IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 March 2023, accepted 4 April 2023, date of publication 11 April 2023, date of current version 24 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3266355

RESEARCH ARTICLE

A CNN-LSTM-Based Fusion Separation Deep Neural Network for 6G Ultra-Massive MIMO Hybrid Beamforming

RAFID UMAYER MURSHED[1], ZULQARNAIN BIN ASHRAF[1], ABU HORAIRA HRIDHON[1], KUMUDU MUNASINGHE[2], (Member, IEEE), ABBAS JAMALIPOUR[3], (Fellow, IEEE), AND MD. FARHAD HOSSAIN[1], (Member, IEEE)

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1205, Bangladesh
[2]School of IT and Systems, University of Canberra, Canberra, ACT 2617, Australia
[3]School of Electrical and Information Engineering, The University of Sydney, Camperdown, NSW 2006, Australia

Corresponding author: Rafid Umayer Murshed (rafid.buet.eee16@gmail.com)

ABSTRACT In the sixth-generation (6G) cellular networks, hybrid beamforming would be a real-time optimization problem that is becoming progressively more challenging. Although numerical computation-based iterative methods such as the minimal mean square error (MMSE) and the alternative manifold optimization (Alt-Min) can already attain near-optimal performance, their computational cost renders them unsuitable for real-time applications. However, recent studies have demonstrated that machine learning techniques like deep neural networks (DNN) can learn the mapping done by those algorithms between channel state information (CSI) and near-optimal resource allocation and then approximate this mapping in near real-time. In light of this, we investigate various DNN architectures for beamforming challenges in the terahertz (THz) band for ultra-massive multiple-input multiple-output (UM-MIMO) and explore their contextual mathematical modelling. Specifically, we design a sophisticated 1D convolutional neural network and long short-term memory (1D CNN-LSTM) based fusion-separation scheme, which can approach the performance of the Alt-Min algorithm in terms of spectral efficiency (SE) for fully connected structures and, at the same time, use significantly less computational effort. Simulation results indicate that the proposed system can attain almost the same level of SE as that of the numerical iterative algorithms while incurring a substantial reduction in computational cost. Our DNN-based approach also exhibits exceptional adaptability to diverse network setups and high scalability. Although the current model only addresses the fully connected hybrid architecture, our approach can also be expanded to address a variety of other beamforming topologies.

INDEX TERMS 6G, CNN, hybrid beamforming, LSTM, UM-MIMO.

I. INTRODUCTION

With daily increase in data traffic requirements ranging from mission-critical to massive machine connectivity, optimism for the sixth generation (6G) cellular network is growing exponentially. In response to the dramatic increase in the use of smartphones and related technologies, as well as the ongoing development of autonomous vehicles and Internet of Things (IoT) devices, the wireless industry has set extremely aggressive performance targets for the 6G systems. These targets will incorporate a peak data rate of at least 1 Tb/s, which is 100 times that of 5G, an over-the-air latency of 10–100 μs, and high mobility (1,000 km/h) [1]. Although specific requirements will vary according to deployment scenarios, it is safe to conclude that 6G will deliver a far more sophisticated user experience than current wireless technologies. Obtaining the desired performance levels often requires addressing complicated optimization problems extremely fast, such as resource allocation, beamforming, precoding, and scheduling. The intriguing question now is how such

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

enormous gains will be accomplished. Three main genres of improvement can be considered as listed below.

- Utilization of additional resources (e.g., gaining access to a greater chunk of the spectrum, including hitherto untapped frequency bands),
- Enhancement of the efficiency of utilization of those resources using a variety of approaches (e.g., hybrid beamforming), and
- Virtualization of the network to the greatest extent possible in order to minimize cost, flexible deployment and better scalability (e.g., network slicing).

Although research has been going on in all these genres, increased utilization of existing resources has seen significant developments in the past couple of years. One of the most interesting prospects being explored for 6G deployment is the utilization of the terahertz (THz) band [2].

It is well-known that the huge accessible bandwidths at THz frequencies are accompanied by substantial propagation losses and power constraints, resulting in minimal communication distances. Several strategies have been proposed to combat this predicament, among which ultra-massive multiple-input multiple-output (UM-MIMO) beamforming seems to be the most promising [3]. Furthermore, multiple antenna communication, a critical component of next-generation wireless technology, may attain higher data rates than single antenna systems. The use of beamforming is essential for obtaining such high rates. These beamforming systems with many antenna elements can significantly enhance the spectral and energy efficiency at THz frequencies by focusing narrow and high-gain beams on a small region, thereby increasing the overall data rate and channel capacity via spatial multiplexing. Beamforming can primarily be classified into three types: analog, digital and hybrid. In analog beamforming, phase shifters are used in the radio frequency (RF) domain at transmitters to send the same signal directed toward a specific receiver. These signals are combined at the receiver, and hence coverage is increased. However, in digital beamforming, each transmitter is equipped with its own RF chain, and multiple independent beams are generated, which significantly increases the data rate. This increase in data rate transpires at the cost of extremely high power consumption and hardware costs. On the other hand, hybrid beamforming (HBF) sacrifices a little accuracy to reduce power consumption and hardware costs significantly. Here, high-dimensional analog beamforming is combined with lower-dimensional digital baseband precoding. Due to its many advantages, hybrid beamforming has become quite popular in research and industry.

Existing beamforming algorithms rely on numerical iteration-based complex computations, which impede real-time resource allocation and generate a large amount of computational overhead, increasing both latency and costs [4]. Furthermore, these algorithms are often valid only for specific network configurations. With the enormous number of antennas required by THz UM-MIMO systems, the computational costs induced by these numerical iteration-based algorithms only get higher and higher. Furthermore, the maximum ratio (MR) and minimal mean square error (MMSE) algorithms utilized in MIMO receivers are not tuned to reduce computational complexity and communication latency. Consequently, there is an incentive to discover a different approach to beamforming problems that is simple, efficient and adaptable. Various algorithms have also been put forward to reduce power consumption through hybrid beamforming [5], which achieve sum rates comparable to those of fully digital beamformers. However, each algorithm developed thus far consumes a significant number of computational resources [5], [6], [7]. Hence, they are impractical for real-time use. Meanwhile, the superior performance of deep neural networks (DNNs) on a range of inference and regression tasks has resulted in significant investment in research, development, and cloud infrastructure deployment for training and running DNNs. Supervised learning has become a valuable tool in training DNNs to achieve accurate and reliable predictions [8]. While the training phase of DNNs might be pretty time-consuming, the inference phase follows a straightforward deterministic execution model. Hence, DNNs are becoming increasingly more appealing for executing real-time inference tasks.

While substantial advances in the application of DL to wireless networks have been made [9], [10], [11], [12], the majority of research in this domain is data-driven and does not make use of the most recent innovations in DL architectures and algorithms. This paper proposes a deep learning-based beamforming scheme with a specially designed fusion-separation network based upon a 1D convolutional neural network and long short-term memory (LSTM) for UM-MIMO beamforming at THz. This work combines domain-specific mathematical modelling with data-centric machine learning (ML) models to create a more sophisticated architecture with improved computational efficiency and robustness. We not only employ a data-driven strategy here but also include domain-specific knowledge and the most advanced ML techniques and architectures. We employ supervised learning to discover end-to-end mappings for MO-Alt-Min. Extensive simulations using synthetic data are conducted to validate our proposed approach.

A. RELATED WORKS

The design of the beamforming matrix is constrained by the expensive mmWave RF chains. Traditional full-digital beamformers need to connect an RF chain for each antenna element, which imposes intolerant power consumption and hardware cost and is no longer suitable for THz UM-MIMO systems. Analog RF beamforming schemes, implemented using analog circuitry introduced in [13], [14], [15], and [16], commonly employ analog phase shifters, which limit the beamformer's components to having a constant modulus. As a result, analog beamforming performs poorly compared to completely digital beamforming methods. To solve this

problem, an HBF architecture was proposed in [7]. It replaces the fully digital beamformer with a low-dimensional digital precoder followed by a high-dimensional analog precoder.

To date, most research focuses on two HBF structures; namely, fully-connected [7] and partially-connected architecture [17], [18], [19]. Authors in [20] studied fully-connected and partially-connected structures and demonstrated that the fully-connected structure has better spectral and energy efficiency than the partially-connected structure when insertion loss is included. Many topologies have been thoroughly researched to overcome the limitation of the number of RF chains. To obtain feasible hybrid beamformers and combiners, some numerical algorithms, such as orthogonal matching pursuit (OMP) [7], Karush-Kuhn-Tucker (KKT) based [5], and manifold optimization (MO) based [21] algorithms, were proposed. The hybrid precoder design was approached in [21] as a matrix factorization issue, which proposed efficient alternating minimization (Alt-Min) methods for the fully-connected and partially-connected hybrid precoding structures, respectively. In particular, an Alt-Min technique based on MO was suggested for the fully connected structure to approach the performance of the highly complex, fully digital precoder. Thus, by imposing an orthogonal restriction on the digital precoder, a low-complexity Alt-Min algorithm was subsequently proposed. It, however, significantly reduced the SE. Therefore, these algorithms either fail to achieve optimal performance in terms of spectral efficiency (SE) or generate a massive computational burden.

