Deep Learning-based Compensation for Doppler Shifts in Hybrid Beamforming for mmWave Communication

Kartik Ramesh Patel and Sanjay Dasrao Deshmukh

MCT's Rajiv Gandhi Institute of Technology, Mumbai, India

https://doi.org/10.26636/jtit.2025.4.2349

Abstract — Millimeter-wave (mmWave) communication is a key enabler of 5G and future wireless systems, providing vast bandwidth for high-speed data transfers. However, high user mobility leads to significant Doppler shifts, which can severely degrade the performance of beamforming – an essential technology for mmWave systems. The traditional hybrid beamforming (HBF) technique faces challenges in adapting to rapid channel variations caused by Doppler effects. Therefore, this paper introduces a deep learning framework to mitigate Doppler-induced channel distortions in hybrid beamforming. We propose a long-short-term memory (LSTM)-based neural network that predicts Doppler shifts and dynamically adjusts the hybrid beamforming vectors to compensate for these variations. This approach proactively addresses channel distortion, enhancing both spectral and energy efficiency. The simulation results and the performance comparison of proposed model against conventional beamforming and state-of-the-art techniques demonstrate the superiority of deep learning-based solution in maintaining robust communication links under high-mobility conditions, showcasing its potential to improve performance in next-generation wireless networks.

Keywords — Doppler shift, hybrid beamforming LSTM, mmWave, spectral efficiency

1. Introduction

The huge growth in mobile data traffic, driven by applications such as ultra-high definition video, autonomous systems, and the Internet of Things (IoT) [1] has resulted in demand for unprecedented data rates and ultra-low communication latency. This demand has propelled wireless communication into the millimeter-wave (mmWave) (30 – 300 GHz) spectrum, [2] a frontier defined by its vast contiguous bandwidth.

However, the mmWave spectrum presents physical challenges, such as severe loss of path loss and susceptibility to blockage [3]. To overcome these limitations, large-scale antenna arrays are employed to achieve high-gain beamforming, focusing signal energy into narrow, directional beams to establish and maintain a robust communication link. To manage costs, hardware complexity and power consumption with fully digital beamforming, hybrid beamforming (HBF) has emerged as a consensus energy-efficient architecture [4]. By partitioning the beamforming task between a high-dimensional analog domain (using phase shifters) and a low-dimensional digital domain (at baseband), HBF provides a balance of array gain and system cost.

Despite its architectural merits, the efficacy of HBF depends on accurate and real-time channel state information (CSI). In high-mobility scenarios, such as vehicle-to-everything (V2X) communication, high-speed rail, and drone networks, this contingency becomes a bottleneck [5]. The movement between the transmitter and the receiver causes significant Doppler shifts manifesting as rapid phase variations in the channel [6]. This phenomenon causes the channel to decompose over time, invalidating the "quasi-static" assumption upon which conventional, reactive beam-tracking algorithms and codebook-based HBF solutions are built. It leads to severe performance degradation, inter-channel interference, and even link failure.

Coherence time shrinks at highway speeds, making reactive beam updates insufficient [2], while classical Kalman filter (KF) and extended Kalman filter (EKF) [7] trackers update angles and recent deep learning works often predict beam indices, they do not incorporate the hybrid analog-digital pair at each slot from a predicted complex channel, thus leaving a gap in Doppler-aware HBF design.

Researchers identified the time-series nature of the problem, applying recurrent neural networks (RNN) [8] and their variants, such as LSTM and gated recurrent units (GRU), to predict future channel states [9].

Despite that, the research gap persists. Most current works focus on the prediction of the full-dimensional, unconstrained CSI matrix. There is still a significant disconnect between this high-dimensional prediction and its practical, real-time application within the constrained HBF architecture. Recently, research has moved to exploring alternatives, such as deep reinforcement learning (DRL) for policy-based beam control [10] and transformer-based models for long-range temporal dependency. Although promising, DRL models can suffer from training instability, and transformers carry a significant computational overhead.

Conventional hybrid beamforming techniques, which rely on the assumption of a quasi-static channel, fail in high-mobility mmWave environments due to rapid channel decorrelation [11]. Although recent deep learning models demonstrate the ability to predict channel variations, there is a re-

search gap in developing a framework that efficiently and proactively integrates the predictive information into the physically constrained and latency-constrained hybrid beamforming architecture in order to maintain robust, high-throughput communication links [12].

In such a context, we propose a pipeline type, where an LSTM predicts $\widehat{\mathbf{H}}(t+1)$ from the past K channels, and next that prediction drives analog and digital stages under constant module constraints.

Recent works on mobility-robust mmWave links mainly tracks beams with state-space filters, i.e. KF/EKF or learns beam indices directly with neural networks, often without redesigning the full hybrid analog-digital chain under motion.

The contribution of this paper can be summarized as follows:

- The impact of high Doppler shifts on the time-varying mmWave channel is modeled and effect quantified on a conventional hybrid beamforming system.
- The predict-then-design deep learning framework is used, centered on an LSTM-based predictive engine, which proactively forecasts the evolution of the mmWave channel and directly computes the required compensatory HBF vectors.
- A dynamic HBF compensation algorithm is implemented that translates the LSTM predictive output into real-time, constrained analog and digital beamforming commands.
- A comprehensive simulation-based performance evaluation is presented, benchmarking the proposed framework against conventional (extended Kalman filter-based) and orthogonal matching pursuit (OMP) methods in terms of spectral efficiency, energy efficiency, and link robustness under various high-mobility scenarios.

The paper is organized as follows. Section 2 reviews related work. Section 3 presents the method concerned, while Section 4 discusses the simulation results and evaluates the performance. Section 5 concludes the paper.

2. Related Work

The authors of [7] provide a insight into extended Kalman filters, which are solution for non-linear tracking problems. In this context, EKF works in a two-step predict-update loop. It uses the system's mobility model to predict the next channel state and the actual received pilot signal to correct the prediction. Particle filters are a more robust, non-parametric alternative to EKF. They represent the channel state not as a single estimate with covariance (like EKF), but as a cloud of thousands of weighted particles [7].

In both approaches, the computational complexity is extremely high. EKFs involve large matrix inversions at every time step, and particle filters are more brittle, relying on an accurate pre-defined mathematical model of the user's movement, which is not suitable in highly dynamic environments.

