

Convolutional neural network-based hybrid beamforming design based on energy efficiency for mmWave M-MIMO systems

Hanane Ayad, Mohammed Yassine Bendimerad, Fethi Tarik Bendimerad

LTT Laboratory, Department of Telecommunication, University Abou Baker Belkaïd, Tlemcen, Algeria

Article Info

Article history:

Received Dec 27, 2024

Revised Jul 25, 2025

Accepted Sep 16, 2025

Keywords:

Convolutional neural network

Deep learning

Energy efficiency

Hybrid beamformers

M-MIMO

ABSTRACT

Millimeter-wave (mmWave) massive multiple-input multiple-output (M-MIMO) technology brings significant improvements in data transmission rates for communication systems. A key to the design of mmWave M-MIMO systems is beamforming techniques, which focus signals toward specific directions but rely on expensive, energy-intensive radio frequency (RF) chains. To address this issue, hybrid beamformers (HB) have been introduced as a partial solution, and deep learning (DL) has proven effective for HB design. However, previous works utilizing machine learning (ML) networks have primarily focused on the spectral efficiency (SE) metric for constructing HB. In this paper, we present a convolutional neural network (CNN) architecture whose loss function is defined to maximize energy efficiency (EE) directly. The network jointly learns analog and digital beamformers by evaluating EE (throughput per total power, including phase shifters, switches, digital-to-analog converters (DACs), and RF chains) and selecting the configuration that yields the highest EE. The CNN takes a channel matrix as input and outputs RF and baseband beamformer matrices. Simulation results validate the effectiveness of the proposed M-MIMO EE scheme, achieving significant EE improvements by optimizing hybrid precoding and reducing RF chain usage.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Hanane Ayad

LTT Laboratory, Department of Telecommunication, Abou Baker Belkaïd University

Tlemcen, 1300, Algeria

Email: ayad.hanane@univ-tlemcen.dz

1. INTRODUCTION

The Massive multiple-input multiple-output (M-MIMO) has become a cornerstone of 5G and is projected to play an equally central role in 6G, where it is expected to address the demands for ultra-high data rates, massive connectivity, and low latency by leveraging large antenna arrays to enhance spectral efficiency (SE), coverage, and link reliability [1], [2]. At millimeter-wave (mmWave) frequencies, fully digital (FD) beamforming demands as many power-hungry RF chains as antennas, making it impractical for real-world systems. Hybrid precoding (HP) addresses this by splitting the beamforming into a low-dimensional digital stage and an analog network of phase shifters or switches, thereby approximating FD performance with significantly fewer RF chains [3]–[4]. Most HP schemes have been developed to maximize SE, including spatially sparse codebooks [3], alternating-minimization algorithms [5], subsequent innovations introduced dynamic-stream assignment for uplink multiuser orthogonal frequency division multiplexing (OFDM) [6], switch-based analog networks [7], and per-RF-chain clustering in centralized cell-free systems [8]. More recent contributions leverage meta-heuristic optimizations such as genetic-algorithm based HP [9], and beam-

division multiple access with zero-forcing and orthogonal user-beam pairing to reach near-digital SE with fewer RF chains [10]. Additionally, relay-reflecting intelligent surfaces have been shown to simultaneously amplify and steer signals, further enhancing SE in cell-free M-MIMO [11]. As networks density, energy efficiency (EE) has emerged as a key design criterion. Techniques such as low-resolution DACs, adaptive subarray allocation, and channel-matrix optimization have significantly improved EE in massive MIMO [12]–[13]. Recent work like hybrid analog and digital Beamformers with low resolution (HANDBALL) [14] further integrates sensing with coarse-quantized beamforming. However, few hybrid precoding studies embed EE directly into their optimization, instead relying on heuristic power reductions following SE-driven design.

Recently, deep learning (DL) has proven effective for HB design. In the context of HP design, CNN-based methods have been employed to jointly optimize analog and digital components under imperfect CSI, improving SE over conventional designs [15], [16]. Further enhancements address EE by integrating adaptive fully-connected networks [17]. Complementary studies explore CNNs for subarray configurations [18], reinforcement learning for dynamic beam selection [19], and deep unfolding for fast, trainable HP [20]. Unsupervised approaches also eliminate the need for labeled data and codebooks, proving effective in distributed and quantized settings [21]–[22]. In parallel, extensive surveys of machine learning in massive MIMO highlight both the opportunities and open challenges in applying DL to hybrid beamforming [23].

To the best of our knowledge, this is the first DL-based HP optimization that directly incorporates EE criteria. Our solution is a CNN based HB design for mmWave M-MIMO that directly targets EE rather than SE. Specifically, we introduce an EE-aware loss function that combines SE and realistic hardware power consumption (phase shifters, switches, DACs, RF chains), steering the network toward beamformer configurations that maximize EE. We embed this hardware power model in the CNN so that both analog and digital weightings intrinsically account for actual energy usage.

The organization of this article is as follows: section 2 presents the theoretical foundations of the study, including the signal model, energy efficiency, and problem formulation. The proposed CNN-based beamforming strategy is also described in this section. Section 3 details the simulation environment and system parameters used in MATLAB, along with the CNN training strategy and hyperparameter configuration. Section 4 reports and analyzes the simulation results, highlighting the performance gains achieved in terms of energy and spectral efficiency under various RF configurations. Finally, section 5 concludes the paper and outlines potential directions for future work.