To this end, there has been much research on the implementation of beamforming in designing efficient and resilient MIMO systems [22], [23], [24], [25]. Existing hybrid beamforming systems frequently assume that infinite resolution phase shifters will be used to construct analog beamformers [26]. However, the components necessary to construct precise phase shifters can be rather costly [27], [28]. In practice, low-resolution phase shifters with a lower cost are frequently utilized. The simplest method for designing beamformers with limited resolution phase shifters is to build the RF beamformer with infinite resolution first and then to quantize the value of each phase shifter to a finite set [29]. However, this approach is not useful in systems with low-resolution phase shifters [30]. A specialized model is necessary for channels with enormous transmitting or receiving arrays or surfaces. However, as characteristics, such as path loss and angle of arrival (AoA), cannot be assumed constant between antennas, the channel model becomes non-stationary in space [31]. Another method for restricting the number of RF chains is to use simple analog switches to achieve antenna subset selection [32], [33], [34]. They cannot, however, achieve complete diversity gain in correlated channels since the antenna selection strategy uses just a subset of channels [35], [36].

Authors in [7] studied mmWave systems with massive antenna arrays. Precoding/combining was formulated as a sparse reconstruction task using mmWave channels' spatial structure. Using basis pursuit, algorithms were created that approximate optimal unconstrained precoders and combiners with low-cost RF hardware. Sohrabi and Yu [5] proposed a hybrid beamforming design with a low-dimensional digital beamformer and an RF beamformer employing analog phase shifters. An architecture with fewer RF chains can approach the performance of a fully digital one. If the number of RF chains is double the number of data streams, the hybrid beamforming structure can implement any completely digital beamformer regardless of antenna elements. In scenarios with fewer RF chains, the hybrid beamforming design problem was examined for both a point-to-point MIMO and a downlink MU-MISO system. However, these proposed systems do not take into account the computational complexity that would further aggravate in THz UM-MIMO.

To achieve near-optimal performance while reducing the long time consumption incurred by the conventional numerical algorithms, some DL-based algorithms were proposed in [37], [38], [39], [40], and [41]. Authors in [37] introduced a DL approach for joint channel estimation and hybrid beamforming for frequency-selective, wideband mm-Wave systems. In particular, a massive MIMO orthogonal frequency division multiplexing (MIMO-OFDM) system was considered, and three different DL frameworks comprising convolutional neural networks (CNNs) were proposed, which accepted the received pilot signal as input and yielded the hybrid beamformers at the output. Authors in [38] proposed PrecoderNet, a deep reinforcement learning (DRL) based hybrid beamforming design for mmWave systems. Using the precoding and combining matrices from the previous learning iteration as the state and the current learning iteration's matrices as the action, the DRL agent learns the near-optimal HBF design policy. However, this approach requires a complex and computationally expensive MO-based approach to calculate the feasible analog beamformers. Paper [39] further proposed a CNN framework for the joint design of precoder and combiners. This network accepts the input of a channel matrix and gives the output of analog and baseband beamformers. The proposed CNN framework does not require the knowledge of steering vectors of array responses, and it provides higher performance in spectral efficiency as compared to the conventional greedy- and optimization-based algorithms. However, as THz UM-MIMO channel matrices are extremely large in dimensions, the computational overhead generated would be too high for real-time implementation. These DL methods require a massive amount of training data in advance, and the dimension of their input data is quite high for UM-MIMO systems. The larger the input size of a neural network, the more computationally intensive it is. In some cases, the training data is very difficult to obtain. When the transmission environment or the system configuration changes, new training data is needed, and the neural network needs to be retrained. In most cases, the entire architecture of the neural network has to be remodelled.

TABLE 1. Summary of Important Abbreviations.

Abbreviation	Definition
Alt-Min	Alternating Minimization
CNN	Convolutional Neural Network
CSI	Channel State Information
DL	Deep Learning
DNN	Deep Neural Network
FPS-Alt-Min	Alt-Min algorithm using Fixed Phase Shifter
GMM	Gaussian Mixture Model
HBF	Hybrid Beamforming
IoT	Internet of Things
LSTM	Long Short-Term Memory
ML	Machine Learning
MMSE	Minimum Mean Squared Error
mmWave	Millimeter-Wave
MO	Manifold Optimization
MR	Maximum Ratio
OMP	Orthogonal Matching Pursuit
PE-Alt-Min	Alt-Min algorithm using phase extraction
QoS	Quality of Service
RF	Radio Frequency
SE	Spectral Efficieny
SER	Symbol Error Rate
THz	Terahertz
ULA	Uniform Linear Array
URLLC	Ultra-Reliable Low Latency Communications
UM-MIMO	Ultra-Massive MIMO

TABLE 2. Summary of System Model Notations.

Notation	Definition
$\mathbf{a}_t$, $\mathbf{a}_r$	Transmit and receive array response vectors
$\mathbf{b}$	Binary symbol vector
C(A)	Column space of A
$\mathbf{e}$	Error vector
f	Carrier frequency
$\mathbf{F}_{\text{opt}}$	Optimal fully digital precoder matrix
$\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$	Analog and digital transmit beamforming matrices
$\mathbf{H}$	Channel matrix
N_{cl}, N_{ray}	Number of clusters and number of rays in each cluster
N_p	Number of propagation paths
N_s	Number of data-streams
N_t, N_r	Number of antennas at the transmitter and the receiver
N_{RF}^t, N_{RF}^r	Number of RF chains at the transmitter and the receiver
P	Angular power delay profile
$\mathbf{s}$	Symbol vector
$\mathbf{W}_{\text{RF}}$, $\mathbf{W}_{\text{BB}}$	Analog and digital beamforming matrices at receiver end
$\mathbf{W}_t$, $\mathbf{W}_r$	Precoding and combining matrices
$\mathbf{x}$	Transmitted signal
α	Path gain
β_n	Complex gain of the nth path
ρ	Transmit power
ϕ_r, θ_r	Receive angles of departure and arrival
ϕ_t, θ_t	Transmit angles of departure and arrival

B. CONTRIBUTIONS AND ORGANIZATION

This paper presents a supervised learning-based hybrid beamforming scheme for point-to-point UM-MIMO. First, training data for the neural network is generated using OMP and Alt-Min algorithms. The network is then trained offline using synthetic data and deployed online. The main contributions of this work are summarized as follows:

- We design a novel fusion-separation network to perform transmit and receive beamforming with reduced computational overhead for fully connected architectures. A single transmitter and receiver communicating at THz frequencies in a UM-MIMO system with multiple data streams are considered. We combine the sequence modelling capabilities of LSTM with the feature extraction capabilities of 1D-CNN to design a highly sophisticated neural network architecture. This beamforming scheme does not require the knowledge of steering vectors of array responses.
- We develop a mathematical modelling framework for the seamless simultaneous integration of complex matrices' real and imaginary components into a neural network's input layer. This allows us to reduce the complexity of the whole system and make it more efficient and lightweight.
- We analyze the performance of the novel fusion-separation network in terms of the achievable SE, varying RF chains, and computational overhead. The performance of the proposed approach is compared to that of the well-known Alt-Min algorithms for fully connected HBF. We also investigate the trade-off between model size and performance by creating three models of the same architecture but with different numbers of parameters.

The remainder of the paper is organized as follows. First, we delineate the system model and formulate the problem in Section II. Then, in Section III, we present the mathematical modelling of the proposed approach before delving into the detailed description of the novel fusion-separation network architecture in Section IV. Section V describes the simulation setup, while Section VI renders the simulation results accompanied by the pertinent analysis. Lastly, we conclude the paper with suggestions for future work in Section VII.

Regarding notation, scalars, matrices, and vectors are represented in lower case, bold upper, and bold lower cases, respectively. Scalar norms, vector L_2 norms, Frobenius norms, and pseudo-inverse are denoted by $|.|$, $||.||$, $||.||_F$ and $\dagger$, respectively. For any general matrix or vector operator $\mathbf{x}$, $\mathbf{x}^{\mathbf{T}}$ and $\mathbf{x}^*$ represent the transpose and conjugate transpose matrices, respectively. E[.], $\mathbb{C}$ and $\mathbb{N}$ denote the expected value, the set of the complex and natural numbers, respectively. Table 1 and Table 2 present the summary of important abbreviations and the system model notations, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we describe the adopted network and channel models. We then move on to formulate the basic beamforming problem.