A hierarchical beam sweeping method is proposed in [12], where instead of estimating the full channel matrix, the system simply tries to find the best pre-set beam. A hierarchical search avoids testing all beams. First, it uses wide beams to find the general direction, and then zooms in with progressively narrow beams. Unfortunately, such methods are difficult to adapt. A high-speed user can travel out the optimal beam in the time it takes the algorithm to complete its sweep-and-search process. Hence, this method is fundamentally reactive. It can only fix a beam misalignment after it has already occurred and caused performance degradation.

The importance of using the HBF architecture is explained in [4]. The analog beamformer has a large matrix of simple, low-cost phase shifters but does the coarse beam steering, while the digital beamforming has an expensive single antenna element, dedicated, and power-hungry radio frequency chain. Therefore, the hybrid beamforming architecture is considered a compromise between analog and digital architecture with few RF chains.

The HBF architecture using codebook-based precoding is described in [13]. The codebook is a predefined set of high-performance analog and digital beamforming vectors. The system tracking job is reduced by selecting the best index of this codebook. The problem with this method lies in the codebooks designed for static and slowly varying channels. The entire codebook is generated offline, assuming the channel is stable. The Doppler effect breaks this assumption. The true optimal beam for high-speed user will likely lie in between any two precalculated beams in the codebook, leading to a constant, suboptimal quantization error.

The authors in [10] mentioned that modern wireless channels are so complex (with blocks, reflections, and mobility) that model-based approaches from [7], [12], are no longer feasible. They conclude that machine learning and deep learning act as universal function approximates to learn the complex mapping from received pilots to optimal beam directly from data, without needing a perfect mathematical model.

Paper [6] is the first attempt to apply deep learning to the problem. A convolutional neural network (CNN) is designed to treat the channel matrix $\mathbf{H}(t)$ as an image. CNNs are good at finding spatial features with the ability to detect, in the channel matrix, patterns such as dominant paths and their angles to determine the best beam at that instant. The downside of this method is that CNNs are not inherently designed to capture temporal dependencies. With no memory of the past, a CNN cannot see the user movement trend and, therefore, cannot predict the future.

The method of time series forecasting based on the requirement of the dynamic wireless channel is presented in [8] and [9], which is exactly why RNNs like LSTM and GRU are built. Unlike CNN, the LSTM has a memory that allows it to process a sequence of past channels to learn the underlying dynamics of the user's motion. The goal of these methods is purely prediction-oriented and they prove to be highly accurate [9].

The critical gap in the methods described in [7] and [8] is that they do not specifically address the integration of these predictions into the hybrid beamforming framework for Doppler mitigation. Both end at the prediction of $\widehat{\mathbf{H}}(t+1)$, but they do not provide an answer as to what is going to be

Tab. 1. Related literature vs. proposed method.

Approach	Predicts/tracks	Uses prediction to design both analog and digital HBF each slot	Mobility focus/Doppler	Remarks
EKF beam tracking [7]	Angles/kinematics	No	Vehicular EKF with Jacobians	Model-driven; low complexity; sensitive to mismatch
DL beam tracking [6]	Beam index	No	Mobility under sounding	High accuracy; not coupled to the hybrid precoder/combiner
Sub-6 mmWave beam/blockage [8]	Beam/blockage	No	Mobility/robustness	Multiband features; no HBF co-design
DL-HBF surveys/reviews [11]	–	Varies	Discuss challenges	Surveys DL for HBF; limited Doppler-aware co-design exemplars
This work (LSTM-HBF)	Channel (complex)	Yes	Explicit Doppler via sequence prediction	Adds the SE-loss bound, complexity, and full reproducibility

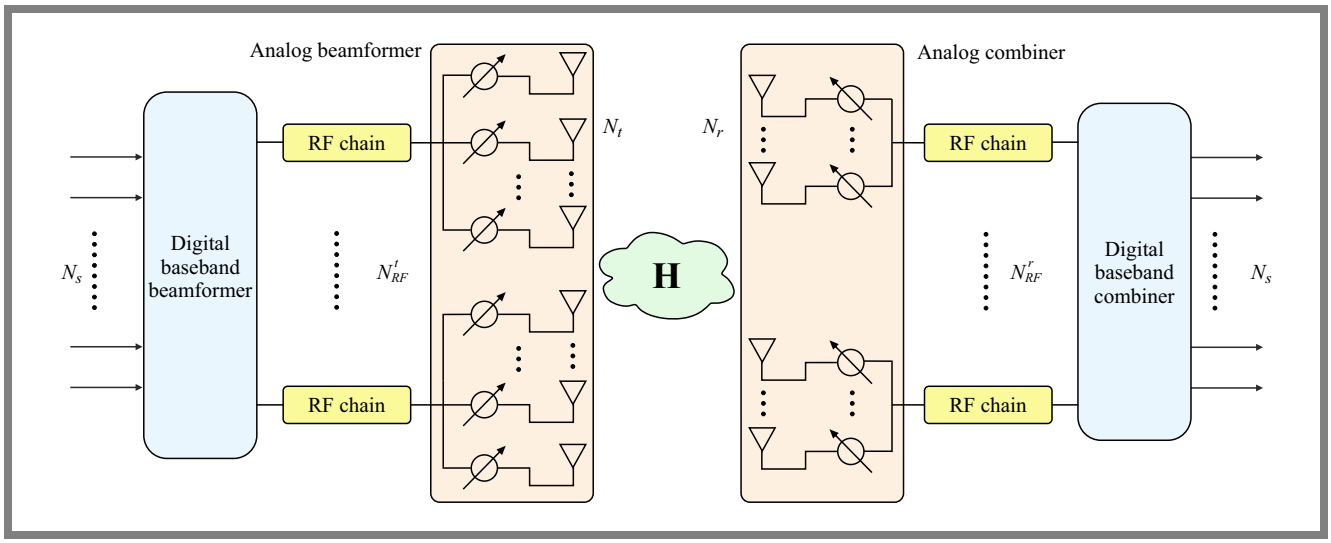

Fig. 1. Hybrid analog-digital architecture with N_t antennas and N_{RF}^t RF chains at BS and N_r, N_{RF}^r at UE side.

done next. There is a gap in how to use this predicted full-digital matrix $\widehat{\mathbf{H}}(t+1)$ to calculate constrained analog $\mathbf{F}_{RF}$ and digital $\mathbf{F}_{BB}$ matrices. These calculations have to be done quickly enough to proactively mitigate the Doppler effect.

The proposed method is based on these foundations, but it is distinct in its focus on proactively mitigating the Doppler effect in hybrid beamforming using a predictive LSTM model. It aims to bridge the gap between deep learning-based channel prediction and practical hybrid beamforming design for high-mobility mmWave systems. Table 1 shows the comparison of the proposed method with other articles.

