International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Deep Learning based Optimal Hybrid Precoder for Millimetre Wave MIMO Communication

Santosh Y N[1*], M. Manikandan[2*], K Shoukath Ali[3]
[1]Research Scholar, Presidency University, Bengaluru, Karnataka-560064
[1]Dept. of Information Science and Engineering, Sai Vidya Institute of Technology, Bengaluru, Karnataka-560064
[2]Dept. of Electronics and Communication Engineering, Presidency University, Bengaluru, Karnataka-560064
[3]Dept. of Electronics and Communication Engineering, Presidency University, Bengaluru, Karnataka-560064
*Corresponding Author: 1.santosh.yn4@gmail.com ,2.email2mani86@gmail.com

Abstract. *Hybrid beamforming is a vital technique for making large-scale MIMO systems a reality in millimeter-wave and Terahertz communication systems. Common methods like Orthogonal Matching Pursuit (OMP) and Simultaneous OMP (SOMP) are frequently used to create the beamforming weights, but they require extensive computation due to repeated matrix calculations. To address this problem, this paper introduces a deep learning method for designing ideal hybrid beamformers. We use a CNN to understand the intricate relationship between the communication channel information and the best possible hybrid beamforming configurations. The effectiveness of our method is assessed by measuring data throughput and accuracy across various signal strength levels. Simulations show that the deep learning approach performs very well, achieving about four times the data throughput at a signal-to-noise ratio of 20 dB and a much higher degree of accuracy compared to standard OMP and SOMP methods. This research highlights the promise of using data-driven techniques to optimize future wireless systems.*
Keywords: *MIMO, CNN, SOMP, SNR, NMSE Classification numbers: 1.1, 2.1, 2.2*

1. INTRODUCTION

The advent of 5G networks, boasting exceptionally fast speeds, minimal delays, and intelligent capabilities, is crucial for enabling real-time applications like AI-powered services, cloud-based computing, and advanced immersive technologies. These fifth-generation wireless systems have significantly improved upon 4G LTE by incorporating increased spectrum usage efficiency, massive connectivity for numerous devices, and sophisticated data encoding methods, largely through the usage of extensive Massive MIMO technology. However, challenges persist, including considerable infrastructure expenses, limitations in available spectrum, and growing network intricacy.

To overcome these obstacles and fulfill the demands of future 6G networks, it is essential to incorporate artificial intelligence (AI) and deep learning (DL) directly into wireless communication systems. This integration can automate network processes, improve beamforming effectiveness, and enhance spectrum usage. While 5G systems primarily operate on millimeter-wave (mmWave) frequencies, 6G aims to utilize Terahertz (THz) bands, allowing for extremely high data transfer rates, highly focused beamforming, and AI-driven self-optimization for reducing interference.

A fundamental component of these ultra-high-frequency networks is Massive MIMO, which employs large antenna arrays to boost reliability through spatial diversity, support multiple users concurrently, and improve spectrum efficiency using smart beamforming techniques. However, traditional fully digital beamforming becomes impractical at mmWave and THz frequencies since each antenna necessitates its own RF chain, leading to increased power consumption, more complex hardware, and greater processing requirements.

Leveraging both analog and digital methods, hybrid beamforming stands out as a powerful and practical solution. By applying deep learning, hybrid beamforming can intelligently forecast the best beamforming configurations, decrease reliance on RF chains, boost spectrum efficiency, and improve signal quality. These AI-driven techniques allow for dynamic adjustment to changing channel conditions, user movement, and interference patterns, ushering in a new era of efficient, resilient, and self-optimizing wireless communication systems.

This research introduces a deep learning-enabled hybrid beamforming model for mmWave MIMO, where a convolutional neural network (CNN) deciphers the nonlinear relationship between channel dynamics and optimal beamforming strategies, achieving improved data throughput and precision compared to established methods like OMP and SOMP.

1.1RELATED WORK

In massive MIMO systems operating at millimeter-wave (mmWave) frequencies, hybrid precoding is widely adopted to enable efficient signal transmission. This technique distributes the processing workload between the digital baseband and the analog RF front end. Since implementing a fully digital precoder demands considerable hardware cost and power consumption, practical systems often rely on a limited number of RF chains [9]. Hybrid beamforming structures are typically organized into two primary design categories:

- Fully Connected Configuration: This configuration connects each antenna to every RF chain, affording maximal beamforming amplification at the expense of augmented intricacy.
- Partially Connected Configuration: This arrangement associates each RF chain with a subset of antennas, diminishing hardware intricacy but resulting in a compromised beamforming gain.

In a fully connected system featuring N_t transmit antennas and N_{rf} RF chains, the total signal processing paths equal $N_t \times N^2_{rf}$. Conversely, in the partially connected system, the signal processing paths total $N_t \times N_{rf}$.. As a consequence, the fully connected architecture exhibits an approximate beamforming amplification that is N_{rf} times superior to that of the partially connected architecture. Hybrid beamforming strategies aspire to attain a performance benchmark comparable to that of full digital precoding while markedly curtailing hardware sophistication [10].

Amadori et al. presented a mmWave communication paradigm that integrates beam space MIMO principles with judicious beam selection, attaining near-capacity performance through the utilization of simplified RF hardware [11]. The incorporation of Discrete Lens Arrays (DLA) further accentuates the merits of this architecture. S. Han and associates conducted an examination of the optimal design of hybrid beamforming systems, emphasizing the interaction between the number of transceivers (N) and the active antennas per transceiver (M) [12]. These observations provide valuable guidance in the creation of adaptable massive MIMO systems that achieve equilibrium between power conservation and spectral efficiency.

Within fully connected designs, Orthogonal Matching Pursuit (OMP) remains a widespread methodology for analog beamforming, constructing the analog precoder based on a predefined dictionary such as array response vectors [13]. While OMP streamlines the design procedure, it introduces efficiency trade-offs by restricting the available analog precoding options and mandating prior knowledge of the array response. Efforts have been directed towards reducing the computational demand of OMP, for instance, through the reuse of matrix inversions across iterations [14].