Notation: a denotes a scalar, $\mathbf{a}$ is a vector and $\mathbf{A}$ is a matrix. For a vector $\mathbf{a}$, the notation $[\mathbf{a}]_i$ denotes its i -th element. Similarly, for a matrix $\mathbf{A}$, $[\mathbf{A}]_{:,i}$ and $[\mathbf{A}]_{i,j}$ represent the i -th column and the (i, j) -th entry, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ indicate transpose, and Hermitian operations. The Frobenius norm is represented by $\|\cdot\|_F$ and $\mathbf{I}_N$ is an identity matrix of size N . In this context, $[\mathbf{A}]_{:,i}$ refers to the full column vector composed of all rows at the i -th column position.

2. THEORETICAL FOUNDATION AND PROPOSED APPROACH

2.1. SIGNAL MODEL

Figure 1 depicts a mmWave M-MIMO system equipped with N_{BS} transmitting antennas that serves a single-user mobile station with N_{MS} receiving antennas. The transmitter provides N streams of data symbols to the receiver over the network. Figure 1 shows how the base station (BS) precodes the data streams using $Nt_{RF} \times N$ digital precoders $\mathbf{F}_D$ and $N_{BS} \times Nt_{RF}$ analog precoders $\mathbf{F}_A$. The discrete-time data streams are represented by the vector $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$. The covariance matrix of $\mathbf{s}$ vector is $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_N/N$ under the assumption of independence and a Gaussian distribution with zero mean and unit variance. Afterward, the transmitted signal is expressed as $\mathbf{x} = \mathbf{F}_A \mathbf{F}_D \mathbf{s}$. The transmitter is subject to power limitation according to the constraint $\|\mathbf{F}_A \mathbf{F}_D\|_F = N$, and the analog beamformers are unitary matrices with equal-norm elements, *i.e.*, $[\mathbf{F}_A]_{:,i} [\mathbf{F}_A]_{:,i}^H = 1/N_{BS}$. For a narrowband block-fading channel, the signal received at the N_{MS} antennas can be written as:

$$\mathbf{r} = \sqrt{\rho} \mathbf{H} \mathbf{F}_A \mathbf{F}_D \mathbf{s} + \mathbf{n} \quad (1)$$

where $\mathbf{r}$ is the $N_{MS} \times 1$ received signal vector, $\mathbf{H}$ is the channel matrix with $N_{MS} \times N_{BS}$ dimensions, ρ is the average received power, and $\mathbf{n}$ is the additive white Gaussian noise (AWGN) *i.e.*, $\mathbf{n} \sim N(0, \sigma_n^2 \mathbf{I}_{N_{MS}})$.

The received signal is processed by $N_{MS} \times Nt_{RF}$ analog combiner $\mathbf{W}_A$ with the constraint $[\mathbf{W}_A]_{:,i} [\mathbf{W}_A]_{:,i}^H = 1/N_{MS}$, and $Nt_{RF} \times N$ digital combiner $\mathbf{W}_D$ as (2):

$$\tilde{\mathbf{r}} = \sqrt{\rho} \mathbf{W}_D^H \mathbf{W}_A^H \mathbf{H} \mathbf{F}_A \mathbf{F}_D \mathbf{s} + \mathbf{W}_D^H \mathbf{W}_A^H \mathbf{n} \quad (2)$$

The power P_{RF} is given by:

$$P_C = P_{LO} + P_{PA} + Nt_{RF}[2P_{DAC} + P_{RF}] + Nt_{RF}N_{BS}P_{PS} \quad (5)$$

Then, the Energy efficiency can be derived as (6):

$$EE = \frac{1}{P_C} \log_2 \left(\left| \mathbf{I}_N + \frac{\rho}{N} \mathbf{R}_n^{-1} \mathbf{W}_D^H \mathbf{W}_A^H \mathbf{H} \mathbf{F}_A \mathbf{F}_D \times \mathbf{F}_D^H \mathbf{F}_A^H \mathbf{H}^H \mathbf{W}_A \mathbf{W}_D \right| \right) \quad (6)$$

where $\mathbf{R}_n^{-1} = \sigma_n^2 \mathbf{W}_D^H \mathbf{W}_A^H \mathbf{W}_A \mathbf{W}_D$ is the noise covariance matrix after the combining block.

2.3. Problem formulation

The joint optimization problem for HB estimation can be written as (7):

$$\begin{aligned} & \underset{\mathbf{F}_A, \mathbf{F}_D, \mathbf{W}_A, \mathbf{W}_D}{\operatorname{argmax}} EE \\ & \text{subject to: } \mathbf{F}_A \in \mathcal{F}_{RF}, \mathbf{W}_A \in \mathcal{W}_{RF}, \\ & \quad \|\mathbf{F}_A \mathbf{F}_D\|_F^2 = N \end{aligned} \quad (7)$$

The sets of analog beamformers that are technically possible under $\mathbf{F}_A$ and $\mathbf{W}_A$ constraints are denoted by $\mathcal{F}_{RF}$ and $\mathcal{W}_{RF}$ respectively. For a more detailed explanation with slightly different notation, please refer to article [16]. To simplify and solve the optimization problem quoted in (7), we divide the combined precoder/combiner design issue into two distinct sub-problems by separately estimating the precoders ($\tilde{\mathbf{F}}_A$ and $\tilde{\mathbf{F}}_D$) and combiners ($\tilde{\mathbf{W}}_A$ and $\tilde{\mathbf{W}}_D$). Initially, to find the estimated precoders, we compute all combinations of L_F paths selected from the entire set of transmission paths, generating all conceivable antenna response vectors to construct a precoder matrix. Subsequently, Nt_{RF} columns must be chosen from $\mathbf{a}_T(\Theta_T^{(ij)})$ which achieves the maximum EE when the combiner is taking FD optimal $\mathbf{W}_{opt}$. Then, the estimated precoders are constructed from $\mathcal{F}_{RF}^{(l_F)}$.