3. Research Methodology

The proposed system consists of a massive mmWave MIMO block with a hybrid beamforming architecture, as shown in Fig. 1. The time-varying channel $\mathbf{H}(t)$ is modeled as a sum of the multiple paths (multipath fading). Each path is characterized by a Doppler shift $f_{D,l}$ which accounts for the relative motion of the UE and the BS. The Doppler shift $f_{D,l}$ for the l-th path is modeled using relative velocity v and angle of arrival θ_i. This is important for high-mobility scenarios, where the Doppler shift significantly impacts the channel. Matrices $\mathbf{F}_{RF}$ (analog precoder) and $\mathbf{W}_{RF}$ (analog combiner) are based on the SVD of the time-varying channel matrix to align the strongest eigen modes.

3.1. System Model

We consider a single-cell downlink scenario with a base station (BS) equipped with a uniform linear array (ULA) N_{BS} or N_t antenna. The mobile user (UE) has a ULA of N_{UE} or N_r antennas. BS uses a hybrid beamforming structure with N_{RF} RF chains, where $N_{RF} \ll N_{BS}$. Let N_t and N_r denote the numbers of transmit and receive antennas, N_{RF}^t and N_{RF}^r the number of RF chains. The N_s number of data streams, where $N_s \leqslant \min N_{RF}^t$.

The analog precoder/combiner is defined as $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_{RF}^t}$, $\mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N_{RF}^r}$ with constant-modulus en-

tries:

$$\left|\left[F_{\mathrm{RF}}\right]_{ij}\right| = \frac{1}{\sqrt{N_t}}, \quad \left|\left[W_{\mathrm{RF}}\right]_{ij}\right| = \frac{1}{\sqrt{N_r}}.$$

The digital baseband precoder/combiner is $\mathbf{F}_{BB} \in \mathbb{C}^{N_t \times N_{RF}^t}$, $\mathbf{W}_{BB} \in \mathbb{C}^{N_{RF}^r \times N_s}$ and the transmit symbol vector $\boldsymbol{s} \in \mathbb{C}^{N_s}, \mathbb{E}\left[ss^H\right] = \frac{P}{N_s} I$ with the noise $n \sim \mathcal{CN}\left(0, \sigma_n^2 I\right)$. The array response vector for a ULA can be defined as:

$$\mathbf{a}_N(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\pi d \sin\theta}, e^{j2\pi d \sin\theta}, \ldots, e^{j\pi(N-1)\sin\theta}\right]^T, \tag{1}$$

where N is the number of antennas and θ is the physical angle of departure or arrival.

Equation (1) defines the response vector for a ULA with N elements, where each element of the array introduces a phase change relative to others. The phase change is proportional to the distance between the antenna's elements, the wavelength of the signal, and the angle of arrival or departure.

The antennas are assumed to be in a uniform linear configuration, which is typical for MIMO systems, and the phase shifts depend on the angle of arrival (or departure) relative to the array axis.

The signal received at the UE can be modeled as:

$$y(t) = \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H}(t) \mathbf{F}_{RF} \mathbf{F}_{BB} \boldsymbol{s}(t) + \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \boldsymbol{n}(t), \tag{2}$$

where:

- $\mathbf{H}(t) \in \mathbb{C}^{N_r \times N_t}$ is the time-varying mmWave channel matrix,
- $\mathbf{F}_{RF} \in \mathbb{C}^{N_{BS} \times N_{RF}}$ and $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ are the analog and digital precoders at the BS, respectively, where N_s is the number of data streams,
- $\mathbf{W}_{RF} \in \mathbb{C}^{N_{UE} \times N_{RF}}$ and $\mathbf{W}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ are the analog and digital combiners at the UE side,
- $\boldsymbol{s} \in \mathbb{C}^{N_s \times 1}$ is the transmitted symbol vector,
- $\boldsymbol{n}(t) \sim \mathcal{CN}\left(0, \sigma_n^2 I\right)$ is the additive white Gaussian noise.

Equation (2) defines the hybrid beamforming system model, where the signal received at the UE is a combination of the transmitted signal through the analog precoding and digital precoding matrices at the BS as well as the analog and combining and digital combining matrices at the UE. Additive white Gaussian noise (AWGN) is assumed at the receiver with zero mean and variance σ_n^2 [13]. The system uses hybrid beamforming with separate analog and digital precoding/combining matrices to achieve power and phase control.

Figure 2 shows the concept of Doppler effect with an object moving at relative velocity v and the path angle is defined as θ between the source and direction of motion of the object.

The Doppler effect causes channel matrix $\mathbf{H}$ to vary over time. We model such a channel using a clustered geometric model, such as the Saleh-Valenzuela (SV) model [14] (Fig. 3). The channel matrix $\mathbf{H}(t)$ at time t is a superposition of the L scattering paths:

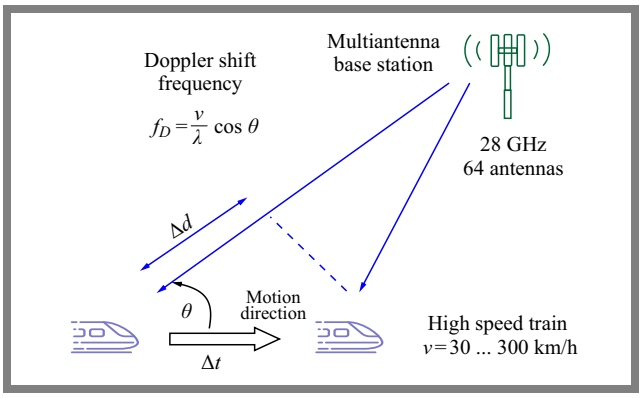

Fig. 2. Relative velocity v, path angle θ and Doppler shift.

$$\mathbf{H}(t) = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l e^{j2\pi f_{D,l} t} \boldsymbol{a}_{N_r}\left(\theta_l^{AOA}\right) \boldsymbol{a}_{N_t}^H\left(\theta_l^{AOD}\right), \tag{3}$$

where L is the number of paths, α_l is the complex gain of the l-th path, $f_{D,l} = \frac{v}{\lambda} \cos\left(\theta_l\right)$ is the Doppler frequency of l-th path, v is the user velocity, λ is the wavelength, and $\boldsymbol{a}(.)$ are the array response vectors.