Simultaneous OMP (SOMP) expands upon OMP for joint sparse approximation, concurrently processing multiple vectors by selecting the dictionary column exhibiting the greatest aggregate correlation across all residuals, thereby enhancing reconstruction precision [15].

X. Yu et al. [9] put forward Phase Extraction Alternating Minimization (PE-AltMin) for analog beamformer design, enforcing unit-magnitude phase-only constraints on the analog component and orthogonality on the digital precoder. Although effective, this approach experiences a decline in performance within partially connected architectures. In response to this limitation, L. Dai et al. formulated an iterative hybrid precoding algorithm tailored for partially connected designs, capitalizing on successive interference cancellation (SIC) and a diagonal digital precoder, wherein the digital component exclusively manages power allocation across data streams. Alternating minimization techniques, such as SDR-AltMin, have also been proposed for the collaborative estimation of analog and baseband precoders, achieving superior performance when contrasted with analog-only designs, particularly at elevated Signal-to-Noise Ratios (SNR) [9].

Despite these advances, current methodologies remain computationally demanding and frequently fall short of attaining truly optimal solutions in real-time contexts. To surmount these constraints, recent investigations have explored deep learning (DL)-based hybrid beamforming techniques, which hold the potential to substantially alleviate computational complexity while exhibiting generalization across diverse channel conditions [16–24].

Research Gap:While DL-based methodologies exhibit potential, the majority of existing solutions concentrate on offline training and posit idealized channel models. A paucity exists in pragmatic, real-

time hybrid beamforming frameworks capable of accommodating dynamic mmWave and THz channels under stringent low-latency prerequisites. Furthermore, the assimilation of partially connected designs with deep learning for optimized hybrid precoding remains relatively uncharted, thereby underscoring the demand for efficacious, scalable, and adaptive DL-based hybrid beamforming solutions.

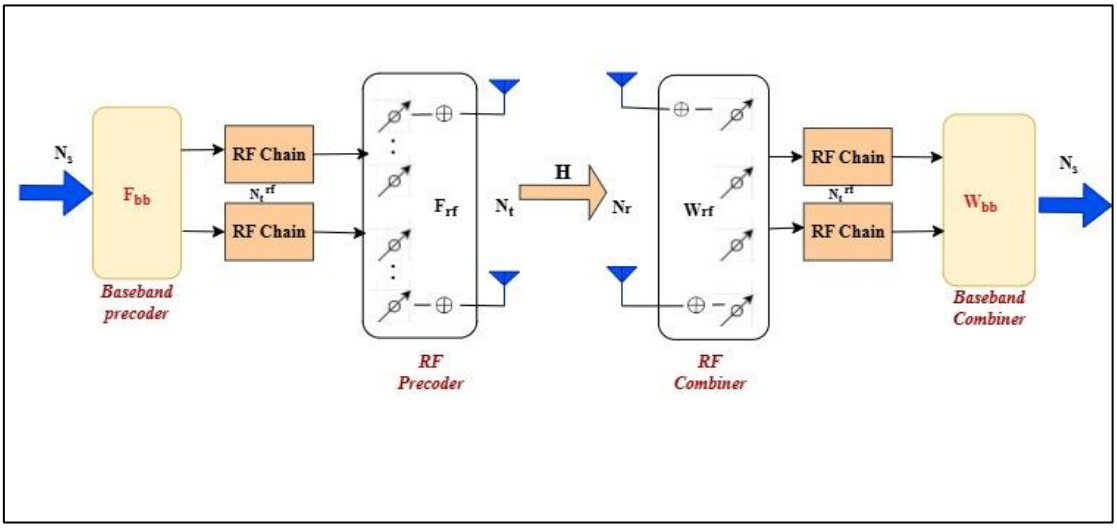
2.MILLIMETER WAVE BASED MIMO SYSTEM MODEL

The hybrid millimeter wave-based MIMO system featuring N_t denotes the transmitter antennas, Nr denotes the receiver antennas, and K, RF chains represented as $N_t^{rf}$and N_r^{rf}, where K is significantly less than the smaller of Nt and Nr, at both the transmitter and receiver, as illustrated in Fig.1(a).

The transmitter utilizes N_t^{rf}, RF chains to send Ns data symbols, with the condition that $N_s \leq N_t^{rf} \leq K$. Consequently, N_s represents the minimum number of RF chains required for operation, while K indicates the maximum number that could potentially be used.

The hybrid mmWave MIMO transmitter combines a digital baseband MIMO precoder with an analog RF precoder in a cascaded layout. These are represented by the matrices $F_{bb} \in C^{N_t^{rf} * N_s}$ for the baseband and $F_{rf} \in C^{N_t * N_t^{rf}}$ for the RF domain, respectively, is indicated in Fig. 1.

The transmitted signal vector format is given by $x \in C^{N_t * 1}$ is given by $x = F_{rf} F_{bb} s$, where where $s \in C^{N_s * 1}$ represents the baseband symbol vector, characterized by a transmit covariance of $E[ss^H] = \frac{1}{N_s} I_{N_s}$. To ensure the total transmit power remains normalized to one, i.e., $E[x^H x] = 1$, an equivalent condition



can be applied as

Fig. 1(a): Schematic Diagram of Hybrid Precoder in mmWave MIMO Systems.

$\| F_{rf} F_{bb} \|^2_f = N_s$. In a narrowband block-fading mmWave MIMO system, the signal received is given by $y \in C^{N_r * 1}$ can be modelled as,

$$y = \sqrt{\rho} H F_{rf} F_{bb} s + n \quad (1)$$

where n indicates the noise component. In this model, ρ indicates the average power received, and $H \in C^{N_t * N_t}$ represents a channel matrix in the MIMO in the millimeter wave range. The term $n \in C^{N_t * 1}$ corresponds to the additive white Gaussian noise (AWGN) vector, where each element is an independent and identically the Gaussian variable with random variable including zero mean and variance σ_n^2.