A. NETWORK MODEL

We consider a single-user UM-MIMO network, where both the base station (BS) and the user employ a uniform linear

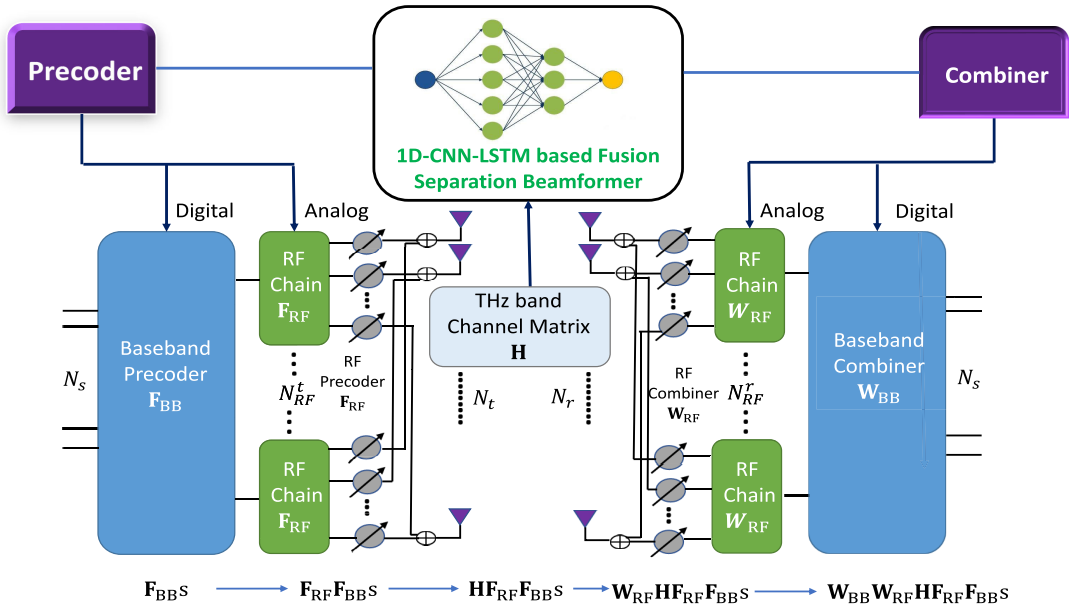

FIGURE 1. System architecture for 1D CNN-LSTM based UM-MIMO transceiver with hybrid beamforming.

array (ULA) of multiple antennas. We assume that there are N_t antennas at the transmitter and N_r antennas at the receiver, as shown in Figure 1. There are $N^t{}_{\mathrm{RF}}$ and $N^r{}_{\mathrm{RF}}$ number of RF chains at the transmitter and the receiver, respectively. We also consider this as a point-to-point multi-stream network with N_s number of data streams. Here, $N_s \leq N^t{}_{\mathrm{RF}} \leq N_t$ and $N_s \leq N^r{}_{\mathrm{RF}} \leq N_r$. Due to the limited number of transmit/receive RF chains, it is not possible to execute completely digital beamforming, which requires one dedicated RF chain per antenna element. As depicted in Figure 1, we instead explore a two-stage hybrid digital and analog beamforming architecture at the BS and user terminal. Now, the transmitted signal can be given as below

$$\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}, \tag{1}$$

where $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N^t{}_{\mathrm{RF}} \times N_s}$ is the $N^t{}_{\mathrm{RF}} \times N_s$ digital baseband precoder, $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times N^t{}_{\mathrm{RF}}}$ is the $N_t \times N^t{}_{\mathrm{RF}}$ analog RF precoder, and $\mathbf{s} \in \mathbb{C}^{N_s}$ is the symbol vector with

$$E[\mathbf{s}\mathbf{s}^H] = (\frac{1}{N_s})\mathbf{I}_{\mathbf{N}_s}. \tag{2}$$

Now, the received signal $\mathbf{r}$ after decoding can be expressed as

$$\mathbf{r} = \sqrt{\rho}\mathbf{W}_{\mathrm{RF}}{}^H\mathbf{W}_{\mathrm{BB}}{}^H\mathbf{H}\mathbf{x} + \mathbf{W}_{\mathrm{RF}}{}^H\mathbf{W}_{\mathrm{BB}}{}^H\mathbf{n}, \tag{3}$$

where $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times N^r{}_{\mathrm{RF}}}$ and $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N^r{}_{\mathrm{RF}} \times N_s}$ are respectively the analog and digital beamforming matrices at the receiver end, ρ is the transmit power, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{N_r})$ denotes the additive white Gaussian noise (AWGN) vector.

B. CHANNEL MODEL

In this research, the Saleh-Valenzuela (S-V) model [42] with a THz-band modification is employed to develop the channel model. The S-V model is a cluster-based, statistical channel model that connects clustering phenomena to stochastic angles of departure and arrival for each beam. Furthermore, HBF architectures can primarily be classified into two types based on the availability of channel state information (CSI), i.e., HBF based on full-instantaneous CSI and HBF based on averaged CSI [43]. We only address the first case in this study and leave the second one for future extensions of our work. In practical terms, CSI can be collected precisely and efficiently by channel estimation at the receiver and then shared with the transmitter using efficient feedback techniques [44]. Now, the channel matrix $\mathbf{H} \in \mathbb{C}^{N^t \times N^r}$ can be given by

$$\mathbf{H} = \sqrt{\frac{N_t\, N_r}{N_{\mathrm{cl}}\, N_{\mathrm{ray}}}} \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray}}} \alpha_{il}\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)\, \mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)^H, \tag{4}$$

where N_{cl} and N_{ray} represent the number of clusters and the number of rays in each cluster, respectively. α_{il} represents the gain of the l^{th} ray in the i^{th} propagation cluster. We assume that α_{il} are independent and identically distributed according to the distribution $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$ and $\sum_{i=1}^{N_{\mathrm{cl}}} \sigma_{\alpha,i}^2 = \hat{\gamma}$, which is the normalization factor to satisfy $\mathbb{E}[||\mathbf{H}||_F^2] = N_t N_r$. Moreover, $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ denote the receive and transmit array response vectors, where $\phi_{il}^r(\phi_{il}^t)$ and $\theta_{il}^r(\theta_{il}^t))$ stand for azimuth and elevation angles of arrival and departure, respectively. In this study, the uniform square planar array (USPA) with $\sqrt{N} \times \sqrt{N}$ antenna elements is investigated. Therefore, the array response vector corresponding to the l^{th} ray in the i^{th} cluster can be written as

$$\mathbf{a}(\phi_{il}, \theta_{il}) = \frac{1}{\sqrt{N}}[1, \ldots, e^{j\frac{2\pi}{\lambda}d(p\sin\phi_{il}\sin\theta_{il} + q\cos\theta_{il})}, \ldots, e^{j\frac{2\pi}{\lambda}d((\sqrt{N}-1)\sin\phi_{il}\sin\theta_{il} + (\sqrt{N}-1)\cos\theta_{il})}]^T, \tag{5}$$

where d and λ represent the antenna spacing and signal wavelength respectively, and $0 \leq p < \sqrt{N}$ and $0 \leq q < \sqrt{N}$ represent the antenna indices in the 2D plane, respectively. This channel model will be utilized in simulations, but our beamformer architecture is also applicable to more general models.

1) AZIMUTH AND ELEVATION ANGLE

Denoting the angular power delay profile of each individual cluster as $P_i(\phi, \theta)_{\text{cluster}}$, the complete power delay profile can be computed as [45]

$$P(\phi, \theta) = \sum_{i=1}^{N_c} P_i(\phi, \theta)_{\text{cluster}}. \tag{6}$$

Furthermore, considering there is no correlation between angular power profiles, the angular power profile for each cluster can be rewritten as

$$P_i(\phi, \theta)_{\text{cluster}} = P_i(\phi)_{\text{cluster}}\, P_i(\theta)_{\text{cluster}}. \tag{7}$$

The azimuth and elevation angle of each ray is assumed to be independent [46]. If the azimuth angle of i^{th} cluster is φ_i^t, and the azimuth angle of j^{th} ray within the i^{th} cluster is Φ_{ij}^t, then the total azimuth angle is expressed as

$$\phi_{ij}^t = \varphi_i^t + \Phi_{ij}^t. \tag{8}$$

Similarly, if the elevation angle of i^{th} cluster is ϑ_i^t, and elevation angle of j^{th} ray within the i^{th} cluster is Θ_{ij}^t, then the total elevation angle is

$$\theta_{ij}^t = \vartheta_i^t + \Theta_{ij}^t, \tag{9}$$

where Φ_{ij}^t and Θ_{ij}^t follow zero-mean second-order Gaussian mixture model that can be expressed as [47]

$$\text{GMM(x)} = \frac{a_1}{2\pi\sigma_1} e^{-\frac{1}{2}(\frac{x-\overline{x_1}}{\sigma_1})^2} + \frac{a_2}{2\pi\sigma_2} e^{-\frac{1}{2}(\frac{x-\overline{x_2}}{\sigma_1})^2}, \tag{10}$$

where $\overline{x_1}, \overline{x_2} = 0$. Azimuth angle φ_i^t and elevation angle ϑ_i^t of each cluster follow the uniform distribution, where $\varphi_i^t \in (-\pi, \pi]$ and $\vartheta_i^t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

C. PROBLEM FORMULATION

The beamforming problem can be segregated into two independent sub-problems, i.e., the precoder design and the decoder design. Their mathematical formulations are almost identical, aside from the former having an additional power constraint. The precoder design problem can be formulated as below

$$\min_{\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}} \left\| \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \right\|_F$$
$$s.t.\ |(\mathbf{F}_{\text{RF}})_{ij}| = 1,\ \ \|\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}\|_F^2 = N_s. \tag{11}$$

It has been proven that the above problem (11) is an analogous formulation for maximizing SE [48]. This can be fairly evident since the optimal hybrid precoders are closely comparable to the unconstrained optimal fully digital precoder.