This is the clustered geometric channel model, commonly used for mmWave channels, especially in high-mobility scenarios. Each path is associated with a complex gain α_l, and the Doppler shift $f_{D,l}$ introduces a time-varying phase shift at the receiver. Vectors $\boldsymbol{a}_{N_r}(\theta_l^{AOA})$ and $\boldsymbol{a}_{N_t}^H(\theta_l^{AOD})$ are the array response vectors for the UE and BS, respectively, corresponding to the angles of arrival $\theta_{r,l}$ and departure $\theta_{t,l}$.

3.2. LSTM-based Channel Prediction

The core of the proposed method is an LSTM network that predicts the $\mathbf{H}(t)$ future state of the channel matrix. The LSTM is particularly suitable for this task due to the ability to model long-term dependence in sequential data, which is essential for predicting channel variations in wireless communication systems. The input to the LSTM at time step t is

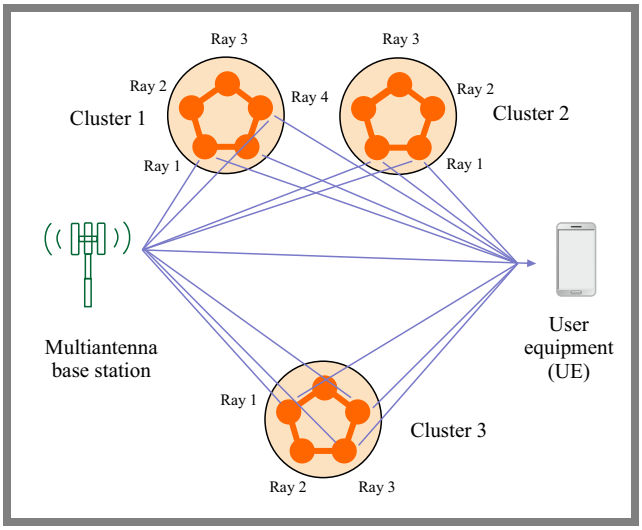

Fig. 3. Physical spread of the signal in the channel.

a sequence of the K most recent channel matrices:

$$X_t = \mathbf{H}(t-K), \mathbf{H}(t-K+1), \ldots, \mathbf{H}(t-1)\,,$$

where $\mathbf{H}(t) \in \mathbb{C}^{N_t \times N_r}$ is the channel matrix at time t, K is the number of previous time steps considered for predicting the channel matrix at the current time step.

These matrices are flattened and concatenated into a 2D tensor which serves as the input to the LSTM network. The LSTM network will use this sequence of channel matrices to predict the future state of the channel matrix.

We use a stacked LSTM architecture which involves multiple LSTM layers to capture both short- and long-term dependencies in the channel evolution:

- Input layer – takes the flattened channel sequence from the previous time steps.
- LSTM layer 1 – contains 128 LSTM units and processes the sequence and captures short-term temporal features. Using tanh as the activation function is standard.
- Layer 2 – with 64 LSTM units helps to learn more complex and longer-term dependencies in the evolution of the channel.
- Dense layer (fully connected) – with 256 neurons with ReLU activation function helps in mapping the learned temporal features to the desired output dimension.
- Output layer – a dense layer with $2 \times N_{UE} \times N_{BS}$ neurons (for real and imaginary parts) and a linear activation function to output the flattened predicted channel matrix $\widehat{\mathbf{H}}(t)$.

Each $\mathbf{H}(t) \in \mathbb{C}^{N_t \times N_r}$ is split into Re and Im and concatenated along the feature axis, yielding an input tensor of shape $(\mathrm{K}, 2N_rN_t)$.
The LSTM outputs $\widehat{\mathbf{H}}(\mathrm{t}+1)$ with the same format, which we reassemble into complex form. Beamformer update is realized using $\widehat{\mathbf{H}}(t+1)$ in following way:

- Compute SVD of: $H_{\text{pred}} = W_{RF}^H \widehat{\mathbf{H}}(\mathrm{t}+1)\mathbf{F}_{RF} \Rightarrow U \sum V^H$,
- Pick $\mathbf{F}_{RF}, \mathbf{W}_{RF}$ via OMP over steering-vector dictionaries to approximate the dominant singular subspace under constant-modulus,
- Set $\mathbf{F}_{BB} = V_{(:,1:N_S)}$ and $\mathbf{W}_{BB} = U_{(:,1:N_S)}$ with a normalization to meet power. This predict-then-design loop repeats every coherence block and K (history length) trades delay for robustness.

For network feature and I/O mapping each complex channel $\mathbf{H}(\mathrm{t}) \in \mathbb{C}^{N_t \times N_r}$ is split into real/imaginary parts and concatenated, yielding an input tensor shape $(\mathrm{K}, 2N_rN_t)$. We use a stacked LSTM with 128 and 64 hidden units (first returns sequences, second returns last state), followed by a dense 250 (ReLU) and an output layer of size $2N_rN_t$ (linear) that reconstructs $\widehat{\boldsymbol{H}}(t+\mathbf{1})\mathbb{C}^{N_r \times N_t}$. Then $\mathbf{H}_{eff} = \mathbf{W}_{RF}^H \widehat{\boldsymbol{H}}(t+1)\boldsymbol{F}_{RF}$ and design $\boldsymbol{F}_{BB}, \boldsymbol{W}_{BB}$ are formed via SVD subject to constant-modulus analog constraints as shown in Fig. 4.

The training process is conducted using the scheme presented below.

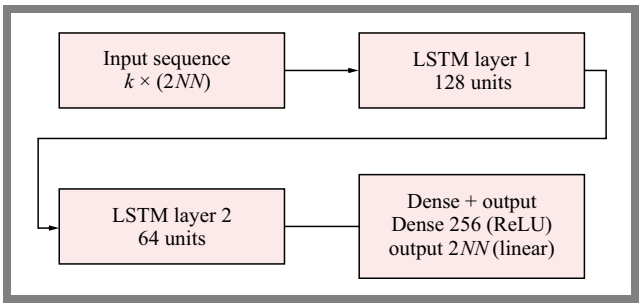

Fig. 4. LSTM network features.

We generate thousands of channel evolution sequences under various user velocities, angles, and scattering environments to create a comprehensive training dataset. This dataset represents the realistic dynamics of the channel matrix over time.

The loss function used to train the LSTM is the mean squared error (MSE) between the predicted channel matrix $\widehat{\mathbf{H}}(t)$ and the actual channel matrix $\mathbf{H}(t)$:

$$\mathcal{L} = \frac{1}{M} \sum_{i=1}^{M} \left\| \mathbf{H}_i(t) - \widehat{\mathbf{H}}_l(t) \right\|_F^2 , \tag{4}$$

where M is the batch size and $\|.\|_F$ is the Frobenius norm which calculates the matrix difference between the predicted and actual channel matrices.

