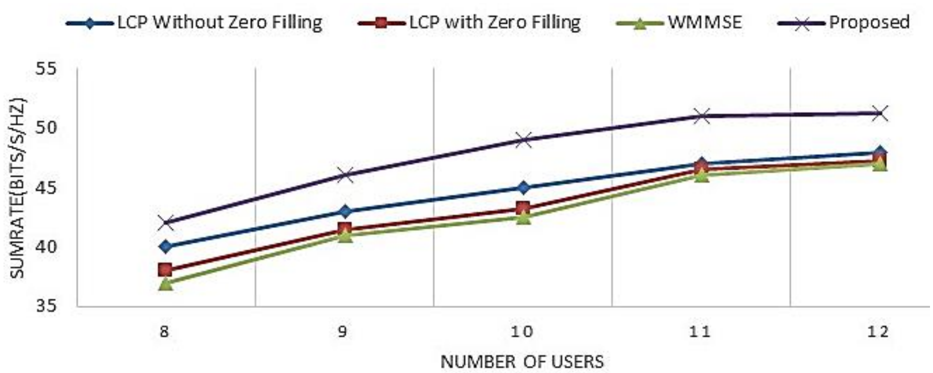

Figure 2. Sum rate performance for varied number of users (training process)

In the next experiment, we measure performance of the proposed approach in terms of sum rate with respect to channel estimation error and actual channel. Figure 3 show the comparative representation of these methods. In this experiment, we have measured the sum rate performance for varied channel estimation error. The proposed approach achieves better performance when compared with the other algorithms. The increase in channel estimation error leads to affect the sum rate performance. The proposed model uses a huge amount of data for training purpose, which helps to reduce the estimation error. For this experiment, the average sum rate is obtained as 28.2, 33.5,33.6, 35.1, and 39.2 by using EZW, WMMSE, LUW, DeepLPC and proposed approach, respectively.

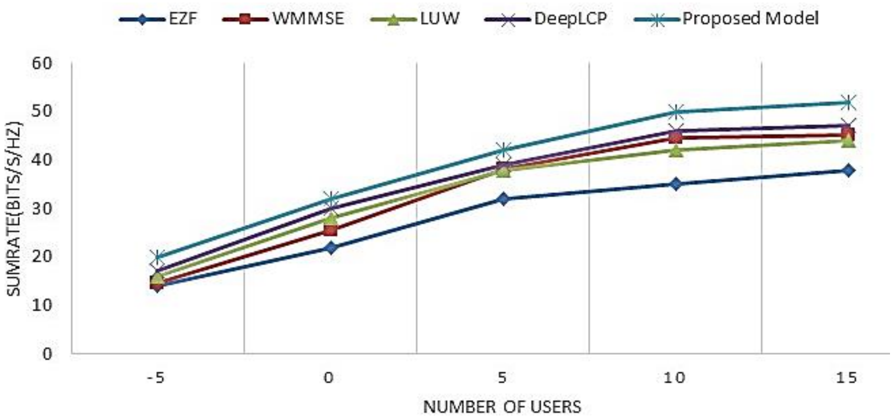

Figure 3. Sum rate performance for imperfect CSI

Further, we measure the SE performance for varied number of training samples and obtained performance is presented in Figure 4 and obtained performance is compared with the existing methods as mentioned in [24], [25]. Long *et al.* [24] authors presented a broad learning method for time-varying channel, similarly, in [25] authors introduced semi-supervised learning method. The average SE performance is obtained as 13.10, 13.20, and 13.60 by using broad learning [24], semi-supervised learning [25], and proposed deep learning approaches, respectively. The robust architecture of the proposed approach helps to achieve the better performance during the initial training process due to consideration of optimized power distribution by using DL approach.