Algorithm 1. Energy-efficiency HB for M-MIMO systems

```

1: Input:  $\mathbf{H}, N_{BS}, N_{MS}, Nt_{RF}, Nr_{RF}, N$ .
2: Output:  $\tilde{\mathbf{F}}_A, \tilde{\mathbf{F}}_D, \tilde{\mathbf{W}}_A, \tilde{\mathbf{W}}_D$ 
3: Compute  $L_F = \binom{N_{cl} N_{ray}}{Nt_{RF}}$  AND  $L_W = \binom{N_{cl} N_{ray}}{Nr_{RF}}$ 
4: for  $l_F = 1 : L_F$  do
5:    $\mathbf{F}_A^{(l_F)} = \mathcal{F}_{RF}^{(l_F)}$ ;  $\tilde{\mathbf{F}}_A = \mathbf{F}_A^{(l_F)}$ ;
6:    $\mathbf{F}_D^{(l_F)} = (\tilde{\mathbf{F}}_A^H \tilde{\mathbf{F}}_A)^{-1} \tilde{\mathbf{F}}_A^H \mathbf{F}_{opt}$ ;  $\tilde{\mathbf{F}}_D = \mathbf{F}_D^{(l_F)}$ 
7:    $EE(l_F) = \frac{1}{P_C(l_F)} \log_2 \left| \mathbf{I}_N + \frac{\rho}{N\sigma_n^2} (\mathbf{W}_{opt}^H \mathbf{W}_{opt})^{-1} \mathbf{W}_{opt}^H \mathbf{H} \tilde{\mathbf{F}}_A \tilde{\mathbf{F}}_D \times \tilde{\mathbf{F}}_D^H \tilde{\mathbf{F}}_A^H \mathbf{H}^H \mathbf{W}_{opt} \right|$ 
8: end for
9:  $[\sim, \tilde{l}_F] = \operatorname{argmax} (EE(:, 1))$ 
10:  $\tilde{\mathbf{F}}_A = \mathbf{F}_A^{(\tilde{l}_F)}$ ,  $\tilde{\mathbf{F}}_D = \mathbf{F}_D^{(\tilde{l}_F)}$ 
11: Use the finding  $\tilde{\mathbf{F}}_A$  and  $\tilde{\mathbf{F}}_D$  for calculate  $\tilde{\mathbf{W}}_A$  and  $\tilde{\mathbf{W}}_D$ 
12: for  $l_W = 1 : L_W$  do
13:    $\mathbf{W}_A^{(l_W)} = \mathcal{W}_{RF}^{(l_W)}$ ;  $\tilde{\mathbf{W}}_A = \mathbf{W}_A^{(l_W)}$ 
14:    $\mathbf{W}_D^{(l_W)} = (\tilde{\mathbf{W}}_A^H \mathbf{Y} \tilde{\mathbf{W}}_A)^{-1} (\tilde{\mathbf{W}}_A^H \mathbf{Y} \mathbf{W}_{opt})$ ;  $\tilde{\mathbf{W}}_D = \mathbf{W}_D^{(l_W)}$ 
15:    $\mathbf{Y} = \frac{\rho}{N} \mathbf{H} \tilde{\mathbf{F}}_A \tilde{\mathbf{F}}_D^H \tilde{\mathbf{F}}_A^H \mathbf{H}^H + \sigma_n^2 \mathbf{I}_{N_{MS}}$ 
16:    $EE(l_W) = \frac{1}{P_C(l_F)} \log_2 \left| \mathbf{I}_N + \frac{\rho}{N\sigma_n^2} (\tilde{\mathbf{W}}_D^H \tilde{\mathbf{W}}_A^H \tilde{\mathbf{W}}_A \tilde{\mathbf{W}}_D \times \tilde{\mathbf{W}}_A^H \tilde{\mathbf{W}}_D^H)^{-1} \times \mathbf{H} \tilde{\mathbf{F}}_A \tilde{\mathbf{F}}_D \times \tilde{\mathbf{F}}_D^H \tilde{\mathbf{F}}_A^H \mathbf{H}^H \tilde{\mathbf{W}}_A \tilde{\mathbf{W}}_D \right|$ 
17: end for
18:  $[\sim, \tilde{l}_W] = \operatorname{argmax} (EE(:, 1))$ 
19:  $\tilde{\mathbf{W}}_A = \mathbf{W}_A^{(\tilde{l}_W)}$ ,  $\tilde{\mathbf{W}}_D = \mathbf{W}_D^{(\tilde{l}_W)}$ 
20:  $[\sim, \tilde{l}_F] = \operatorname{argmax} (EE(:, 1))$ 