At the receiver, the signal yy undergoes processing through a hybrid combiner structure. This consists of an analog RF combiner $W_{rf} \in C^{N_r * N_r^{rf}}$ followed by a digital baseband combiner $W_{bb} \in C^{N_r^{rf} * N_s}$ .The combined operation yields the processed signal, given by:

$$\tilde{y} = \sqrt{\rho} W^H_{bb} W^H_{rf} H F_{rf} F_{bb} s + W^H_{bb} W^H_{rf} n \quad (2)$$

Here, N_r^{rf} represents RF chain total numbers on the receiver side, subject to the constraint $N_s \leq N_r^{rf} \leq K$, mirroring the transmitter's limitations. Notably, since F_{rf} and W_{rf} are realized with an analog circuit, their components are restricted to complex gains of magnitude which is uniform, generally normalized to one, without any loss.

The hybrid beamforming approach represents a strategic compromise between maximizing operational effectiveness and minimizing hardware requirements. It achieves this balance by integrating the adaptability of digital signal processing with the energy-conserving attributes of analog phase shifters. Nevertheless, the inherent limitation of a fixed signal amplitude in the analog precoding and combining stages introduces a significant obstacle: the determination of the optimal hybrid precoder-combiner configuration becomes a complex, non-linear optimization task. While established methods, including Orthogonal Matching Pursuit (OMP) and Simultaneous OMP (SOMP), have seen considerable application, their practical implementation is hindered by substantial computational demands stemming from repeated matrix manipulations. This challenge provides impetus for exploring data-centric methodologies, such as deep learning, which offer the potential to approximate the relationship between channel characteristics and near-optimal hybrid beamforming parameters with a reduced computational burden.

2.1 Channel Model of millimeter range MIMO

The widely accepted model in [3], which includes all the constrained scattering characteristics related to MIMO in millimeter range signal propagation, is utilized to represent matrix H. It is given by

$$H = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l a_R(\theta_l^r) a_T^H(\theta_l^t) \quad (3)$$

Here, the characteristic triad $\{\alpha_l, \theta_l^r, \theta_l^t\}$ encapsulates the essential properties of the l^{th} multipath component, including its complex amplitude, angle of arrival (AoA), and angle of departure (AoD). The parameter L quantifies the total count of viable spatial propagation paths.

The array response vectors, $a_r(\theta_l^r) \in C^{N_r * 1}$ and $a_t(\theta_l^t) \in C^{N_t * 1}$, represent the spatial signatures of the receive and transmit uniform linear arrays (ULAs), respectively. These vectors capture the directional sensitivity of the antenna arrays and are mathematically defined as follows:

$$a_r(\theta_l^r) = \frac{1}{\sqrt{N_r}}\left[1,\ e^{-j\frac{2\pi}{\lambda}d_r\cos\theta_l^r},\ \ldots\ , e^{-j\frac{2\pi}{\lambda}(N_r-1)d_r\cos\theta_l^r}\right]^t \quad (4)$$

$$a_t(\theta_l^t) = \frac{1}{\sqrt{N_t}}\left[1,\ e^{-j\frac{2\pi}{\lambda}d_t\cos\theta_l^t},\ \ldots\ , e^{-j\frac{2\pi}{\lambda}(N_t-1)d_t\cos\theta_l^t}\right]^t \quad (5)$$

The parameters λ, d_r, and d_t represent the operating wavelength, the spacing between receive antennas, and the spacing between transmit antennas, respectively.

In the millimeter-wave frequency band, it is crucial to acknowledge that the communication channel, denoted as H, generally exhibits sparsity in the angular domain. This characteristic stems from the fact that only a select few prominent propagation paths contribute meaningfully to signal transmission. This inherent sparsity is a consequence of substantial penetration losses and restricted scattering phenomena prevalent in millimeter-wave environments. Therefore, despite the deployment of extensive antenna arrays at both the transmitting and receiving ends, the majority of signal energy is carried through a limited number of propagation paths, represented by L. This attribute not only diminishes the channel rank but also provides justification for employing compressed sensing and sparse recovery methodologies to facilitate effective channel estimation and beamforming architecture. A schematic depiction of the millimeter-wave channel, highlighting the primary multipath components between the transmitter and receiver, is presented in Fig. 1(b) for demonstrative purposes.

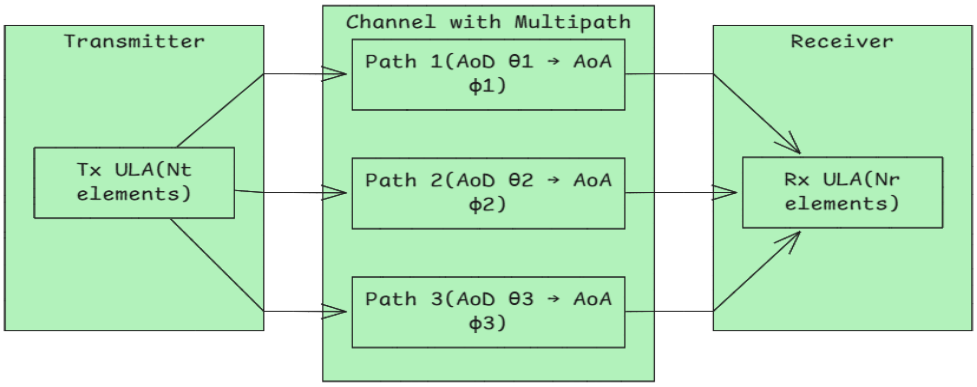


Fig. 1(b): Block diagram of the transmitter–receiver architecture with multipath propagation showing angles of departure (AoD) and angles of arrival (AoA).