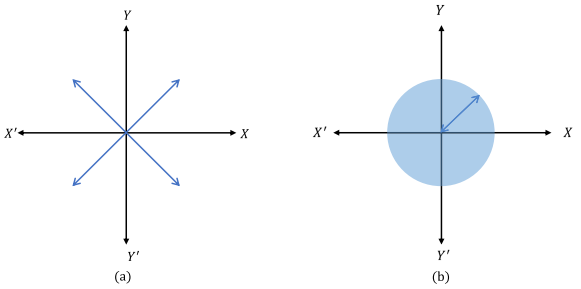

FIGURE 2. 2D projection of beamforming matrices: (a) Possible directions in which $\mathbf{F_{RF}F_{BB}}$ can point, directions are limited due to constraints on $\mathbf{F_{RF}}$; (b) $\mathbf{F_{opt}}$ can point in any direction.

It is well-known that a MIMO channel can be decomposed into several parallel independent channels [49]. If we have N_s independent channels, by multiplexing N_s independent data streams onto these channels, the data rate is increased by N_s times compared to a system with a single antenna at the transmitter and receiver. However, we must select N_s independent beam directions with the highest channel gains to obtain the maximum capacity. Now, the singular value decomposition (SVD) of the channel can be utilized to find optimal unconstrained beamformers (which maximize channel capacity), i.e., N_s singular vectors corresponding to the highest singular values of the channel matrix can be used to find N_s optimal beam directions [50]. As a result, the optimal fully digital precoder matrix $\mathbf{F}_{\text{opt}} \in \mathbb{R}^{N_t \times N_s}$ contains the first N_s columns of $\mathbf{V}$. Here, columns of $\mathbf{V}$ are the right singular vectors and are generated from the SVD of the channel $\mathbf{H}$, i.e., $\mathbf{H} = \mathbf{U} \sum \mathbf{V}^{\text{H}}$. The columns of $\mathbf{U}$ are the left singular vectors.

Once the beamforming matrices are calculated, SE is found by

$$R = \log_2 |\mathbf{I}_M + \frac{\rho}{N_s} \mathbf{W}_t (\mathbf{W}_t^H \mathbf{W}_t)^{-1} \mathbf{W}_t^H \mathbf{H} \mathbf{V}_t \mathbf{V}_t^H \mathbf{H}^H|, \tag{12}$$

where $\mathbf{W}_t = \mathbf{W}_{\text{RF}} \times \mathbf{W}_{\text{BB}}$, $\mathbf{V}_t = \mathbf{F}_{\text{RF}} \times \mathbf{F}_{\text{BB}}$ and ρ is the transmit power.

III. MATHEMATICAL MODELLING

Before delving into the details of our proposed system, let us take a quick look at how we mathematically model the problem and what our DNN's approach to solving it is. Let us first think of $\mathbf{F}_{\text{opt}}$ and $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ matrices as vectors in a high-dimensional complex plane, which we hypothetically project onto a 2D real plane here for a quick analogy. Due to the unit modulus constraint on $\mathbf{F}_{\text{RF}}$, the vector of $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ cannot point in all directions as shown in Figure 2(a). However, $\mathbf{F}_{\text{opt}}$, which is the unconstrained optimal beamforming matrix, can point in any given direction, as shown in Figure 2(b). Since $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ cannot point in every direction like $\mathbf{F}_{\text{opt}}$, $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ will not be equal to $\mathbf{F}_{\text{opt}}$ most of the time. Hence, there will be an error vector $\mathbf{e}$ between these two vectors, as shown in Figure 3. Our goal is to minimize this $\mathbf{e}$ in a much higher-dimensional complex plane. In other words, we need to find

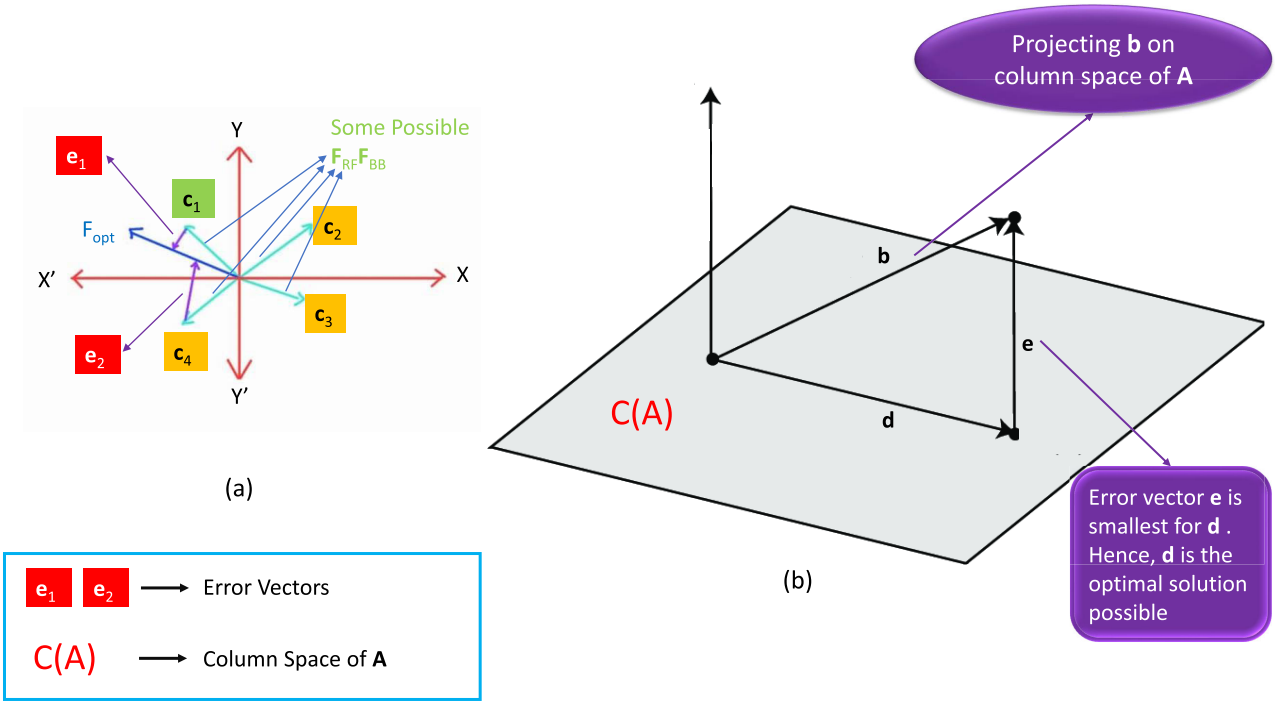

FIGURE 3. Hypothetical vector representation of beamforming matrix columns.

the $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ which is closest to $\mathbf{F}_{\text{opt}}$. In this analogy, from Figure 3, we observe that $\mathbf{c}_1$ is the closest possible choice because its projection on $\mathbf{F}_{\text{opt}}$ is the largest and hence its error $\mathbf{e}_1$ is also the smallest. We can also consider the beamforming problem (11) as a system of matrix equations of the form $\mathbf{Ax} = \mathbf{b}$, which we can solve for an exact $\mathbf{x}$ only when $\mathbf{b}$ is in the column space of $\mathbf{A}$. We can consider each column of $\mathbf{F}_{\text{BB}}$ and $\mathbf{F}_{\text{opt}}$ as vectors in ${N^t}_{\text{RF}}$ and N_t dimensional space respectively. This can be expressed in matrix form as

$$\mathbf{AX} = \mathbf{B}, \tag{13}$$

where

$$\mathbf{A} = \begin{bmatrix} A_{11} & \cdots & A_{1N_{\text{RF}}} \\ A_{21} & \cdots & A_{2N_{\text{RF}}} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ A_{N_t1} & \cdots & A_{N_tN_{\text{RF}}} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1N_s} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_{N_{\text{RF}}1} & \cdots & x_{N_{\text{RF}}N_s} \end{bmatrix}$$

and $\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1N_s} \\ b_{21} & \cdots & b_{2N_s} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ b_{N_t1} & \cdots & b_{N_tN_s} \end{bmatrix}$.