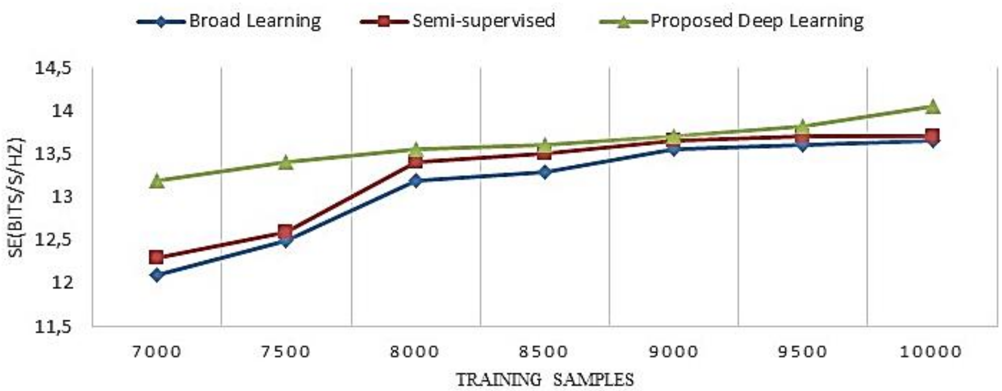

Figure 4. SE performance

In the next experiment, we measure the SE performance for 200 realizations and compare the obtained performance with standard CNN, supervised BL, semi supervised broad learning. We also measure the optimum SE performance for 200 realizations. Figure 5 shows the obtained SE performance. In the Figure 5, the optimal SE for 200 realizations is obtained as 13.72 (bits/s/Hz) and other methods obtained the overall SE as 13.61, 13.6 and 12.8 as mentioned in [24], [25] whereas the proposed approach reported SE performance as 13.65 which is close to the optimum SE. This shows the robustness of the deep learning model.

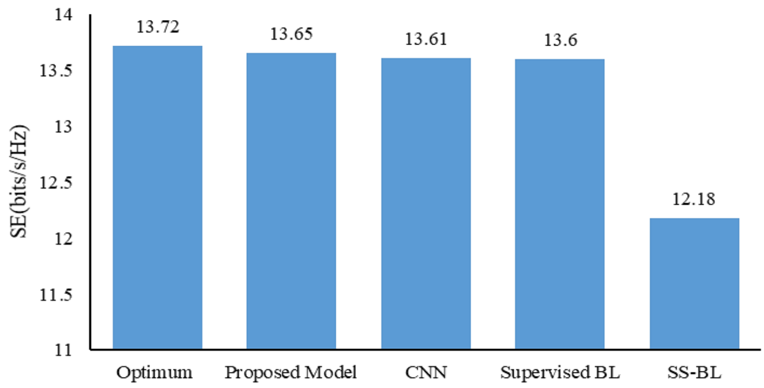

Figure 5. SE performance via diverse approaches

5. CONCLUSION

In this article, we have focused on massive MIMO communication systems and identified the importance of precoding scheme. However, the traditional precoding methods face several challenges therefore; deep learning based methods have gained huge importance to cater issues in the current communication environment. Several researchers have adopted deep learning based hybrid precoding but the performance of these systems is degraded due to computational complexity and inappropriate power allocation. To overcome these issues, we have introduced a combined problem formulation model which considers the sum rate and power allocation performance. The experimental analysis shows that the proposed DL based model achieves better performance when compared with the existing hybrid precoding methods.

ACKNOWLEDGMENT

We also acknowledge Dr. B. R. Patagundi, Principal S G Balekundri Institute of Technology, Belagavi, Karnataka, India for his constant support in conduction of research work in the organization and We thank Dr. Suresh Akkole, Hod, Department of Electronics and Communication Engineering for his motivation and guidance.

REFERENCES

[1] D. C. Araújo, T. Maksymyuk, A. L. F. Almeida, T. Maciel, J. C. M. Mota, and M. Jo, “Massive MIMO: survey and future research topics,” *IET Communications*, vol. 10, no. 15, pp. 1938–1946, Oct. 2016, doi: 10.1049/iet-com.2015.1091.