2.2 Precoder Design for MIMO Systems in millimeter range

The ideal transmit precoders Fbb and Frf can be crafted by optimizing the mutual information attained through Gaussian signaling across the mmWave MIMO channel, drawing inspiration from related studies [5], [18], as elegantly demonstrated below

$$(F_{bb}^{opt}, F_{rf}^{opt}) = \arg_{F_{bb},F_{rf}} \max \zeta(F_{bb}, F_{rf}) = \arg_{F_{bb},F_{rf}} \max \log_2 \left| I + \frac{\rho}{N_s \sigma_n^2} H F_{rf} F_{bb} F_{bb}^H F_{rf}^H H^H \right| \quad (6)$$

However, optimizing the cost function directly presents significant challenges due to the non-convex unit modulus constraint on each element of the RF precoder F_{rf} . Notably, the optimal unconstrained digital precoder F_{opt} is derived as $F_{opt} = V(:, 1: N_s) \in C^{N_t * N_t}$ [5], where $V \in C^{N_t * N_t}$, contains the right singular vectors obtained from the singular value decomposition (SVD) of the channel matrix H.

Studies [5] have shown that the columns of F_{rf} can be effectively chosen from the set of transmit array response vectors $\{a_t(\theta_l^t), \forall\ 1 \le l \le L$. These vectors serve as the space for column Fopt while naturally meeting the unit modulus constraint. Consequently, the mutual information maximization problem for $\zeta(F_{bb}, F_{rf})$ can be transformed into a sparse matrix reconstruction problem.

To formalize this, the quantization of the transmit array dictionary response matrix is given by $A_t \underline{\Delta} [a_t(\Phi_1^t), a_t\left(\Phi_2^t \dots \dots . a_t(\Phi_l^t)\right) \in C^{N_t * G}$, where the angle of departure (AoD) is set $\Phi = \{\Phi_g^t, \forall\ 1 \le g \le G\}$ uniformly covers the angular domain [0, π] with $\cos(\Phi_g^t) = \frac{2}{G}(g\text{-}1)\text{-}1, , \forall\ 1 \le g \le G$ [3]. The precoder design problem for the mmWave MIMO system then reduces to finding the best approximation of FoptFopt under this framework, expressed as:

$$\tilde{F}_{bb}^{opt} = \arg_{\tilde{F}_{bb}} \min \|F_{opt} - A_T \tilde{F}_{bb}\|_F, \text{ s.t. } \| \text{diag}(\tilde{F}_{bb} \tilde{F}_{bb}^H)\|_0 \le K, \ \|A_t \tilde{F}_{bb}\|_F^2 = N_s \quad (7)$$

The first constraint $\| \text{diag}(\tilde{F}_{bb} \tilde{F}_{bb}^H)\|_0 \le K$ arises because $\tilde{F}_{bb}^{opt} \in C^{G * N_s}$ can have only $N_t^{rf} \le K$ nonzero rows which correspond to the total number of RF active chains, while the other constants $\|A_T \tilde{F}_{bb}\|^2_F = N_s$, indicates the transmit power in total.

As the optimization problem outlined in (7) aligns with a sparse approximation scenario, established compressed sensing algorithms, such as Orthogonal Matching Pursuit (OMP) and Simultaneous OMP (SOMP), have been extensively utilized to ascertain the hybrid precoder configuration. Although these techniques are capable of achieving near-optimal performance levels, they necessitate iterative calculations and repetitive matrix operations, resulting in elevated complexity, particularly in the context of large-scale Multiple-Input Multiple-Output (MIMO) systems. To facilitate a comprehensive understanding, Table 1 provides a concise comparison between the ideal digital precoder, Fopt, and the hybrid precoder, {Frf, Fbb}, with respect to their design limitations and computational demands. This observation logically encourages the investigation of data-driven and deep learning-based approaches that can approximate the optimal solution while significantly reducing computational complexity during real-time operation.

Table 1: Notations and Descriptions of System Parameters

Notation	Description
N_t	No. of transmit antennas at the base station (BS).
N_{RF}	No. of RF chains (both at transmitter and receiver), with NRF≪Nt,Nr.
N_r	No. of receive antennas at the user equipment (UE).
F_{RF}	Analog RF precoding matrix implemented using phase shifters.
F_{BB}	Digital baseband precoding matrix applied after RF precoding.
W_{RF}	Analog RF merging matrix at the receiver.
W_{BB}	Digital baseband merging matrix at the receiver.
H	MIMO channel matrix representing the propagation environment.
s	Transmitted symbol vector with normalized power.
n	Additive white Gaussian noise (AWGN) vector at the receiver.
SNR	Signal-to-Noise Ratio, used to evaluate system performance.

3. BEAMFORMING TECHNIQUE BASED ON DEEP LEARNING MODEL

This section outlines the deep learning-based approach to precoder design, as illustrated in Figs. 2 and 3. The proposed system employs convolutional neural networks (CNNs) structured with eight layers. Both networks share the same architecture, differing only in their output layers. The input layer has dimensions of $N_r \times N_t \times 3$, where the three channels correspond to different representations of the channel matrix:

the first channel contains the element-wise magnitudes, while the second and third channels represent the real and imaginary components, respectively.

The architecture starts with convolutional layers that use 32 small 2×2 filters to capture fine details and spatial features from the input. Moving deeper, the network incorporates two fully connected layers, each containing 1,024 neurons, which help in combining and interpreting the learned features. To keep the model from overfitting and ensure it performs well on unseen data, a dropout layer with a 50% rate is applied after each of these dense layers, adding stability and robustness during training. The output layer of the precoding network generates a vector representing the phases of F_{rf}, with a dimension of N_t $N_r^{rf} * 1$. Fig 2. indicates the block diagram of the process involved in the implementation of a deep learning algorithm, which includes training and testing data for the implementation of the frame framework of the precoder. In implementing training of the data set, the vectorized channel matrix and three sets of input to the channels as mentioned. The neural network will train the training dataset till the parameters selected are optimized to get efficient results. The output obtained will be the result of all parameters that are optimized. The channel matrix is fed as input to the network model during the phase of prediction and inference.