```

The FD optimal precoder and combiner are denoted $\mathbf{F}_{opt}$ and $\mathbf{W}_{opt}$, respectively. The channel matrix $\mathbf{H}$ is processed to singular value decomposition (SVD), in order to make $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H$. Leveraging the mentioned decomposition, [3] indicates that $\mathbf{F}_{opt}$ corresponds to the first N columns of $\mathbf{V}$ as $\mathbf{F}_{opt} = \mathbf{V}^{(1)}$ which can be used to obtain the FD optimal precoding matrix and $\mathbf{W}_{opt}$ can be calculated using the unconstrained beamformer $\mathbf{F}_{opt}$ as (8):

$$\mathbf{W}_{opt} = \left(\frac{1}{\rho} \left(\mathbf{F}_{opt}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{opt} + \frac{N\sigma_n^2}{\rho} \mathbf{I}_N \right)^{-1} \mathbf{F}_{opt}^H \mathbf{H}^H \right)^H \quad (8)$$

We then estimate hybrid combiners using the findings of the hybrid precoders obtained through the same methodology. The estimated combiners are constructed from $\mathcal{W}_{RF}^{(IF)}$. The optimization sub-problem for HB is described in Algorithm 1.

2.4. CNN-based beamformers design

In this section, we present the architecture of the proposed network for joint analog precoder and combiner optimization, depicted in Figure 2. This figure illustrates the architecture comprising two CNNs. One of them focuses on optimizing precoders by extracting essential features from the channel matrix. It generates optimal phases for the precoder, steering signal power in desired directions, ultimately enhancing EE. The other CNN is dedicated to combiner design. It processes channel data through convolutional layers, generating phases to optimize combiners. This leads to efficient signal reception, contributing to overall EE improvement. Each CNN comprises eight layers and has a similar structure, except for the final layer. In data generation, to enhance the processing capability of imperfect channel state information (CSI), we first construct randomly P_{ch} perfect channel matrices $\mathbf{H}^{(P_{ch})}$ for different user locations. Then, we adopt L noisy channel matrices for each generated perfect channel matrix, introducing element-wise synthetic noise. The level of synthetic noise is determined by the formula:

$$SNR_{TRAIN} = 20 \log_{10} \left(\frac{|\mathbf{H}_{i,j}|^2}{\sigma_{Train}^2} \right),$$

where σ_{Train}^2 represents the variance of the noise in the training phases associated with the channel component $[\mathbf{H}]_{i,j}$. To account for variations in the wireless environment, we use three different levels of SNR_{TRAIN} . These values (15, 20, and 25 dB) reflect realistic training scenarios ranging from low to moderate SNR conditions, as adopted in [16]. Prior to feeding the complex-valued channel matrix $\mathbf{H}$ into the real-valued neural networks, we further modify it to facilitate feature extraction and enhance performance. Thus, we go for an input layer of size $N_{MS} \times N_{BS} \times 3$. The first input $[\mathbf{X}]_{:,1} = |\mathbf{H}|_{i,j}$ is the element-wise absolute value of the channel matrix. $[\mathbf{X}]_{:,2} = \text{Re}\{\mathbf{H}\}_{i,j}$ and $[\mathbf{X}]_{:,3} = \text{Im}\{\mathbf{H}\}_{i,j}$ are respectively, the second and third inputs. They represent the real and imaginary parts of the channel matrix. The matrix $[\mathbf{X}]_{:,1:3}$ is the three-dimensional input data of the network.

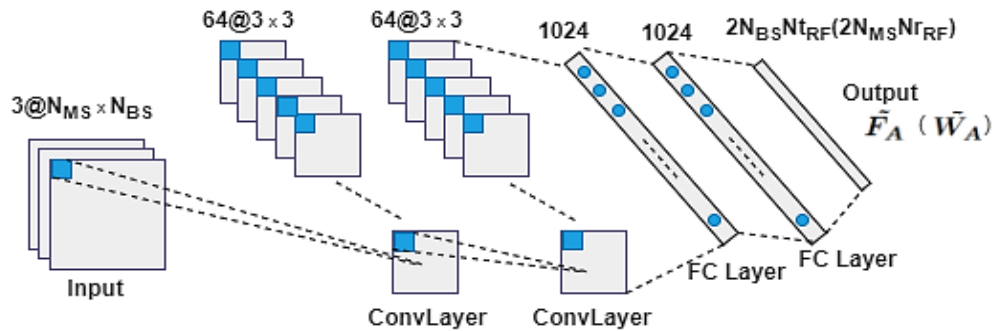


Figure 2. Proposed CNN model for joint beamformers design

The rest of the network structure includes convolutional layers with N_{filt} filters of size $k \times k$ used in the second and third layers to extract and select data feature vectors. The activation functions in the convolutional layers are all rectified linear unit (ReLU) functions. The fourth and sixth layers are FC with N_{FC} units. To prevent overfitting, dropout layers with a p_{drop} probability are included after FC layers, specifically in the fifth and seventh layers. The output layer of the CNN_F is based on the vectorized form of the $\mathbf{F}_A$ phases, resulting in a size of $N_{BS} N_{t_{RF}} \times 1$. Similarly, the output layer of CNN_W has a size of $N_{MS} N_{r_{RF}} \times 1$. The generated data for $P_{ch} = x_1$ and $L = x_2$ feed the CNNs during the training and validation

phases. The training and validation datasets are created by randomly dividing the total data into σ_{train} for training and $1 - \sigma_{train}$ for validation. The Adam optimization Algorithm is used to train the model.