The Adam optimizer is used for training the LSTM [15]. Adam is a popular optimization algorithm due to its adaptive learning rate and is well suited for training deep networks [16]. The LSTM-based predictive hybrid beamforming is provided as Algorithm 1, while Fig. 5 shows the flow of the predict-then-design hybrid beamforming method.

Algorithm 1 LSTM-based predictive hybrid beamforming

Input: Past K channels $\{\mathbf{H}(t-K+1), \ldots, \mathbf{H}(t)\} \in \mathbb{C}^{N_r \times N_t}$

1: **Complex → real stack:** $[\Re\{\mathbf{H}(.)\}, \Im\{\mathbf{H}(.)\}] \in \mathbb{R}^{K \times (2N_rN_t)}$
2: **LSTM prediction:** output $\widehat{\mathbf{H}}(t+1)\mathbb{C}^{N_r \times N_t}$
3: **Analog dictionaries:** $\mathcal{A}_t = \{a_{N_r}(\theta_m)\}, \mathcal{A}_r = \{a_{N_t}(\phi_n)\}$ (ULA, $d = \lambda/2$)
4: **OMP** (analog stage): select $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_{RF}^t}, \mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N_{RF}^r}$ with $\left|[\mathbf{F}_{\mathrm{RF}}]_{ij}\right| = \frac{1}{\sqrt{N_t}}, \left|[\mathbf{W}_{\mathrm{RF}}]_{ij}\right| = \frac{1}{\sqrt{N_r}}$ to best approximate the dominant subspaces of $\widehat{\mathbf{H}}(t+1)$
5: **Digital stage:** $\mathbf{H}_{eff} = \mathbf{W}_{RF}^H \widehat{\mathbf{H}}(t+1)\mathbf{F}_{RF}$ take SVD $= U\Sigma V^H$; $\mathbf{F}_{BB} = V_{(:,1:N_s)}$ and $\mathbf{W}_{BB} = U_{(:,1:N_s)}$
Normalize $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_s$
6: **Apply:** use $\mathbf{F}_{RF}, \mathbf{F}_{BB}, \mathbf{W}_{RF}, \mathbf{W}_{BB}$ at slot $t+1$

Complexity: OMP $\mathcal{O}\left(N_s\left(|\mathcal{A}_t|\, N_t + |\mathcal{A}_r|\, N_r\right)\right)$;
SVD of $N_{RF}^r \times N_{RF}^t$
LSTM inference $\mathcal{O}\left(K, N_rN_td_{LSTM}\right)$

3.3. Predictive Hybrid Beamforming

The objective of predictive hybrid beamforming is to use predicted future channel states $\widehat{\mathbf{H}}(t)$ in order to adjust the beamforming matrices in a proactive manner, thus aiming to maximize spectral efficiency. This approach is especially beneficial in dynamic environments where the channel evolves

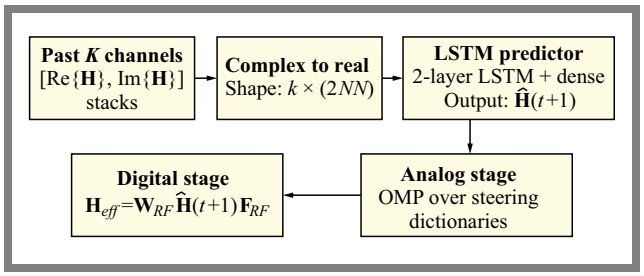

Fig. 5. Predict-then-design pipeline.

over time, such as high-mobility scenarios with significant Doppler shifts. By predicting the future state of the channel matrix, we can preemptively adjust the beamforming vectors to better align with the expected future channel conditions, improving the system performance.

In hybrid beamforming, the goal is typically to maximize the spectral efficiency, which quantifies the data rate achievable over a given bandwidth.

Let us define the effective channel as:

$$\mathbf{H}_{eff}(t) = \mathbf{W}_{RF}^{H}\,\mathbf{H}(t)\,\mathbf{F}_{RF}.$$

With digital processing, the per-slot spectral efficiency is:

$$R(t) = \log_2 \det\left(N_s + \frac{\rho}{N_s}\left(\mathbf{W}_{BB}^{H}\,\mathbf{W}_{BB}\right)^{-1}\cdot \mathbf{W}_{BB}^{H}\,\mathbf{H}_{eff}(t)\,\mathbf{F}_{BB}\,\mathbf{F}_{BB}^{H}\,\mathbf{H}_{eff}^{H}(t)\,\mathbf{W}_{BB}\right), \quad (5)$$

where $R(t)$ is the spectral efficiency in bits/sec/Hz and N_s is the number of data streams and $\rho = \frac{P}{\sigma_n^2}$.

In Eq. (5) the expression inside the logarithm is the effective channel capacity of the system, taking into account both transmit power and noise. The determinant represents the total capacity available in the system, considering the effective channel gain, interference, and noise.

The formulation assumes AWGN at the receiver and perfect channel state information at the transmitter. Term $\frac{P}{N_s\,\sigma_n^2}$ is the SNR per data stream. The determinant and logarithmic form comes from the Shannon capacity formula for MIMO systems, where the capacity grows logarithmically with the SNR, and the determinant represents the overall gain from the eigenvalues of the system which are captured by the SVD of the channel matrix.

We define $\mathbf{F}_{RF}, \mathbf{F}_{BB}, \mathbf{W}_{RF}, \mathbf{W}_{BB}$ using the predicted $\widehat{\mathbf{H}}(t+1)$ from the LSTM, subject to constant-modulus constraints on $\mathbf{F}_{RF}, \mathbf{W}_{RF}$ and a transmit-power constraint $||\mathbf{F}_{RF}\,\mathbf{F}_{BB}||_F^2 = N_s$.

The optimization problem is non-convex due to the constant-modulus constraint on the analog beamforming matrices [17], [18]. The constant-modulus constraint arises because the analog beamforming matrix (implemented using phase shifters) can only adjust the phase of each antenna element, but cannot control the amplitude.

The iterative approach is as follows:

- The first step is to design analog precoder $\mathbf{F}_{RF}$ and analog combiner $\mathbf{W}_{RF}$ to align with the dominant channel paths. This is achieved by selecting the columns from array response matrix $\alpha(\phi)$ from Eq. (1), which describes the response of the antenna array to different angles that maximize the projected channel gain.
- Once the analog beamformers $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$ are fixed, the effective channel matrix at the receiver is:

$$\widehat{\mathbf{H}}_{eff} = \mathbf{W}_{RF}^{H}\widehat{\mathbf{H}}\,\mathbf{F}_{RF}\,. \quad (6)$$

This effective channel matrix represents the combined effect of analog beam formation at both the transmitter and the receiver.