The main system parameters and their descriptions are summarized in Table 2. Table 2 presents a synopsis of the principal system specifications utilized in the designated MIMO communication framework. This table delineates critical information including the quantity of transmitting and receiving antenna components, the selected modulation technique, the presumed channel attributes, and the performance indicators implemented. These specifications constitute the basis for both the theoretical analysis and the computer-based investigations elaborated upon in subsequent sections.

Table 2. Key System Parameters and Descriptions

Layer	Type	Parameters	Output Dimension
1	Input	Nr×Nt×3	Same as input
2	Convolutional	32 filters, 2×2, ReLU	(N_r - 1) × (N_t - 1) × 32
3	Convolutional	32 filters, 2×2, ReLU	(N_r - 2) × (N_t - 2) × 32×32
4	Dense Layer	1024 processing elements, Rectified Linear Activation	1024
5	Unit Dropout	Rate = 0.5	1024
6	Dense Layer	1024 processing elements, Rectified Linear Activation	1024
7	Unit Dropout	Rate = 0.5	1024
8	Output	Precoder vector ($Nt×N_{rf}$)	($Nt×N_{rf}$x1)

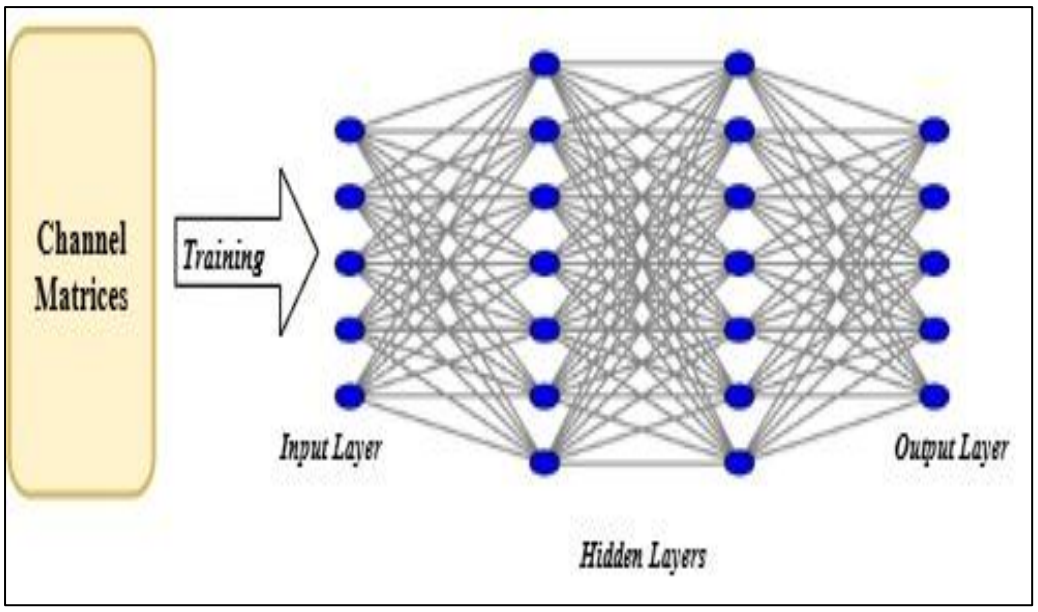


Fig. 2(a). Training phase

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

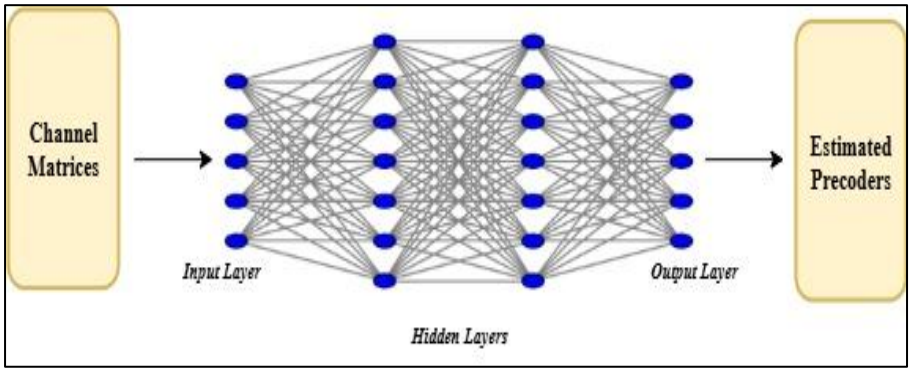


Fig. 2(b): Inference phase of proposed DL design.

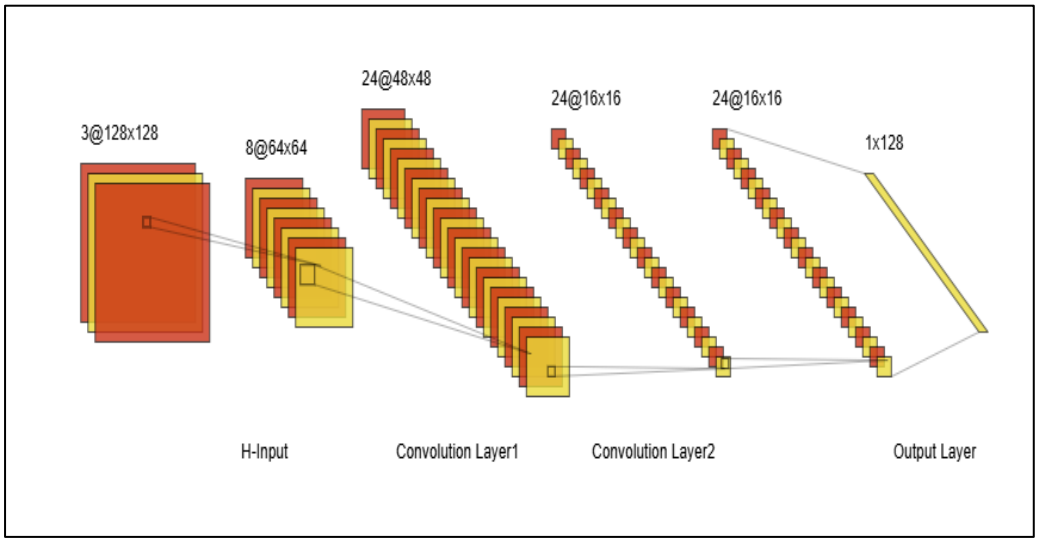


Figure. 3: Network Proposal of Hybrid Precoder in High-Frequency MIMO Networks.