Algorithm 2. CNN-EE based HB optimizer

```

1: Input:  $L, P_{ch}, N_{BS}, N_{MS}, Nt_{RF}, Nt_{RF}, SNR_{TRAIN}$ .
2: Output: Training data  $CNN_F$  and  $CNN_W$ 
3: Generate  $\{\mathbf{H}^{(p)}\}_{p=1}^{P_{ch}}$ 
4: for  $1 \leq p \leq \mathbf{H}^{(p_{ch})}$  and  $1 \leq l \leq L$  do
5:  $[\mathbf{H}^{(l,p)}]_{i,j} \sim \mathcal{N}([\mathbf{H}^l]_{i,j}, \sigma_{TRAIN}^2)$ .
6: Use Algorithm 1 to get  $\tilde{F}_A^{(l,p)}$ , and  $\tilde{W}_A^{(l,p)}$  as labels
7:  $labels_F^{(l,p)} = \angle vec\{\tilde{F}_A\}$ ,  $labels_W^{(l,p)} = \angle vec\{\tilde{W}_A\}$ ,
8: Input data:
9:  $[[X^{(l,p)}]_{:,1}]_{i,j}$ ,  $[[X^{(l,p)}]_{:,2}]_{i,j}$  and  $[[X^{(l,p)}]_{:,3}]_{i,j}$ 
10: Construct the input-output pair  $(X^{(l,p)}, labels_F^{(l,p)})$  for  $CNN_F$  and  $(X^{(l,p)}, labels_W^{(l,p)})$  for  $CNN_W$ 
11: end for  $p$  and  $l$ 
12: training data for  $CNN_F$  and  $CNN_W$ 

```

3. METHODOLOGY

3.1. Simulation setup

This work is implemented in MATLAB R2022a, where a hybrid Beamforming using DL approaches based on the EE criterion (M-MIMO EE) system designed using the following parameters $N_{BS} = 64$, $N_{MS} = 16$, $SNR_{TEST} = 10$ dB and $f_c = 28$ GHz. The propagation channel environment is modeled with $N_{cl} = 4$ and $N_{ray} = 4$ for each cluster, and $\sigma^2 = 5^\circ$ for all transmit and receive azimuth and elevation angles, randomly chosen within the intervals $[-60^\circ, 60^\circ]$ and $[-20^\circ, 20^\circ]$. This paper utilizes actual values $P_{LO} = 22.5$ mW, $P_{MX} = 0.3$ mW, $P_{LPF} = 14$ mW, $P_{HB} = 3$ mW and $\eta = 0.27$ [12]. The study examines the effect of different RF chain configurations, specifically testing systems with 4 RF chains and 6 RF chains at both the transmitter and receiver sides. The simulation evaluates the system's performance based on EE criterion (M-MIMO EE) system with FD precoding (iFullOPT) solution and SE deep learning based hybrid beamforming design solution based on SE (M-MIMO SE) as presented in [16]. Additionally, we perform a comparative analysis of various MIMO setups, including 4×4 MIMO, 9×9 MIMO, and a large-scale 64×16 M-MIMO system.

3.2. CNN training and implementation

The CNN input dataset, previously introduced in section 2.4, is split into a 70/30 training-validation ratio. The model is trained using the Adam optimizer, with the configuration summarized in Table 1.

Table 1. CNN training hyperparameters and structural settings

Symbol	Description	Value
P_{ch}	Perfect channel realizations	8.6
L	Noisy augmentations per realization	100
σ_{train}	Training data fraction	70 %
$1 - \sigma_{train}$	Validation data fraction	30 %
μ	Adam learning rate	0.0005
B	Batch size	100
N_{epoch}	Training epochs	200
N_{filt}	Conv. filters per layer	64
$k \times k$	Filter kernel size	3×3
N_{FC}	Neurons in fully connected layers	1024
p_{drop}	Dropout probability	50 %

4. RESULTS AND DISCUSSION

This section begins by analyzing the scalability of M-MIMO and provides insights into the trade-offs between RF chain optimization and system efficiency, emphasizing the advantages of M-MIMO for enhancing energy performance in modern communication systems. Figure 3 presents the performance of a M-MIMO EE system with $N_{BS} = 64$ and $N_{MS} = 16$ under two RF chain configurations: $Nt_{RF} = 4$ and $Nt_{RF} = 6$. Figure 3(a) compares the SE versus SNR, while Figure 3(b) depicts the EE versus SNR for both configurations. Figure 3(a) illustrates that SE increases linearly with the SNR, as expected. A zoomed-in region highlights that the configuration with 4 RF chains achieves slightly higher SE compared to 6 RF

chains, emphasizing the impact of RF chain optimization on system performance. Figure 3(b) plots EE versus SNR for the same configurations, revealing that $Nt_{RF} = 4$ outperforms $Nt_{RF} = 6$ due to the reduced power consumption associated with fewer RF chains. This analysis highlights a trade-off: reducing the number of RF chains improves EE without significantly compromising SE, making $Nt_{RF} = 4$ a more EE and practical choice for M-MIMO systems.