Next, we develop the digital precoder $\mathbf{F}_{BB}$ and combiner $\mathbf{W}_{BB}$ by performing singular value decomposition (SVD) on the effective channel $\widehat{\mathbf{H}}_{eff}$:

$$\widehat{\mathbf{H}}_{eff} = \mathbf{U}_{eff}\,\Sigma\,\mathbf{V}^{\mathrm{H}}{}_{eff}\,\Sigma\,. \quad (7)$$

where $\mathbf{U}_{eff}$ and $\mathbf{V}_{eff}$ are the left and right singular matrices $\Sigma = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_{N_S})$.

The digital precoder $\mathbf{F}_{BB}$ and digital combiner $\mathbf{W}_{BB}$ are derived from the right and left singular vectors of the effective channel:

$$\mathbf{F}_{BB} = \mathbf{V}_{eff}\,, \quad (8)$$

$$\mathbf{W}_{BB} = \mathbf{U}_{eff}\,. \quad (9)$$

These digital matrices align the signal with the strongest eigen modes of the effective channel and ensure optimal data stream transmission.

The key advantage of the predictive hybrid beamforming approach is that the beamforming matrices $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$ are adjusted proactively based on the predicted future channel matrix $\widehat{\mathbf{H}}(t)$. This allows the system to track the mobile user effectively even in the presence of significant Doppler shifts.

By predicting the future state of the channel, the system can pre-emptively compensate for the variations caused by mobility, leading to more robust communication in high-mobility environments.

Per slot LSTM inference is $\mathcal{O}\,(\boldsymbol{K}, N_r\,N_t\,\boldsymbol{d}_{LSTM})$. OMP on steering dictionaries scales as $\mathcal{O}\,(N_s\,(|\mathcal{A}_t|\,N_t + |\mathcal{A}_r|\,N_r))$. The digital SVD runs on $\mathbf{H}_{eff} \in \mathbb{C}^{N_{RF}^r \times N_{RF}^t}$ (RF chain domain), costing $\mathcal{O}\big(\min\big\{(N_{RF}^r)^2\,N_{RF}^t, (N_{RF}^t)^2\,N_{RF}^r\big\}\big)$ not on full $N_r \times N_t$. Unlike full-digital beamforming, no large matrix inversion at antenna dimension is required, power normalization uses small $N_S \times N_S$ matrices.

4. Results and Discussion

Simulations are carried out in Matlab. Table 2 shows the simulation parameters and their values used to generate the results and discuss the concepts presented in this section.

The baseline models used for comparison are as follows.

- Orthogonal matching pursuit (OMP) based hybrid beamforming – a well-known iterative algorithm for hybrid precoding that does not account for Doppler effects [13],

Tab. 2. Simulation setup parameters.

Parameter	Value
Transmitting antenna N_t	64, tunable from 16 to 128
Receiving antenna N_r	16, tunable from 4 to 4
Transmitter RF chains N_{RF}^t	4, tunable from 2 to 8
Receiver RF chains N_{RF}^r	4, tunable from 2 to 8
Data streams N_s	4
Carrier frequency	28 GHz
Antenna spacing	$\frac{\lambda}{2}$
Number of channel paths	5
Mobility	Standardized mobility model (e.g. 3GPP urban mobility) 30, 120, 300 km/h

- Kalman filter-based hybrid beamforming (KF-HBF) – a conventional approach for tracking time-varying channels [7].
- Perfect CSI (upper bound) – an ideal case where instantaneous CSI is perfectly known at the transmitter [19].

We consider a mmWave system operating at 28 GHz with a 100 MHz bandwidth. The BS has a 64-element uniform linear array (ULA), and the UE has a 16- and 64-element ULA. We simulate the user velocities from 30 km/h to 300 km/h for different Doppler spreads.

Figure 6 shows the spectral efficiency (SE) plot as a function of SNR. At low SNR (–10 dB), all methods exhibit nearly equal performance (~ 4.5 bits/s/Hz). This convergence is expected due to noise dominance and limited beamforming gain. As SNR increases, divergence in performance becomes visible. At 0 dB, LSTM-HBF at 30 km/h achieves approximately 8.1 bits/s/Hz, which is slightly higher than KF-HBF (~ 7.9 bits/s/Hz) and OMP (~ 7.8 bits/s/Hz).

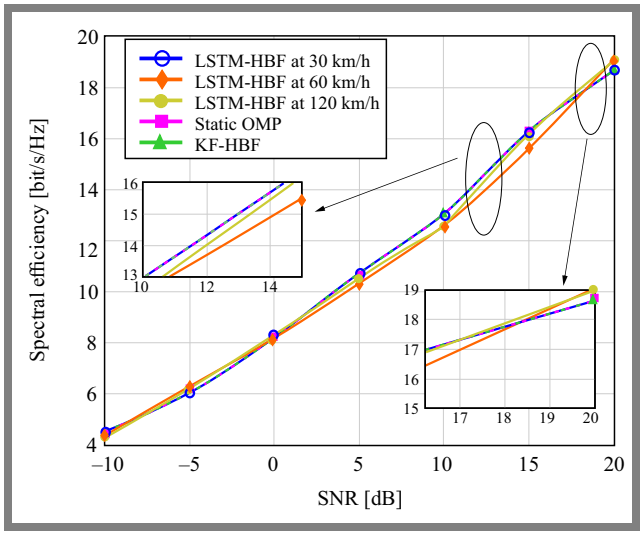

Fig. 6. Comparison of spectral efficiency for 64T and the 16R configuration and velocity.

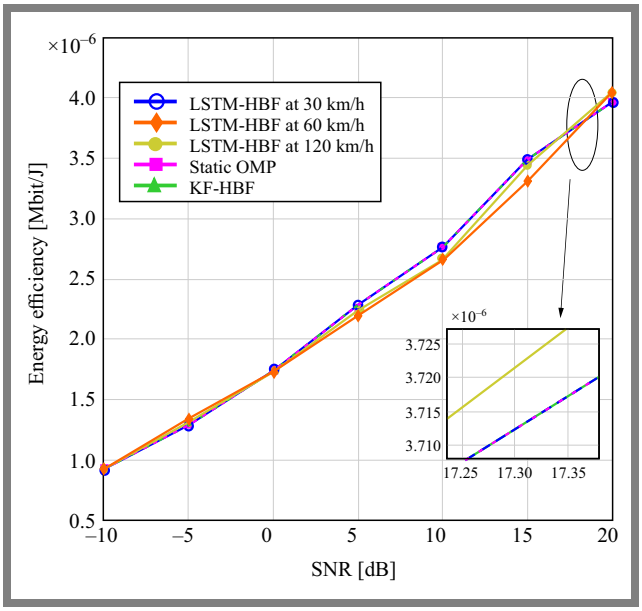

Fig. 7. Comparison of energy efficiency for 64T and 16R configuration and velocity.