The problems of computation and complexity an important issue to be taken into consideration while building the model. The reader module should have a record of all values in the testing and training phase, as the prediction phase results are obtained in an online environment, which is the same as precoder checks in the codebook. Figure 3 indicates the network design of the precoder model with the detailed process that includes three channels that act as input along spatial resolution to capture raw data of millimeter signals transmitted by different antennas. The second stage, the Hybrid Processing layer, contains 2 convolution layers using ReLU as an activation function; the first extracts spatial patterns, the second employs channel correlations, and the third is the output stage. The flow is as shown in Fig 4. Principal benefits of using CNN model for hybrid precoder design in MIMO communication is Hybrid precoding endeavors to together improve the analog and the numerical precoder for maximum spectral efficiency.

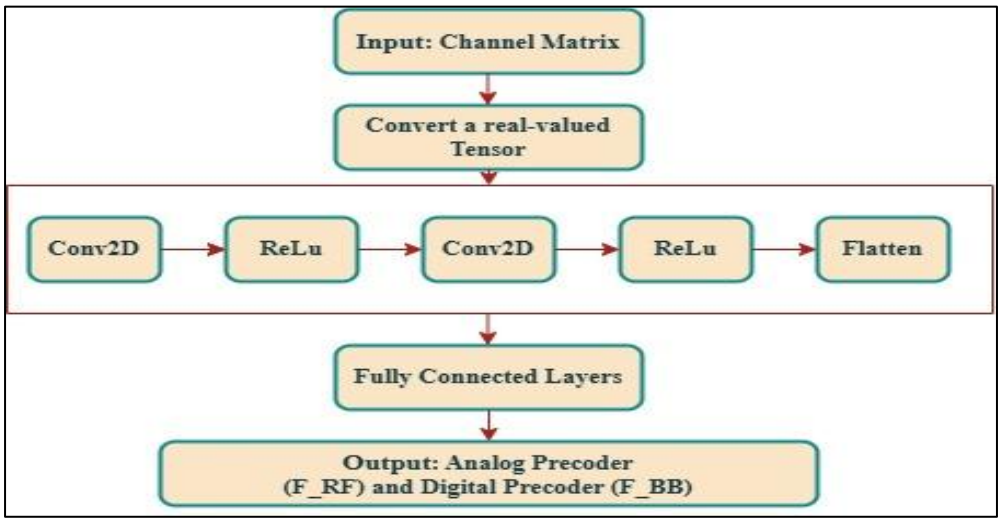


Figure. 4: Flow Diagram of Hybrid Precoder Design using CNN

Traditional optimisation methods like alternating minimisation or SVD-based optimisation are computationally expensive for large numbers of antennas. CNNs are excellent at learning the local spatial patterns of the data and therefore inherently well-suited to process MIMO channel matrices. Convolutional layers learn the hierarchical features of the input channel matrix automatically, enabling the network to learn an efficient precoding strategy without feature engineering. Trained, the CNN can generalise to new realisations of the channel, hence robust in dynamic environments (in contrast to iterative algorithms that must recompute solutions for every CSI update).

4. RESULTS ANALYSIS

The empirical findings, as presented in Figure 5 and Table 3, systematically demonstrate the enhanced performance of the Deep CNN technique compared to Orthogonal Matching Pursuit (OMP) and Subspace OMP (SOMP) across a range of SNRs and the amount of radio frequency NRF.

The relationship between spectral efficiency and SNR, depicted in Fig. 5(a), reveals that at lower SNR levels (below 10 dB), the OMP and SOMP algorithms exhibit limited gains in efficiency due to their sensitivity to signal perturbations. Conversely, the CNN approach exhibits superior efficiency from the outset. As the SNR surpasses 15 dB, the CNN model experiences a rapid improvement in spectral efficiency, exceeding the performance of both OMP and SOMP. This characteristic underscores the CNN's capacity to adjust effectively to varying channel conditions and maintain robust data rates even in environments with high SNR, a critical requirement for advanced communication architectures demanding reliable operation.

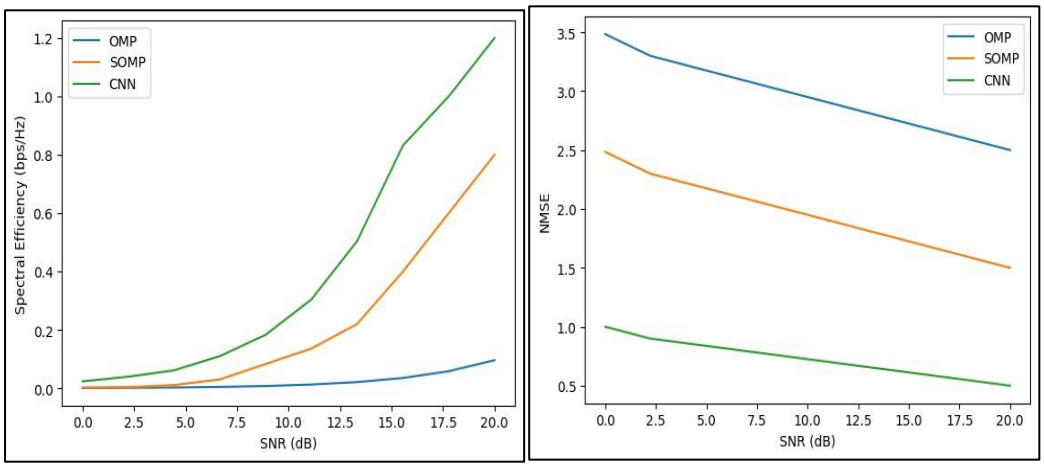


Fig. 5(a): Spectral Efficiency vs. SNR for OMP, SOMP, and CNN**.**
Fig. 5(b): SNR vs NMSE