Here, matrix $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{X}$ corresponds to $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{opt}}$ and $\mathbf{F}_{\text{BB}}$ respectively. We now consider each column of $\mathbf{B}$ and $\mathbf{X}$ separately and resolve (13) into N_s equations of the form:

$$\mathbf{Ax} = \mathbf{b}, \tag{14}$$

where

$$\mathbf{x} = \begin{bmatrix} x_{1j} \\ \cdot \\ \cdot \\ \cdot \\ x_{N_{\text{RF}}j} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \cdot \\ \cdot \\ \cdot \\ b_{N_tj} \end{bmatrix} \text{ and } j = 1,2,3,\ldots,N_s.$$

Here, (14) can only be solved for $\mathbf{x}$ when the vector $\mathbf{b}$ is in the column space of the matrix $\mathbf{A}$. Unfortunately, due to the unit modulus constraint, our column space of $\mathbf{A}$ is severely limited. In simple words, the dimension C($\mathbf{A}$) will be much less than the dimension of $\mathbf{b}$. Therefore, we can rewrite (14) as

$$\mathbf{Ay} = \mathbf{d}, \tag{15}$$

where $\mathbf{d}$ will be the projection of $\mathbf{b}$ on the column space of $\mathbf{A}$ and

$$\mathbf{y} = \begin{bmatrix} y_{1j} \\ \cdot \\ \cdot \\ \cdot \\ y_{N_{\text{RF}}j} \end{bmatrix}, \mathbf{d} = \begin{bmatrix} d_{1j} \\ d_{2j} \\ \cdot \\ \cdot \\ \cdot \\ d_{N_tj} \end{bmatrix} \text{ and } j = 1,2,3,\ldots,N_s.$$

Now problem (11) involves solving $\mathbf{Ay} = \mathbf{d}$ as given by (15) for every column in $\mathbf{F}_{\text{BB}}$. The points below show that (11) essentially requires finding the best possible $\mathbf{d}$ for every $\mathbf{b}$ in $\mathbf{F}_{\text{opt}}$, i.e., minimizing $\mathbf{e}$ for every $\mathbf{b}$ as depicted in Figure 3(b).

- $\mathbf{e} = \mathbf{b} - \mathbf{d}$ and minimizing $\mathbf{e}$ means minimizing $|\mathbf{b} - \mathbf{d}|$.

- Again, $\mathbf{d} = \mathbf{Ay}$. Hence, we must minimize $|\mathbf{b} - \mathbf{Ay}|$ for every $\mathbf{b}$ and $\mathbf{y}$.
- But, all the $\mathbf{b}$'s are the columns of $\mathbf{F}_{\text{opt}}$, all the $\mathbf{y}$'s are the columns of the optimal $\mathbf{F}_{\text{BB}}$ and $\mathbf{A}$ is the $\mathbf{F}_{\text{RF}}$.
- Hence, minimizing $|\mathbf{b} - \mathbf{Ay}|$ for every b is the same as minimizing the Frobenius norm of $\mathbf{F}_{\text{opt}}$ - $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$, which is required by (11).

IV. PROPOSED DNN-BASED BEAMFORMING SYSTEM

This section presents the specifics of the proposed DNN system for hybrid beamforming. First, we describe the procedure by which simulation data for training, validation, and testing of the proposed system is created. The design and architecture of the neural network utilized to construct the beamforming matrices are then discussed. Then, we delineate how the input data and features travel through the network architecture to produce the desired outputs. Finally, we demonstrate how the beamforming matrices are computed using the neural network's outputs.

A. DATA GENERATION AND PRE-PROCESSING

Our proposed DNN actually solves the beamforming problem by essentially solving the $\mathbf{Ay} = \mathbf{d}$ problem as given in (15). Each training sample consists of one column of $\mathbf{F}_{\text{BB}}$(=$\mathbf{y}$) and the corresponding column from $\mathbf{F}_{\text{opt}}$(=$\mathbf{b}$) as illustrated below

$$\begin{bmatrix} A_{11} & \cdots & A_{1N_{\text{RF}}} \\ A_{21} & \cdots & A_{2N_{\text{RF}}} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ A_{N_{t1}} & \cdots & A_{N_tN_{\text{RF}}} \end{bmatrix} \times \begin{bmatrix} y_{11} & \cdots & y_{1N_s} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ y_{N_{RF1}} & \cdots & y_{N_{\text{RF}}N_s} \end{bmatrix} \approx \begin{bmatrix} b_{11} & \cdots & b_{1N_s} \\ b_{21} & \cdots & b_{2N_s} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ b_{N_{t1}} & \cdots & b_{N_tN_s} \end{bmatrix}. \tag{16}$$

Here, the column $[y_{11}, y_{21}, \ldots, y_{N_{\text{RF}}1}]^{\text{T}}$ of $\mathbf{F}_{\text{BB}}$ and corresponding column $[b_{11}, b_{21}, \ldots, b_{N_t1}]^{\text{T}}$ of $\mathbf{F}_{\text{opt}}$ constitute one training sample. Now, we look at exactly how the ground truth of our DNN is generated. As we can see, we treat every column of $\mathbf{F}_{\text{opt}}$ and $\mathbf{F}_{\text{BB}}$ as an individual training sample. Our DNN learns by looking at thousands of possible pairs of $\mathbf{y}$ and $\mathbf{b}$. We also observe that each column of $\mathbf{F}_{\text{opt}}$ is related to only the corresponding column of $\mathbf{F}_{\text{BB}}$. Our DNN learns to predict columns of $\mathbf{F}_{\text{BB}}$ by looking at columns of $\mathbf{F}_{\text{opt}}$. Finally, $\mathbf{F}_{\text{RF}}$ is calculated from $\mathbf{F}_{\text{opt}}$ and $\mathbf{F}_{\text{BB}}$ through pseudo-inverse.

In order to produce simulation data for training, validation, and testing, we start by generating thousands of instances of the channel matrix $\mathbf{H}$ from the channel model described in section II (Figure 4 (1)). Then, corresponding to each instance of the channel matrix, we calculate the fully digital optimal beamforming matrix $\mathbf{F}_{\text{opt}}$ from the SVD of $\mathbf{H}$ (Figure 4 (2)). We then utilize the existing numerical iteration-based algorithms such as the OMP and Alt-Min to find the beamforming matrices $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$, $\mathbf{W}_{\text{RF}}$, $\mathbf{W}_{\text{BB}}$ (Figure 4 (3)). Since most neural networks are not configured to work with complex numbers, the complex numbers are resolved into their polar form. Furthermore, we split up the phase and magnitude of the complex matrices (Figure 4 (4)). Now, these matrices can be treated in the same way as ordinary real number matrices. Finally, following standard practice, the input data is normalized before being passed onto the network. Mathematically, the magnitude normalization can be expressed as

$$v_M(i) = \frac{u_M(i)}{\max(u_M)}. \tag{17}$$

Here, $u_M(i)$ denotes the magnitude of any element of the matrix $\mathbf{u} \in [\mathbf{F}_{\text{opt}}, \mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}]$, max(.) gives the global maximum, and $v_M(i)$ denotes the normalized magnitude input that would be passed to the neural network. Similarly, the phase is also normalized and is given by

$$v_P(i) = \frac{u_P(i) + \pi}{2\pi}, \tag{18}$$

where $u_P(i)$ denotes the phase of any element of the matrix $\mathbf{u} \in [\mathbf{F}_{\text{opt}}, \mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}]$ and $v_P(i)$ denotes the normalized phase input that would be passed to the neural network. Moreover, according to the mathematical modelling presented in section III, we split up the matrices $\mathbf{F}_{\text{opt}}$, $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$ into their constituent columns, that is we vectorize them (Figure 4 (5)). Finally, these resolved columns of $\mathbf{F}_{\text{opt}}$ are passed onto the input layer of the neural network (Figure 4 (6)). This entire workflow is summarized in Figure 4.

B. ARCHITECTURE OF THE PROPOSED DNN

The primary goal of our neural network-based approach is to minimize latency as well as achieve a high sum rate. Here, we refer to latency in the context of inference time, the time the neural network needs to estimate the beamforming matrices (during real-time operation). Since the transmitter and receiver employ hybrid beamforming to send and receive data, the latency is related to the time required to generate the beamforming matrices. Consequently, our objective is to reduce the amount of time necessary to generate these matrices utilizing our DNN-based approach, which we refer to as minimizing the latency. With that in mind, we propose a lightweight 1D CNN-LSTM-based architecture with a comparatively smaller number of parameters than other contemporary networks and estimate the beamforming matrices with high accuracy. A vivid schematic of our proposed network is shown in Figure 5. Here, the dropout layers are omitted for visual clarity. The input to the neural network is two vectors of size $(N_t, 1)$.[1] The normalized resolved components of $\mathbf{F}_{\text{opt}}$ columns constitute the inputs. The phase and magnitude inputs are separately passed onto two independent input layers. In each branch, after the input layer, we have

[1] At the receiver end, the input would be $(N_r, 1)$.