At 10 dB, LSTM-HBF reaches ~ 13.7 bits/s/Hz, outperforming OMP and KF-HBF by ~ 0.5 bits/s/Hz, with a consistent margin across mobility variations. At 20 dB, LSTM-HBF peaks at approximately 19.2 bits/s/Hz, while OMP and KF-HBF saturate closer to 18.6 bits/s/Hz, suggesting that LSTM-based learning preserves marginal advantages even under high-SNR ceilings. Classical OMP/KF beamforming indeed performs well at high SNRs with low mobility. Our experiments target high-mobility Doppler where analog beams lag; the predictive design preserves alignment, yielding non-trivial throughput gains even when baselines appear near-optimal at 20 dB. This is consistent with the coherence-time limits at mmWave and supports proactive design rather than purely reactive tracking.

All methods are evaluated on identical channel realizations with identical power normalization. The mild divergence near 15 dB is the transition region from noise-limited to interference/quantization-limited operation, where analog dictionary granularity and prediction error interact, producing slightly different slopes across various methods.

Figure 7 shows energy efficiency (EE) as a function of SNR. At –10 dB SNR, all models operate near 0.9 Mbits/J, with LSTM-HBF at 60 km/h being slightly lower due to marginal underperformance in SE. From 0 dB onward, the EE curve shows more separation, at 5 dB, LSTM-HBF at 30 km/h achieves ~ 2.3 Mbits/J, leading OMP and KF-HBF by approximately 0.1 – 0.2 Mbits/J.

At 10 dB, LSTM-HBF continues its linear climb to ~ 3.4 Mbits/J, surpassing traditional methods by ~ 0.3 Mbits/J. By 20 dB, the advantage of EE becomes more noticeable with LSTM-HBF reaching ~ 4.2 Mbits/J, compared to OMP (~ 4.0) and KF-HBF (~ 4.05). The EE of LSTM-HBF at 120 km/h closely matches the results at 30 and 60 km/h, which is a strong indicator of model generalization under different Doppler conditions. KF-HBF shows slight degradation at

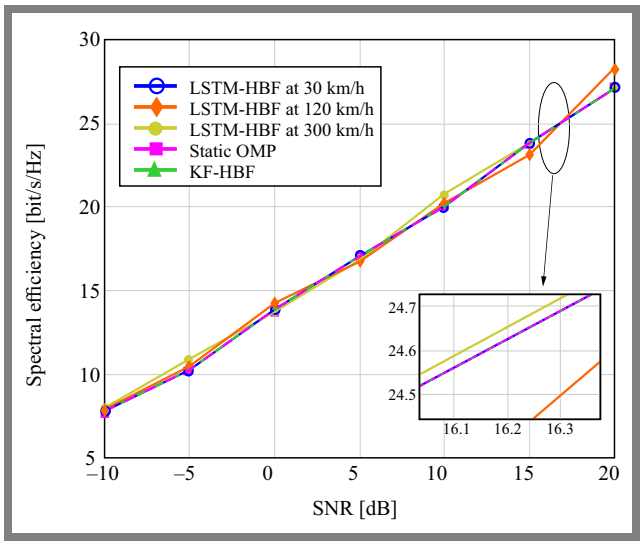

Fig. 8. Comparison of spectral efficiency for 64T and 64R configuration and velocity.

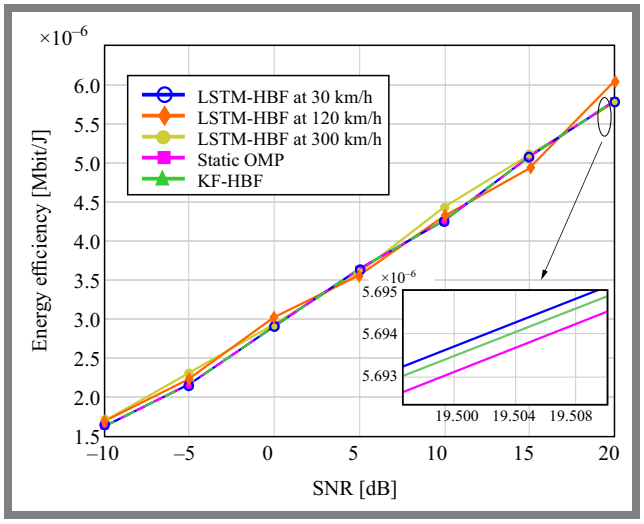

Fig. 9. Comparison of energy efficiency for 64T and 64R configuration and velocity.

higher SNRs, likely due to estimation errors accumulating over time, especially under high mobility.

Figure 8 illustrates the variation of spectral efficiency (SE) with respect to SNR for the proposed LSTM-HBF model at three different vehicular speeds: 30 km/h, 120 km/h, and 300 km/h, and compares it with two baseline methods: KF-HBF and static OMP. At an SNR of –10 dB, all schemes exhibit nearly identical SE, approximately 7.8 bits/s/Hz, as the system performance is primarily noise-limited. However, as the SNR increases, the LSTM-HBF method begins to demonstrate superior performance. At 0 dB, LSTM-HBF operating at 30 km/h achieves approximately 13.7 bits/s/Hz, slightly outperforming KF-HBF and OMP, which reach 13.3 and 13.2 bits/s/Hz, respectively. The performance gap becomes more pronounced at higher SNR values. At 20 dB, the LSTM-HBF model at 120 km/h attains a spectral efficiency of approximately 28.6 bits/s/Hz, outperforming KF-HBF and static OMP by roughly 1.1 and 1.3 bits/s/Hz, respectively.