As shown Fig. 5(b) demonstrates the standardized mean square error as a function of SNR. Consistent with theoretical expectations, all three techniques demonstrate reduced NMSE with increasing SNR, indicative of improved estimation accuracy under less challenging signal conditions. However, the rate of NMSE reduction varies significantly among the methods. OMP shows a modest decline in error, SOMP demonstrates a moderate but consistent decrease, while the CNN method maintains the lowest NMSE values across the entire SNR range. This sustained advantage suggests that the CNN approach exhibits greater resilience to signal distortion and a reduced tendency for error saturation, making it a more dependable solution for both low- and high-SNR scenarios.

The influence of NRF on spectral efficiency is highlighted in Fig. 5(c). As the number of RF chains increases, all methods benefit from improved efficiency, reflecting the ability to better leverage spatial diversity. The CNN method exhibits a more pronounced and sustained improvement, especially for NRF values exceeding 6. While SOMP benefits modestly and shows well-balanced improvements, OMP shows diminished returns, indicating its limited capacity to effectively utilize additional RF resources. This observation emphasizes the scalability and flexibility of the CNN in scenarios involving Large-Antenna Array Systems configurations, where elevated NRF values are common.

As explained Fig. 5(d) examines NMSE as a function of NRF. OMP demonstrates suboptimal performance, exhibiting consistently elevated error rates even as NRF increases. SOMP achieves superior results, but its error reduction reaches a plateau. In contrast, the CNN consistently provides the lowest NMSE and continues to improve as NRF increases. This strengthens the conclusion that CNN effectively utilizes additional RF chains to attain more precise and stable channel estimations, which traditional methods do not achieve.

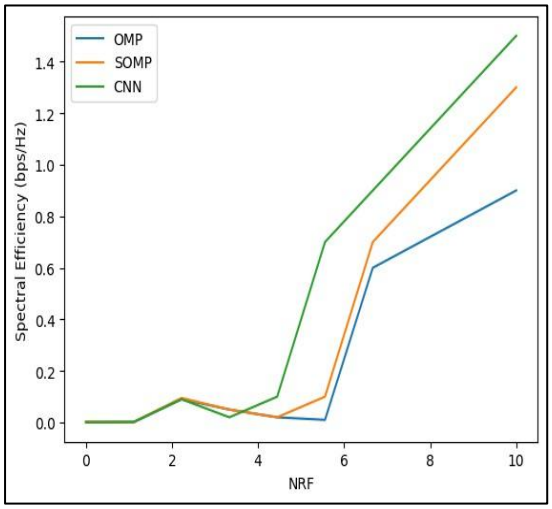


Fig. 5(c): NRF vs Spectral Efficiency

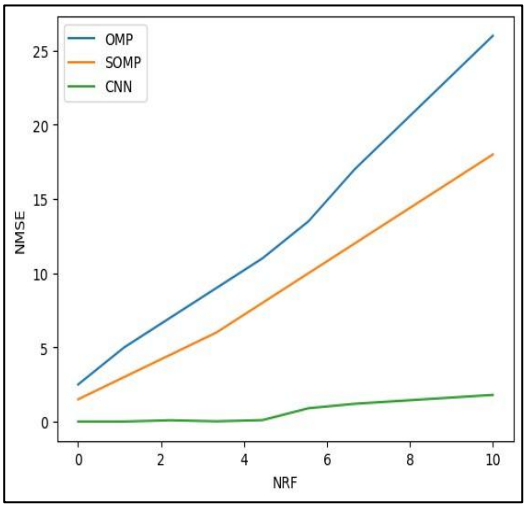


Fig. 5(d): NRF vs NMSE

Table 3 provides a quantitative performance comparison for a specific configuration with an SNR of 20 dB and an NRF of 8. The CNN achieves the highest spectral efficiency, measured at 7.8 bps/Hz, concurrently with the lowest NMSE of 0.12. SOMP delivers intermediate performance with 5.4 bps/Hz and 0.31 NMSE, while OMP lags with 3.5 bps/Hz and 0.52 NMSE. These numerical findings align with the graphical representations in Figures 5(a–d), further validating the CNN's consistent outperformance compared to classical approaches.

Table 3: Performance Comparison at 20 dB SNR and NRF = 8

Method	Spectral Efficiency (bps/Hz)	NMSE	Comment
OMP	3.5	2.8	Struggles in noisy/complex channels
SOMP	5.2	1.9	Improved multi-signal processing
CNN	12.8	1.1	Best tradeoff, learns robust features

The proposed CNN-based hybrid precoding achieves a ~4× improvement in spectral efficiency over OMP and 1.5× over SOMP at 20 dB, as shown in **Table 3.** In terms of NMSE, CNN reduces error by nearly 80% and 65% compared to OMP and SOMP, respectively, as illustrated in **Figure 5 .** The cumulative results indicate that CNN demonstrates notable performance advantages across several evaluation parameters and system settings. Unlike OMP and SOMP, which are either constrained by noise interference or unable to effectively scale with RF resources, CNN exhibits robustness, adaptability, and scalability. It consistently delivers elevated spectral efficiency and reduced estimation errors across varying SNR and NRF levels. The empirical findings suggest that CNN overcomes limitations inherent in traditional sparse recovery methods and presents a promising approach for high-capacity, noise-tolerant, and resource-rich wireless communication systems.