[2] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, “Massive MIMO linear precoding: a survey,” *IEEE Systems Journal*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018, doi: 10.1109/jsyst.2017.2776401.

[3] D. Mercer, “Global connected and IoT device forecast update,” *Strategy Analytic research services*, May 2019.

[4] J. L. Vora “Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G,” *International journal of modern trends in engineering and research*, 2(10), 281-290, 2015.

[5] A. K. Bashir, R. Arul, S. Basheer, G. Raja, R. Jayaraman, and N. M. F. Qureshi, “An optimal multitier resource allocation of cloud RAN in 5G using machine learning,” *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 8, May 2019, doi: 10.1002/ett.3627.

[6] A. Habbal, S. I. Goudar, and S. Hassan, “A context-aware radio access technology selection mechanism in 5G mobile network for smart city applications,” *Journal of Network and Computer Applications*, vol. 135, pp. 97–107, Jun. 2019, doi: 10.1016/j.jnca.2019.02.019.

[7] C. F. Lam, S. Yin, and T. Zhang, “Advanced fiber access networks,” *Academic Press*, 2022, doi: 10.1016/c2020-0-02217-6.

[8] N. Slamnik-Krijestorac, H. Kremo, M. Ruffini, and J. M. Marquez-Barja, “Sharing distributed and heterogeneous resources toward end-to-end 5G networks: A comprehensive survey and a taxonomy,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1592–1628, 2020, doi: 10.1109/comst.2020.3003818.

[9] X. Zhang, L. Zhu, T. Li, Y. Xia, and W. Zhuang, “Multiple-user transmission in space information networks: Architecture and key techniques,” *IEEE Wireless Communications*, vol. 26, no. 2, pp. 17–23, Apr. 2019, doi: 10.1109/mwc.2019.1800274.

[10] A. Agarwal and S. N. Mehta, “Design and performance analysis of MIMO-OFDM system using different antenna configurations,” *2016 International Conference on Electrical, Electronics, and Optimization Techniques* (ICEEOT), Mar. 2016, doi: 10.1109/iceeot.2016.7754908.

[11] J. O. Nielsen, A. Karstensen, P. C. F. Eggers, E. D. Carvalho, G. Steinbock, and M. Alm, “Precoding for TDD and FDD in Measured Massive MIMO Channels,” *IEEE Access*, vol. 8, pp. 193644–193654, 2020, doi: 10.1109/access.2020.3033166.

[12] A. Grassi *et al.*, “Massive MIMO interference coordination for 5G broadband access: Integration and system level study,” *Computer Networks*, vol. 147, pp. 191–203, Dec. 2018, doi: 10.1016/j.comnet.2018.10.012.

[13] S. Mumtaz, J. Rodriguez, and L. Dai, “Introduction to mmWave massive MIMO,” *mmWave Massive MIMO*, pp. 1–18, 2017, doi: 10.1016/b978-0-12-804418-6.00001-7.

[14] A. Fotouhi, M. Ding, and M. Hassan, “DroneCells: Improving spectral efficiency using drone-mounted flying base stations,” *Journal of Network and Computer Applications*, vol. 174, p. 102895, Jan. 2021, doi: 10.1016/j.jnca.2020.102895.