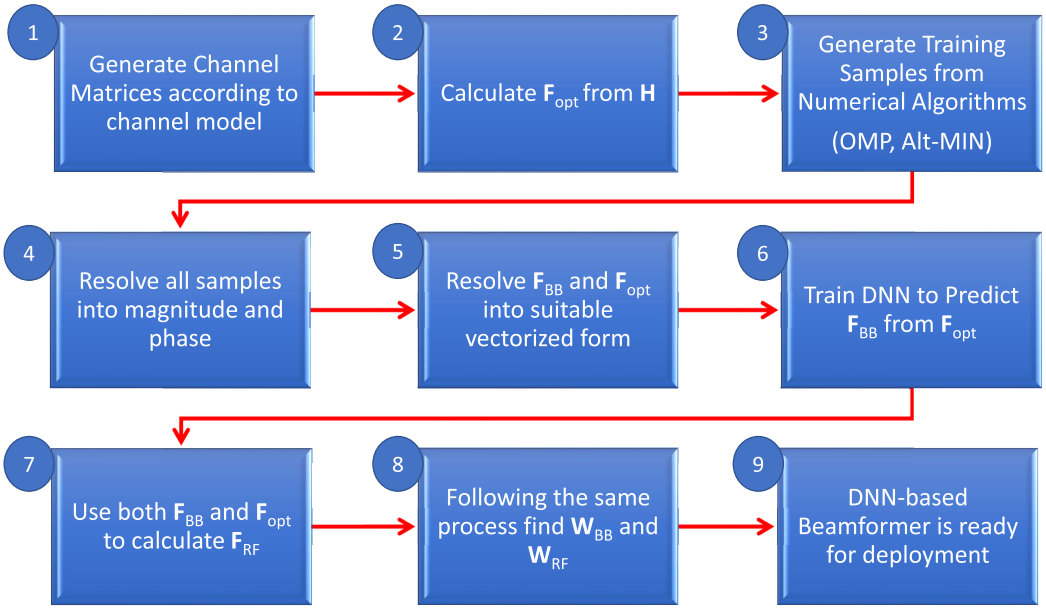

FIGURE 4. Workflow of the proposed hybrid beamforming system.

three blocks of 1D convolution, where each block has two convolution layers followed by a max-pooling layer. The first convolutional block does not have zero padding to avoid superfluous noise insertion in the initial convolutional layers. Nevertheless, the rest of the convolutional blocks apply zero padding to prevent the feature dimensions from becoming too small. After performing thousands of instances of model training, it is observed that using fewer than three blocks reduces the accuracy of the DNN. Conversely, using four or more blocks generates an unnecessary computational burden and gradient underflow with no noticeable improvement in accuracy. Then we use a 1D depth-wise separable convolution block, which again consists of two depth-wise separable convolution layers followed by a pooling layer. We perform extensive trials and errors to determine the specific number of separable convolutional layers used in the model.

After extracting features from the convolutions, we flatten the branches and add a dense layer. We then concatenate (fusion) both branches and pass them through two more dense layers. These dense layers change the dimension of the output features from the convolutional network and appropriately shape the inputs to the subsequent LSTM layers. We then use a bidirectional LSTM block with two LTSM layers. Empirically, after performing thousands of training simulations, we recognize that only two LSTM layers are sufficient to provide optimal accuracy, as observed in Figure 6. Since LSTM layers are computationally pretty expensive, using as few as possible is desirable. Subsequently, the two branches are again separated, and each branch is processed through a block of dense layers to generate the final outputs. The final outputs are two vectors of size (N_{RF}, 1), one for the magnitude and one for the phase.[2]

[2]Here, N_{RF} would be $N^t{}_{\mathrm{RF}}$ and $N^r{}_{\mathrm{RF}}$ at the transmitter and receiver end, respectively.

The outputs of each convolution layer are passed through a Swish activation function, which performs slightly better than the ReLU for Convolution [51]. The swish activation function for any input ξ is given by

$$\mathrm{swish}(\xi) = \xi \times \frac{1}{1 + e^{-\beta\xi}}, \tag{19}$$

where β is a trainable parameter.

Additionally, we use ReLU activation after each dense layer. After every convolution block, we use a 1D spatial dropout layer with a dropout value set to 0.03, while after every dense block, we use Alpha-dropout with a dropout value of 0.04. These dropout layers ensure better generalization and avoid overfitting [52]. In addition, batch-normalization layers are used after each block, and L1-L2 regularizers are incorporated into each layer to accelerate the training process [53]. It can be observed that the separation network for the phase has more depth. This is because, empirically, we observe that SE is relatively more sensitive to slight changes in phase than in magnitude after much trial and error. Therefore, estimating the phase more precisely using a deeper network is vital. We also formulate two other lighter network variants, with all the layers being the same as described above, the only difference being the number of neurons in the layers. Hence, the three variants have a different number of total parameters. The largest one has 2.4 million parameters, which we call DNN-Large. The other variants with 1.3 million and 864,000 parameters are aptly named DNN-Medium and DNN-Small, respectively. The number of neurons, kernel sizes, pool sizes and the number of filters in the layers of each model are determined through extensive simulations.

C. DATAFLOW IN THE NETWORK

In order to efficiently and precisely compute the beamforming matrices $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$, we compute their constituent

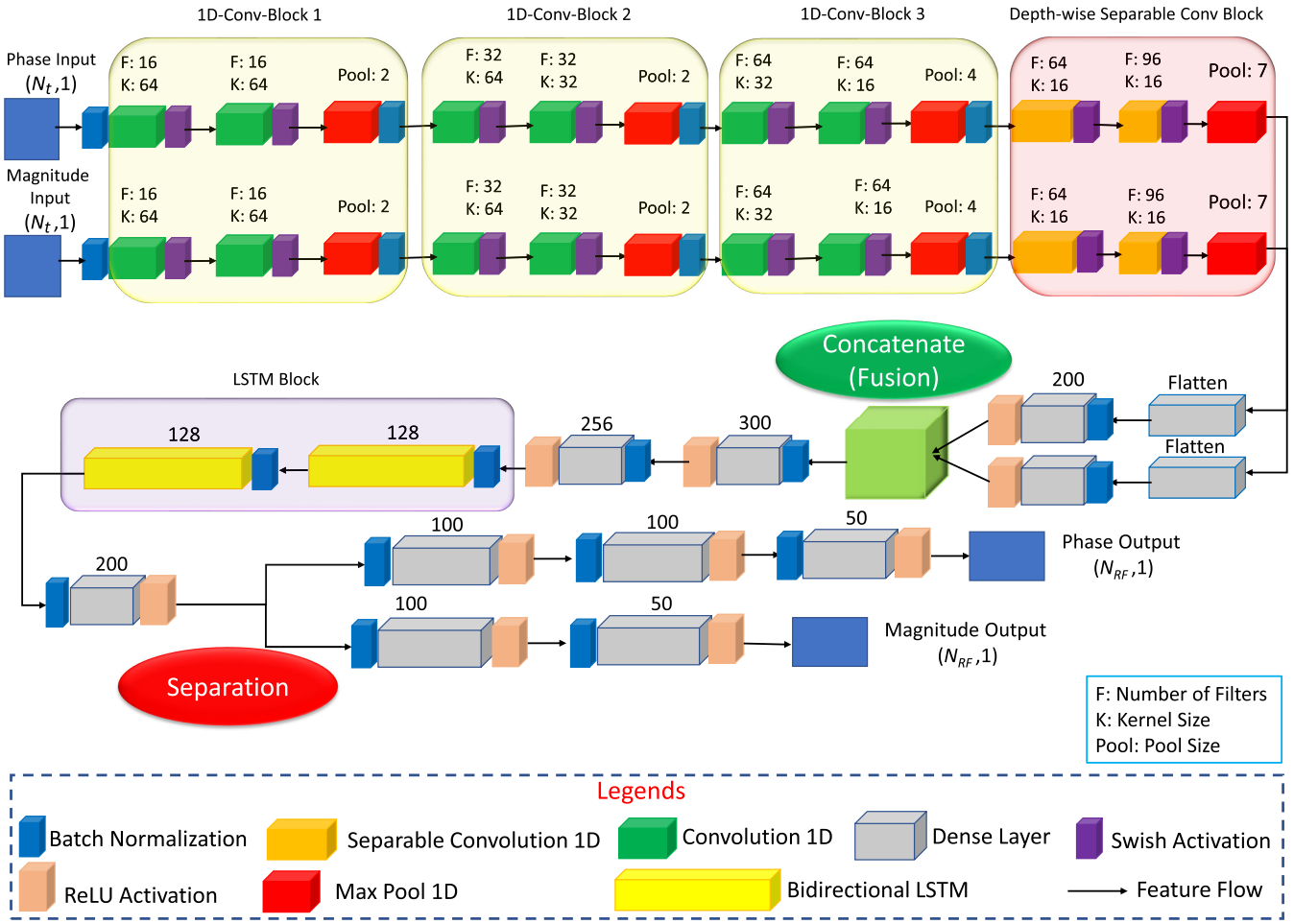

FIGURE 5. Architecture of the proposed 1D CNN-LSTM-based fusion separation DNN.