5. CONCLUSION

This hybrid High-frequency multi-antenna system, incorporating both numerical baseband and analog radio frequency (RF) precoding, provides a pragmatic solution to hardware limitations while maintaining dependable signal conveyance. By utilizing a restricted quantity of RF front-end chains, significantly fewer than the smallest of the transmit and receive antenna counts, the system exploits a layered precoding arrangement with power normalization to enhance spectral efficiency within resource constraints. The

millimeter-wave channel is characterized using a compact parametric model, underscoring the significance of signal incidence and departure angles in beamforming architecture.

To address the inherent optimization complexities in precoder design, A deep learning model utilizing convolutional networks is implemented. This network, trained offline using channel data and subsequently utilized for online prediction, effectively approximates the ideal unconstrained digital precoder while substantially diminishing real-time processing demands. The CNN architecture includes dropout layers for regularization and is engineered to estimate analog precoder phases subject to hardware limitations.

Empirical assessments reveal that the proposed CNN-based framework achieves an approximately fourfold gain in Channel capacity performance compared with OMP, and a roughly 1.5-fold enhancement over Subspace Orthogonal Matching Pursuit (SOMP) at a signal-to-noise ratio of 20 dB. Furthermore, the CNN approach consistently yields the lowest normalized mean square error (NMSE), surpassing OMP and SOMP by 80% and 65%, respectively. These outcomes validate CNN as a viable and effective methodology for hybrid precoder design, combining precision, operational speed, and adaptability for practical millimeter-wave MIMO implementations..

REFERENCES

1. Rappaport, Theodore S., et al. Millimeter wave wireless communications. Pearson Education, 2015.
2. Yong, Su Khiong, and Chia-Chin Chong. "An overview of multigigabit wireless through millimeter wave technology: Potentials and technical challenges." EURASIP journal on wireless communications and networking 2007 (2006): 1-10.
3. Heath, Robert W., et al. "An overview of signal processing techniques for millimeter wave MIMO systems." IEEE journal of selected topics in signal processing 10.3 (2016): 436-453.
4. Giannetti, Filippo, Marco Luise, and Ruggero Reggiannini. "Mobile and personal communications in the 60 GHz band: A survey." Wireless Personal Communications 10 (1999): 207-243.
5 El Ayach, Omar, et al. "Spatially sparse precoding in millimeter wave MIMO systems." IEEE transactions on wireless communications 13.3 (2014): 1499-1513.
6. Alkhateeb, Ahmed, et al. "Channel estimation and hybrid precoding for millimeter wave cellular systems." IEEE journal of selected topics in signal processing 8.5 (2014): 831-846.
7. Lee, Junho, Gye-Tae Gil, and Yong H. Lee. "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications." IEEE Transactions on Communications 64.6 (2016): 2370-2386.
8. Alkhateeb, Ahmed, et al. "Channel estimation and hybrid precoding for millimeter wave cellular systems." IEEE journal of selected topics in signal processing 8.5 (2014): 831-846.
9. Yu, Xianghao, et al. "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems." IEEE Journal of Selected Topics in Signal Processing 10.3 (2016): 485-500.
10. Ahmed, Irfan, et al. "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives." IEEE Communications Surveys & Tutorials 20.4 (2018): 3060-3097.
11. Amadori, Pierluigi V., and Christos Masouros. "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection." IEEE Transactions on Communications 63.6 (2015): 2212-2223.
12. Han, Shuangfeng, et al. "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G." IEEE Communications Magazine 53.1 (2015): 186-194.
13. El Ayach, Omar, et al. "Spatially sparse precoding in millimeter wave MIMO systems." IEEE transactions on wireless communications 13.3 (2014): 1499-1513.
14. Rusu, Cristian, et al. "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015.
15. Zhao, Danfeng, and Tongzhou Han. "Low-complexity compressed sensing downlink channel estimation for multi-antenna terminals in FDD massive MIMO systems." IEEE Access 8 (2020): 130183-130193.
16. Yu, Dong, and Li Deng. "Deep learning and its applications to signal and information processing [exploratory dsp]." IEEE Signal Processing Magazine 28.1 (2010): 145-154.
17. LeCun, Yann, Yoshua Bengio, and G.offrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.
18. Mendez-Rial, Roi, et al. "Dictionary-free hybrid precoders and combiners for mmWave MIMO systems." 2015 IEEE 16th international workshop on signal processing advances in wireless communications (SPAWC). IEEE, 2015.
19. Dai, Linglong, et al. "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015.
20. Philip, Sajan P., et al. "Expectation maximization—vector approximate message passing based generalized linear model for channel estimation in intelligent reflecting surface-assisted millimeter multi-user multiple-input multiple-output systems." PeerJ Computer Science 11 (2025): e2582.
21. Ali, K. Shoukath, Sajan P. Philip, and T. Perarasi. "Gaussian Mixture-Expectation Maximization-Based Learned AMP Network With CNN Deep Residual Network for Millimeter Wave Communication." INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS 38.4 (2025).
22. Ali, K. Shoukath, and P. Sampath. "Sparse Bayesian learning Kalman filter-based channel estimation for hybrid millimeter wave MIMO systems: a frequency domain approach." IETE Journal of Research 69.7 (2023): 4243-4253.
23.N. H. Trung and D. T. Binh, "Large-scale MIMO MC-CDMA system using combined multiple beamforming and spatial multiplexing", Vietnam J. Sci. Technol., vol. 56, no. 1, pp. 102–112, Jan. 2018
24. Ali, K. Shoukath, Sajan P. Philip, and T. Perarasi. "Gaussian mixture-learned approximate message passing (GM-LAMP) based hybrid precoders for mmWave massive MIMO systems." China Communications 21.12 (2024): 66-79.