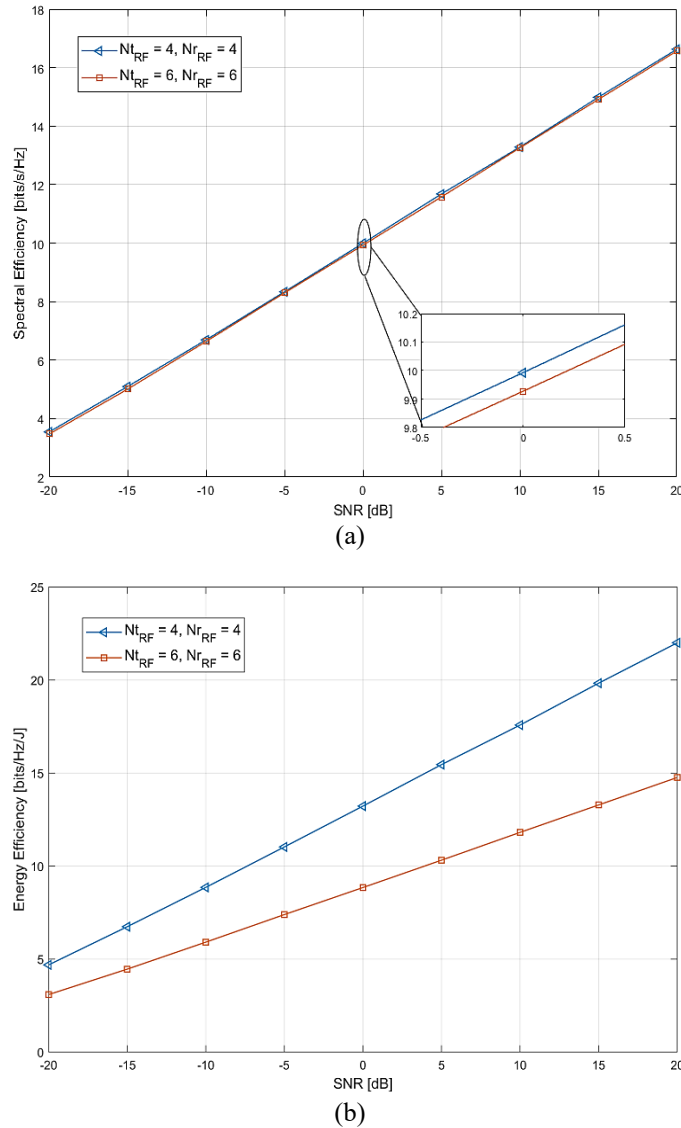


Figure 1. Performance analysis of different RF chain configurations in M-MIMO with $N_{BS} = 64$, $N_{MS} = 16$. (a) SE versus SNR and (b) EE versus SNR

Figure 4 shows the EE performance versus SNR for M-MIMO EE, M-MIMO SE, and iFullOPT. The M-MIMO EE method achieves the best EE across all SNR levels, highlighting its superior optimization for energy consumption. In contrast, iFullOPT demonstrates very low EE due to the high-power consumption of RF chains required for digital precoding. This highlights the advantage of M-MIMO EE, which leverages HP to reduce RF chain usage and improve EE.

Figure 5 compares the EE performance of 4×4 MIMO, 9×9 MIMO, and 64×16 M-MIMO. The results clearly show that M-MIMO EE with a larger antenna array (64×16) outperforms smaller configurations, achieving higher EE across all SNR values. This improvement is particularly significant at low SNR, where M-MIMO EE benefits from HP to optimize energy usage and reduce RF chain consumption. As the SNR increases, the advantage of M-MIMO EE becomes even more apparent, proving its efficiency and scalability for large-scale MIMO systems.