Figure 9 presents the energy efficiency results for the same configurations. At an SNR of –10 dB, all models achieve comparable EE values of ∼1.6 to 1.7 Mbits/J. This similarity is expected given the low throughput and high relative power cost in this regime. As SNR increases, the LSTM-HBF model exhibits a more rapid improvement in EE. At 10 dB, the LSTM-HBF at 300 km/h achieves an EE of approximately 4.2×10^{-6} Mbits/J, slightly ahead of KF-HBF and OMP, which remain around 3.9×10^{-6} and 3.8×10^{-6} Mbits/J, respectively. By 20 dB, the LSTM-HBF model at 120 km/h reaches the highest energy efficiency of approximately 6.1×10^{-6} Mbits/J. In contrast, KF-HBF and static OMP attain 5.9×10^{-6} and 5.8×10^{-6} Mbits/J, respectively. These results highlight the dual advantage of LSTM-HBF in maximizing both spectral and energy efficiency.

5. Conclusions

This paper presents an LSTM-based hybrid beamforming (LSTM-HBF) approach that aims to improve performance of mmWave communication under varying mobility and SNR conditions. The method was benchmarked against traditional approaches such as static OMP and KF-HBF across various spectral and energy efficiency metrics.

The results demonstrate that LSTM-HBF consistently achieves higher performance. Especially, at 20 dB SNR, it delivers a spectral efficiency of up to 28.6 bits/s/Hz, compared to 27.5 bits/s/Hz for KF-HBF and 27.3 bits/s/Hz for static OMP. In terms of energy efficiency, LSTM HBF reaches 6.1×10^{-6} Mbits/J, surpassing the closest benchmark by a margin of $\sim 0.3 \times 10^{-6}$ Mbits/J. The LSTM-HBF model exhibits robust performance across all mobility profiles, spanning from 30 to 300 km/h, with minimal deviation in both SE and EE, indicating strong resilience to Doppler effects and time-varying channel conditions. Future work could explore more complex neural network architectures and investigate the impact of imperfect channel estimation on the performance of the proposed model.

By integrating a predictive LSTM network into a hybrid beamforming framework, we have shown through detailed modeling and simulated results that significant gains in both spectral and energy efficiency are achievable, particularly in high-mobility scenarios where traditional methods fail. This proactive approach ensures a robust communication link, making reliable multi-gigabit mobile communication a practical reality.

The future of this work involves real-life signal analysis with available datasets by extracting temporal channel slices along user trajectories and by using the same analog/digital design steps to compute SE/EE with identical power normalization.

References

[1] M. Shafi *et al.*, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice", *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 1201–1221, 2017 (https://doi.org/10.1109/JSAC.2017.2692307).

[2] T.S. Rappaport *et al.*, "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond", *IEEE Access*, vol. 7, pp. 78729–78757, 2019 (https://doi.org/10.1109/ACCESS.2019.2921522).

[3] R.W. Heath *et al.*, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems", *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 436–453, 2016 (https://doi.org/10.1109/JSTSP.2016.2523924).

[4] A.F. Molisch *et al.*, "Hybrid Beamforming for Massive MIMO: A Survey", *IEEE Communications Magazine*, vol. 55, pp. 134–141, 2017 (https://doi.org/10.1109/MCOM.2017.1600400).

[5] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter Wave Communication in Vehicular Networks: Challenges and Opportunities", *2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Thessaloniki, Greece, 2017 (https://doi.org/10.1109/MOCAST.2017.7937682).

[6] A. Alkhateeb *et al.*, "Deep Learning Coordinated Beamforming for Highly-mobile Millimeter Wave Systems", *IEEE Access*, vol. 6, pp. 37328–37348, 2018 (https://doi.org/10.1109/ACCESS.2018.2850226).

[7] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/non-Gaussian Bayesian Tracking", *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002 (https://doi.org/10.1109/78.978374).

[8] H. Ye, G.Y. Li, and B.-H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems", *IEEE Wireless Communications Letters*, vol. 7, pp. 114–117, 2018 (https://doi.org/10.1109/LWC.2017.2757490).

[9] E. Tuna and A. Soysal, "LSTM and GRU Based Traffic Prediction Using Live Network Data", *2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021 (https://doi.org/10.1109/SIU53274.2021.9478011).

[10] C. Jiang *et al.*, "Machine Learning Paradigms for Next-generation Wireless Networks", *IEEE Wireless Communications*, vol. 24, pp. 98–105, 2017 (https://doi.org/10.1109/MWC.2016.1500356WC).

[11] W. Shahjehan *et al.*, "A Review on Millimeter-wave Hybrid Beamforming for Wireless Intelligent Transport Systems", *Future Internet*, vol. 16, art. no. 337, 2024 (https://doi.org/10.3390/fi16090337).

[12] A. Alkhateeb, O. El Ayach, G. Leus, and R.W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems", *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, 2014 (https://doi.org/10.1109/JSTSP.2014.2334278).

[13] O. El Ayach *et al.*, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems", *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, 2014 (https://doi.org/10.1109/TWC.2014.011714.130846).

[14] I. Marinovic, I. Zanchi, and Z. Blazevic, "Estimation of Channel Parameters for 'Saleh-Valenzuela' Model Simulation", *2005 18th International Conference on Applied Electromagnetics and Communications*, Dubrovnik, Croatia, 2005 (https://doi.org/10.1109/ICECOM.2005.204926).

[15] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO", *IEEE Transactions on Signal Processing*, vol. 63, pp. 6169–6183, 2015 (https://doi.org/10.1109/TSP.2015.2463260).

[16] Junyi Wang *et al.*, "Beam Codebook Based Beamforming Protocol for Multi-Gbps Millimeter-wave WPAN Systems", *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1390–1399, 2009 (https://doi.org/10.1109/JSAC.2009.091009).

[17] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer", *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, pp. 563–575, 2017 (https://doi.org/10.1109/TCCN.2017.2758370).

[18] 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, pp. 416–419, 2019 (https://doi.org/10.1109/LWC.2018.2874264).

[19] S.H. Lim, S. Kim, B. Shim, and J.W. Choi, "Deep Learning-based Beam Tracking for Millimeter-wave Communications Under Mobility", *IEEE Transactions on Communications*, vol. 69, pp. 7458–7469, 2021 (https://doi.org/10.1109/TCOMM.2021.3107526).

Kartik Ramesh Patel, M.E.
Dep. of Electronics and Telecommunication Engineering
https://orcid.org/0000-0001-8391-5537
E-mail: kartik@somaiya.edu
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
https://www.mctrgit.ac.in

Sanjay Dasrao Deshmukh, Ph.D.
Dep. of Electronics and Telecommunication Engineering
https://orcid.org/0009-0004-0382-9831
E-mail: sanjay.deshmukh@mctrgit.ac.in
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
https://www.mctrgit.ac.in