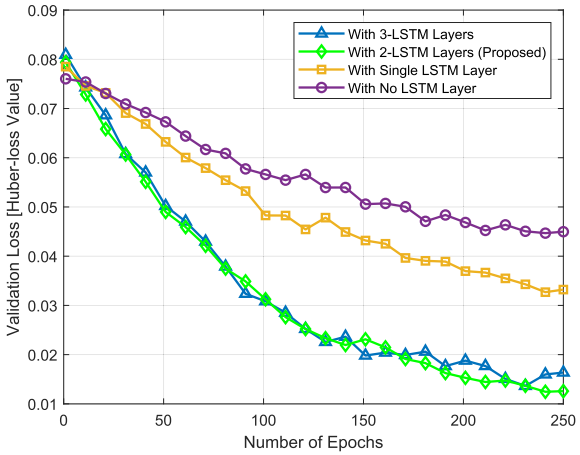

FIGURE 6. Ablation study for the LSTM layers.

columns separately. We first pass the columns of $\mathbf{F}_{\text{opt}}$ as input to our neural network. The phase and magnitude of elements of $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$ are independently processed through a 1D-Convolutional network first. This convolutional network extracts unique features from phase and magnitude autonomously. Mathematically, for each feature map output $\mathbf{o}[i]$, this can be expressed as

$$\mathbf{o}[i] = \sum_{n=0}^{N} (\boldsymbol{\chi}[i+n] \times \mathbf{k}[n]), \tag{20}$$

where $\boldsymbol{\chi}$ is the input data, $\mathbf{k}$ is the kernel coefficients, and N is the length of the kernel.

Then a depth-wise separable convolutional network is used to simultaneously reduce the total parameters of the network and extract more distinct features from the input data. Since the elements of $\mathbf{F}_{\text{BB}}$ are related to both the phase and the magnitude of $\mathbf{F}_{\text{opt}}$, the independently extracted features from both the phase and magnitude branches must be combined. This is done in a fusion network consisting of dense and LSTM layers. The LSTM layers are particularly adept at extracting complex relations between sequential and fragmented data. Since the properties of a beamforming matrix depend not only on the values of its constituent elements but also on the position in which the elements are present, we can think of the elements as a sequence of values. The LSTM layers help

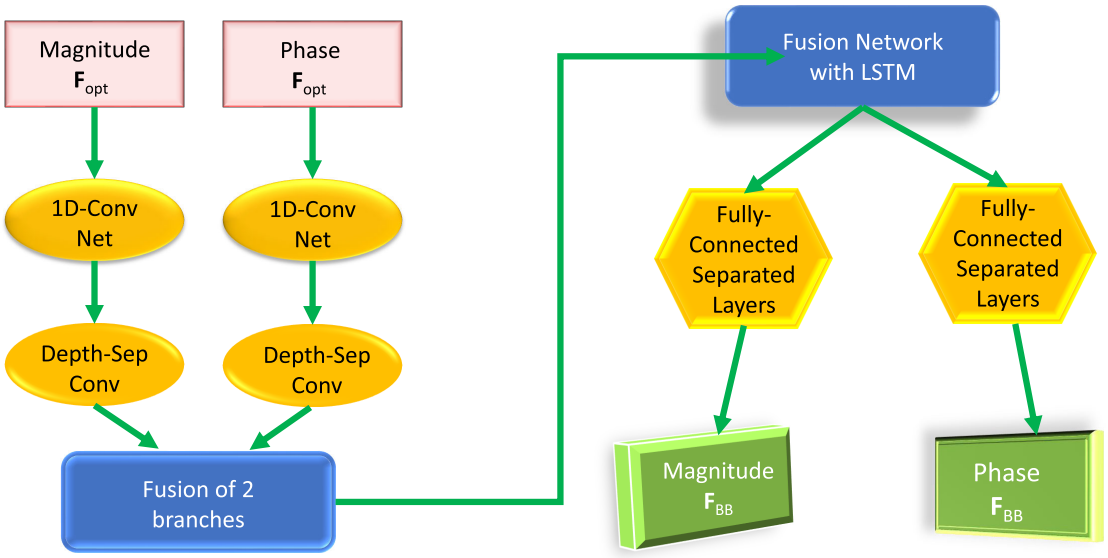

FIGURE 7. Data-flow in the proposed 1D CNN-LSTM-based architecture.

to extract these very sequential features from the data. The performance of the proposed DNN degrades considerably without the use of LSTM layers, as can be observed from the loss curve in Figure 6. In this fusion network, features extracted from the phase and magnitude of $\mathbf{F}_{\text{opt}}$ are combined and generate new components, which are now passed on to two separate branches. This separation network is used to predict the phase and magnitude of elements of $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$ independently. Due to the use of 1D CNN and LSTM along with fusion and separation of two discrete branches as shown in Figure 7, we name our network as 1D CNN-LSTM fusion-separation network.

D. POST PROCESSING AND GENERATING BEAMFORMING MATRICES

The outputs of the neural networks are normalized vectors which are, in fact, the phase and magnitude of columns of $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$. In order to generate the true baseband matrices, we need to reverse the normalization done in the pre-processing. Then the elements of the columns of $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{BB}}$ are given by

$$\begin{aligned} z_P(i) &= k_P(i) \times 2\pi - \pi \\ z(i) &= k_M(i) \cos z_P(i) + j k_M(i) \sin z_P(i), \end{aligned} \tag{21}$$

where $k_P(i)$ and $k_M(i)$ denote the ith normalized phase and magnitude output from DNN, respectively. $z(i)$ represents the ith element of a single column of the digital baseband beamforming matrices. After that, we concatenate the columns together to form the initial matrices $\mathbf{F}_{\text{BB}i}$ and $\mathbf{W}_{\text{BB}i}$. Now, we find the analog beamforming matrices $\mathbf{F}_{\text{RF}}$ and $\mathbf{W}_{\text{RF}}$ using:

$$\mathbf{F}_{\text{RF}i} = \mathbf{F}_{\text{opt}} \dagger \mathbf{F}_{\text{BB}i} \tag{22}$$

and

$$\mathbf{W}_{\text{RF}i} = \mathbf{W}_{\text{opt}} \dagger \mathbf{W}_{\text{BB}i}. \tag{23}$$

However, to ensure the element-wise unit modulus constraints, we apply the following transformation

$$\mathbf{F}_{\text{RF}} = \cos \angle \mathbf{F}_{\text{RF}i} + j \sin \angle \mathbf{F}_{\text{RF}i} \tag{24}$$

and

$$\mathbf{W}_{\text{RF}} = \cos \angle \mathbf{W}_{\text{RF}i} + j \sin \angle \mathbf{W}_{\text{RF}i}. \tag{25}$$

Here, $\angle F_{\text{RF}i}$ and $\angle W_{\text{RF}i}$ denote the phase of initial analog beamforming matrices. Again, to satisfy the normalized transmit power constraint of (11) we utilize the following transformation

$$\mathbf{F}_{\text{BB}} = \frac{\sqrt{N_s} \times \mathbf{F}_{\text{BB}i}}{\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}i}\|_F} \tag{26}$$

and

$$\mathbf{W}_{\text{BB}} = \frac{\sqrt{N_s} \times \mathbf{W}_{\text{BB}i}}{\|\mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}i}\|_F} \tag{27}$$

Thus, we obtain all four beamforming matrices.

V. SIMULATION SETUP

A. SIMULATION ENVIRONMENT

Google Colaboratory [54], a Python development environment that runs in the browser utilizing Google Cloud and provides free access to strong graphical processing units (GPU), is used to conduct all the simulations. Our proposed 1D-CNN-LSTM system and accompanying peripherals are implemented in Python 3.7.11 using TensorFlow 2.7.0 [55] and a Tesla T4 GPU supplied by Google Collaboratory. To ensure a true and fair comparison of the execution time of various processes, all codes are executed using the same configuration.