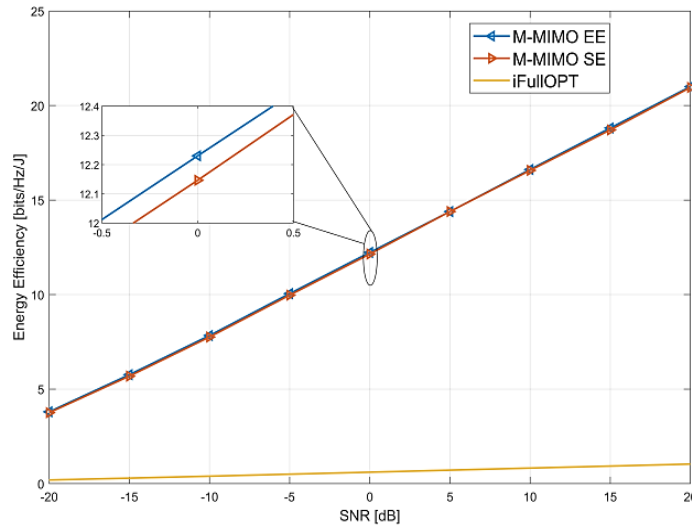


Figure 4. EE performance vs SNR for M-MIMO EE, M-MIMO SE, and iFullOPT

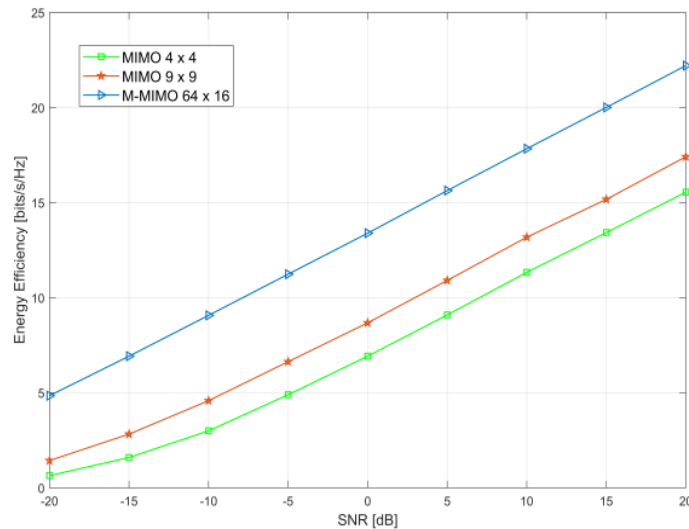


Figure 5. Comparative evaluation of MIMO systems: 4×4, 9×9, and 64×16 configurations

5. CONCLUSION

In this work, we presented an energy-efficient approach for hybrid beamforming in mmWave M-MIMO systems, with a particular focus on maximizing EE. Simulation results demonstrate that the proposed M-MIMO EE configuration particularly with a 64×16 antenna setup significantly outperforms smaller MIMO systems in terms of EE. Moreover, the analysis showed that optimizing the number of RF chains, such as using 4 RF chains instead of 6, offers a notable improvement in EE without a substantial loss in SE. These findings underline the potential of EE-based approaches in enhancing the performance and scalability of Massive MIMO systems, making them a promising solution for future energy-efficient communication networks. For future work, the consideration of multiuser scenarios with reconfigurable intelligent surfaces can be explored, with attention to the energy contribution of each element.

ACKNOWLEDGMENTS

We express our deep appreciation to the LTT Laboratory at the University Abou Bakr Belkaïd, Tlemcen, Algeria, for providing the facilities and collaborative environment that made this research possible. Also, the authors gratefully acknowledge the support of CEDRIC Laboratory (LAETITIA team) of the Conservatoire National des Arts et Métiers (CNAM) in Paris, France.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hanane Ayad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓
Mohammed Yassine Bendimerad	✓	✓		✓	✓	✓				✓		✓	✓	
Fethi Tarik Bendimerad				✓	✓					✓		✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing -Original Draft

E : Writing - Review &Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

- [1] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: a comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021, doi: 10.1109/OJCOMS.2021.3057679.
- [2] N. Telagam, N. Kandasamy, A. K. Manoharan, P. Anandhi, and R. Atchudan, "Beyond 5G: exploring key enabling technologies, use cases, and future prospects of 6G communication," *Nano Communication Networks*, vol. 43, p. 100560, 2025, doi: 10.1016/j.nancom.2024.100560.
- [3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014, doi: 10.1109/TWC.2014.011714.130846.
- [4] J. Lv, T. Wang, and S. Wang, "Optimal analog precoder design for hybrid beamforming is possible," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 9573–9578, Jul. 2023, doi: 10.1109/TVT.2023.3245091.
- [5] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in Millimeter Wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, Apr. 2016, doi: 10.1109/JSTSP.2016.2523903.
- [6] H.-H. Tseng, Y.-F. Chen, and S.-M. Tseng, "Hybrid beamforming and resource allocation designs for mmWave multi-user massive MIMO-OFDM systems on Uplink," *IEEE Access*, vol. 11, pp. 133070–133085, 2023, doi: 10.1109/ACCESS.2023.3335278.
- [7] H. Nosrati, E. Aboutanios, X. Wang, and D. Smith, "Switch-based hybrid beamforming for massive MIMO communications in mmWave bands," *Signal Processing*, vol. 200, p. 108659, Nov. 2022, doi: 10.1016/j.sigpro.2022.108659.
- [8] S. Kamiwatari, I. Kanno, T. Hayashi, and Y. Amano, "RF chain-wise clustering schemes for millimeter Wave cell-free massive MIMO With centralized hybrid beamforming," *IEEE Access*, vol. 12, pp. 19682–19693, 2024, doi: 10.1109/ACCESS.2024.3360716.
- [9] S. M. Bahri and A. Bouacha, "Hybrid digital and analog beamforming design using genetic algorithms," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 6, pp. 6389–6400, 2024, doi: 10.11591/ijece.v14i6.pp6389-6400.
- [10] H. S. Vu, K. T. Truong, and M. T. Le, "Beam division multiple access for millimeter wave massive MIMO: Hybrid zero-forcing beamforming with user selection," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 445–452, 2022, doi: 10.11591/ijece.v12i1.pp445-452.
- [11] N. T. Nguyen, V.-D. Nguyen, H. Van Nguyen, H. Q. Ngo, S. Chatzinotas, and M. Juntti, "Spectral efficiency analysis of hybrid relay-reflecting intelligent surface-assisted cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3397–3416, May 2023, doi: 10.1109/TWC.2022.3217828.
- [12] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, May 2018, doi: 10.1109/JSTSP.2018.2824762.
- [13] Z. Amirifar and J. Abouei, "An improvement to hybrid beamforming precoding scheme for mmWave massive MIMO systems based on channel matrix," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, p. 285, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp285-292.
- [14] A. M. Elbir, A. Celik, and A. M. Eltawil, "Hybrid beamforming for integrated sensing and communications with low resolution DACs," *IEEE Wireless Communications Letters*, vol. 14, no. 1, pp. 103–107, 2025, doi: 10.1109/LWC.2024.3489632.
- [15] O. Islam, M. Rihan, M. Elhefnawy, and S. Eldolil, "Deep learning based hybrid precoding technique for millimeter-Wave massive MIMO systems," in *2021 International Conference on Electronic Engineering (ICEEM)*, 2021, pp. 1–5, doi: 10.1109/ICEEM52022.2021.9480386.
- [16] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Communications Letters*, vol. 23,