[15] J. Jiang, X. Wang, G. A. S. Sidhu, L. Zhen, and R. Gao, "Clustering-based codebook design for MIMO communication system," *ICC 2019 - 2019IEEE International Conference on Communications* (ICC), May 2019, doi: 10.1109/icc.2019.8761374.
[16] P. Liang, J. Fan, W. Shen, Z. Qin, and G. Y. Li, "Deep learning and compressive sensing-based CSI feedback in FDD massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9217–9222, Aug. 2020, doi: 10.1109/tvt.2020.3004842.
[17] Z. Li, S. Han, S. Sangodoyin, R. Wang, and A. F. Molisch, "Joint optimization of hybrid beamforming for multi-user massive MIMO downlink," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3600–3614, Jun. 2018, doi: 10.1109/twc.2018.2808523.
[18] S. Wang *et al.*, "A joint hybrid precoding/combining scheme based on equivalent channel for massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 2882–2893, Oct. 2022, doi: 10.1109/jsac.2022.3196099.
[19] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, Oct. 2018, doi: 10.1109/lwc.2018.2818160.
[20] 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 Communications Letters*, vol. 8, no. 2, pp. 416–419, Apr. 2019, doi: 10.1109/lwc.2018.2874264.
[21] Q. Sun, H. Zhao, J. Wang, and W. Chen, "Deep learning-based joint CSI feedback and hybrid precoding in FDD mmWave massive MIMO systems," *Entropy*, vol. 24, no. 4, p. 441, Mar. 2022, doi: 10.3390/e24040441.
[22] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, Nov. 2019, doi: 10.1109/ieeeconf44664.2019.9048966.
[23] M. Zhang, J. Gao, and C. Zhong, "A deep learning-based framework for low complexity multiuser MIMO precoding design," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11193–11206, Dec. 2022, doi: 10.1109/twc.2022.3190435.
[24] Y. Long, Z. Chen, and S. Murphy, "Broad learning based hybrid beamforming for Mm-Wave MIMO in time-varying environments," *IEEE Communications Letters*, vol. 24, no. 2, pp. 358–361, Feb. 2020, doi: 10.1109/lcomm.2019.2952072.
[25] Y. Long, H. Ding, and S. Murphy, "Semi-supervised learning based hybrid beamforming under time-varying propagation environments," *Digital Communications and Networks*, Jan. 2023, doi: 10.1016/j.dcan.2023.01.018.

BIOGRAPHIES OF AUTHORS

Shilpa Bhairanatti completed her M. Tech from Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, India, in 2010. And registered for the Ph.D. in CMR University. Bangaluru in the year 2021. Currently, she is working as an Assistant Professor in the Department of ECE, S.G. Balekundri Institute of Technology, Belagavi, India. She is having 13 years of Teaching and 2 years of research experience. Her Research interest includes Digital Communication, Networking, Antenna design, Network Security, 6G Communication Technologies. She received Young Researcher Award – 2022 from IARDO and Young Technology Author by National Trailblazers Triumph Award for excellence in book writing in 2023. She was a SPOC of Salesforce Certified Platform Developer I (Centre of Excellence). She completed NPTEL and SWAYAM courses on Introduction to Research, Python for Data Science, Data Science for Engineers, Road Map for Patent creation, Microprocessor and Microcontrollers, Digital Circuits. She has published 2 papers on 6G communications during the year 2022-23. She has guided 24 UG projects where 2 projects funded by KSCST. Design patent titled "Intelligent wireless device for detection of cancerous cells" filed in the year 2023. She can be contacted at email: shilpasky2121@gmail.com.

Dr. Rubini P. Ph. D. is a professor and Head/CSE, School of Engineering and Technology, CMR University, Bangalore. She has completed her Ph.D. in Information Science and Engineering from Anna University Chennai. Having 13 years of experience in teaching and research. She has authored a book chapter "Exploiting locality in distributed SDN controller", which has been published in the WORLDS THICKEST BOOK entitled "WORLD 2023 – wide outcomes on Research and Latest Development" this is an unique Guinness Record Attempt, Specification of the book: 5.80m(19 ft 0.34 in). She has filed 25 patents in various IoT sensor techniques and 1 patent was granted in IoT. She has presented papers in various national and international conferences. She has published 40+ papers in various reputed SCI journals. She is an editor and reviewer for several international journals. She is a certified Neo4J Graph Databases professional. Presently she is guiding research scholars associated with CMR University and involved in developing projects using No-SQL databases for data storage and analysis in social networking and associated with consultancy projects in AI using machine learning techniques. She can be contacted at email: rubini.p@cmr.edu.in.