- no. 7, pp. 1240–1243, Jul. 2019, doi: 10.1109/LCOMM.2019.2915977.
- [17] F. Liu, L. Zhang, X. Yang, T. Li, and R. Du, "DL-based energy-efficient hybrid precoding for mmWave massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6103–6112, May 2023, doi: 10.1109/TVT.2022.3230931.
 - [18] K. Chen, J. Yang, Q. Li, and X. Ge, "Sub-array hybrid precoding for massive MIMO systems: A CNN-based approach," *IEEE Communications Letters*, vol. 25, no. 1, pp. 191–195, Jan. 2021, doi: 10.1109/LCOMM.2020.3022898.
 - [19] Q. Wang, K. Feng, X. Li, and S. Jin, "PrecoderNet: hybrid beamforming for millimeter Wave systems with deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1677–1681, Oct. 2020, doi: 10.1109/LWC.2020.3001121.
 - [20] N. T. Nguyen *et al.*, "Deep unfolding hybrid beamforming designs for THz massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3788–3804, 2023, doi: 10.1109/TSP.2023.3322852.
 - [21] Z. Liu, Y. Yang, F. Gao, T. Zhou, and H. Ma, "Deep unsupervised learning for joint antenna selection and hybrid beamforming," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1697–1710, Mar. 2022, doi: 10.1109/TCOMM.2022.3143122.
 - [22] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Flexible unsupervised learning for massive MIMO subarray hybrid beamforming," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec. 2022, pp. 3833–3838, doi: 10.1109/GLOBECOM48099.2022.10001155.
 - [23] P. K. Gkonis, "A survey on machine learning techniques for massive MIMO configurations: application areas, performance limitations and future challenges," *IEEE Access*, vol. 11, pp. 67–88, 2023, doi: 10.1109/ACCESS.2022.3232855.
 - [24] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: phase shifters or switches?," in *2015 Information Theory and Applications Workshop (ITA)*, 2015, pp. 90–95, doi: 10.1109/ITA.2015.7308971.
 - [25] W. Bin Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter Wave receiver efficiency: a comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8131–8146, 2017, doi: 10.1109/TWC.2017.2757919.
 - [26] K. Gao, N. J. Estes, B. Hochwald, J. Chisum, and J. N. Laneman, "Power-performance analysis of a simple one-bit transceiver," in *2017 Information Theory and Applications Workshop (ITA)*, 2017, pp. 1–6, doi: 10.1109/ITA.2017.8023454.

BIOGRAPHIES OF AUTHORS



Hanane Ayad    is currently pursuing a Ph.D. in telecommunications at the University of Tlemcen, Algeria. She received her M.Sc. degree in networks and telecommunications in 2018. Her research interests include HP, mmWave massive MIMO, and deep learning techniques for wireless communications. She completed a research internship at CNAM, Paris, under the TASSILI ATOME5+ project, and has authored papers presented at international conferences. She can be contacted at email: ayad.hanane@univ-tlemcen.dz.



Mohammed Yassine Bendimerad    received his B.Sc. degree in electrical and electronic engineering, as well as his M.Sc. and Ph.D. degrees in telecommunication and wireless communication technologies from the University of Tlemcen, Algeria, in 2010, 2012, and 2016, respectively. In 2016, he joined the University of Bechar as an assistant professor. He is currently an associate professor in the Telecommunications Department and a member of the Digital Communication Team at the LTT Laboratory of Telecommunications, University of Tlemcen. His research interests include wireless communication and mobile networks, energy efficiency in wireless systems, and artificial intelligence for optimizing communication protocols. He has authored several papers in IEEE-indexed journals and conferences and participates in national and international research projects (PRFU, PHC-TASSILI, ACADEMY). He can be contacted at email: yassine.bendimerad@univ-tlemcen.dz.



Fethi Tarik Bendimerad    received the Engineering degree in electronics from the University of Science and Technology of Oran (USTO), Algeria, in 1983, the Diplôme d'Études Approfondies (DEA) in telecommunications from the University of Nice–Sophia Antipolis, France, in 1984, and the Ph.D. degree in telecommunications from the same university in 1989. His Ph.D. was officially recognized as equivalent to the Doctorat d'État in June 1992. He is currently a professor with the Faculty of Engineering, University of Tlemcen, Algeria. He is also the Director of the Telecommunications Research Laboratory at the University of Tlemcen, and previously served as the director of the Institute of Electronics at the same university. His research interests include wireless communications, signal processing, and advanced telecommunication systems. He has supervised several Ph.D. students and contributed to numerous national and international research projects. He is the author of several publications in reputable journals and conferences and regularly serves as a reviewer for scientific journals. He can be contacted at email: fethitarek.bendimerad@univ-tlemcen.dz.