B. DATASET

We start by generating 100,000 channel realizations ($\mathbf{H}$) and corresponding $\mathbf{F}_{\text{opt}}$, $\mathbf{F}_{\text{BB}}$, $\mathbf{F}_{\text{RF}}$, $\mathbf{W}_{\text{opt}}$, $\mathbf{W}_{\text{BB}}$ and $\mathbf{W}_{\text{RF}}$. Then, each $\mathbf{F}_{\text{opt}}$, $\mathbf{F}_{\text{BB}}$, $\mathbf{W}_{\text{opt}}$ and $\mathbf{W}_{\text{BB}}$ are resolved into N_s columns. The total number of training samples thus becomes 500,000 for $N_s = 5$. Generating this huge amount of data using the MO-Alt-Min algorithm would require almost an eternity. Hence, most of the data (90%) are generated using the OMP algorithm, and the model is first trained on that data. After that, transfer learning is used to carry the weights from this initial training phase, and the model is then trained on MO-Alt-Min generated data, constituting only 10% of the entire dataset. We also generate 10,000 channel realizations to produce the test dataset. Unless otherwise mentioned, all experiments were conducted with $N_t = N_r = 256$, $N_s = 5$, $N_{\text{RF}} = 5$, $\lambda = 0.1$ cm ($f = 0.3$ THz), and $d = 0.05$ cm.

C. SIMULATION PARAMETERS AND MODEL TRAINING

After data generation is finished, we split the generated training data into a training data set and a validation data set in the typical ratio of 8:2. The training process optimizes the weight of the trainable network parameters through back-propagation, and we choose the Adagrad algorithm as the optimization algorithm. Adagrad, also known as adaptive gradient, permits the learning rate to adapt based on specified parameters. It executes larger changes for infrequent parameters and more minor changes for frequently updated ones. This makes it well-suited for sparse data [56]. Additionally, we use Huber loss as our loss function, which is quite well suited to regression tasks as it fuses the best attributes of the mean square error and mean absolute error losses [57]. For each value ϵ in error = y_{true}-y_{pred}, loss is given by

$$\text{loss} = \begin{cases} 0.5 \times \epsilon^2 & \text{if } |\epsilon| \leq \psi \\ 0.5 \times \epsilon^2 + \psi \times (|\epsilon| - \psi) & \text{if } |\epsilon| > \psi \end{cases}, \quad (28)$$

where 'ψ' is a constant parameter that determines the exact behavior of the loss towards outliers. We choose a typical value of $\psi = 1$ for all simulations.

We choose a batch size of 1024 through empirical trial and error. The learning rate, however, is varied throughout the whole training process, and we use a stepwise constant decay scheduler. Figure 8 shows how the proposed scheduler adapts the learning rate based on the training progress. Scheduling the learning allows the model to reach a stable minima much faster than a constant learning rate. As we can observe from figure 8, the learning rate also varies according to the size of the model. The largest model, with 2.4 million parameters, has a higher learning rate because it starts at a higher point in the loss curve than the smaller models, as shown in Figure 9. Although our network has dropout layers, we also apply both L1 and L2 penalties for further regularization [53]. Here, the values of L1 and L2 are empirically chosen to be 0.000001. We monitor the learning progress of the network through appropriate validation data set performance metrics.

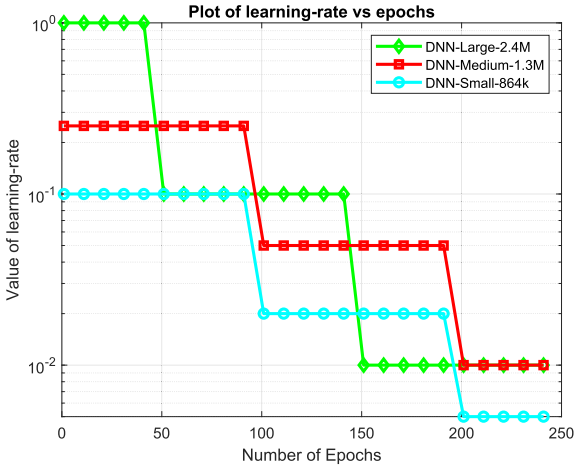

FIGURE 8. Learning rate used for different model sizes with the number of epochs.

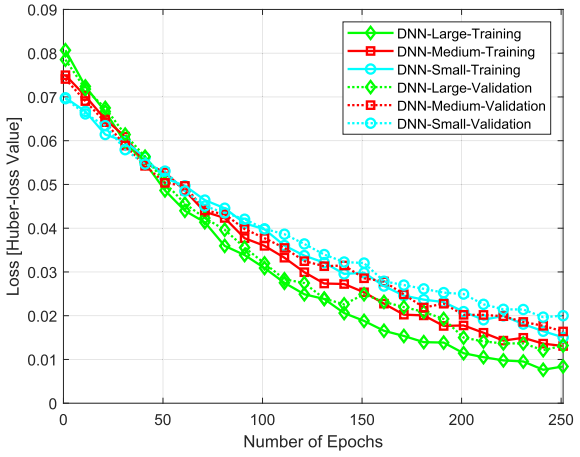

FIGURE 9. Training and validation loss of different model sizes with the number of epochs.

The performance metrics for the validation set are the mean square error and the mean absolute percentage error.

VI. RESULTS

In this section, we present the simulation results and corresponding analysis to validate the proposed beamforming system. We evaluate the performance of our proposed system in terms of SE and computational time. Moreover, we compare our approach to the already established beamforming algorithms and analyze their advantages and drawbacks.

A. SPECTRAL EFFICIENCY COMPARISON

In this segment, we compare the SE performance of our proposed system with the existing beamforming algorithms such as the MO-Alt-Min, OMP, PE-Alt-Min, etc. We also show the SE achieved by fully digital beamforming for benchmarking purposes.

In order to find the SE, we first calculate the beamforming matrices using the existing beamforming algorithms, such

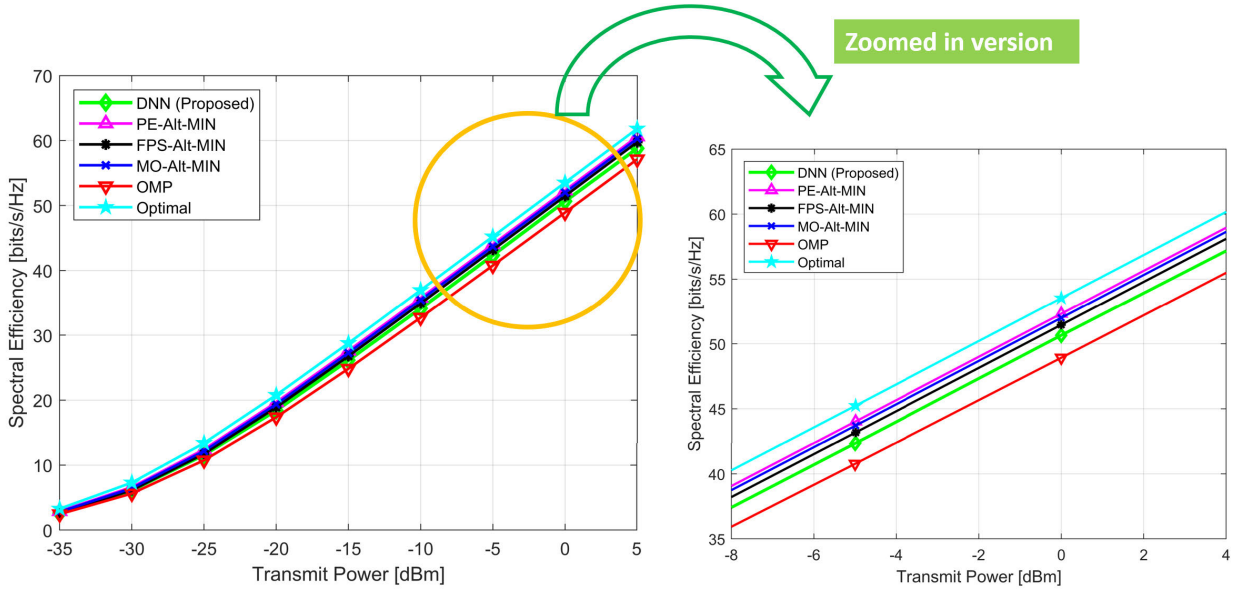

FIGURE 10. SE of different algorithms for varying transmit power with $N_t = N_r = 256$ and $N_s = N_{RF} = 5$.

as the MO-Alt-Min, OMP, PE-Alt-Min, and FPS-Alt-Min for all the 10,000 test channel realizations. Additionally, the fully digital optimal beamforming matrices are found simply by the singular value decomposition of the channel matrices. Subsequently, the beamforming matrices for our proposed method are generated using the approach described in section IV. Now, the SE of all the algorithms is computed using their respective matrices. We plot the SE along with the transmit power to show the potency of our proposed scheme, which is presented in Figure 10. It can be seen that the proposed method achieves almost the same SE as that of the Alt-Min algorithms and outperforms the OMP algorithm by quite a margin. To observe the slight differences in achieved SE, we take a closer look at the plot and see that our neural network-based beamformer approximately achieves only 1.6 bits/s/Hz less than the most spectrally efficient MO-Alt-Min and PE-Alt-Min. Unsurprisingly, we also observe that as the transmit power increases, the SE of all the algorithms increases, but they are always less than the SE achieved by the fully digital beamformer.[3] The fully digital beamformer requires as many RF chains as the number of antennas it has. This is both costly in terms of hardware and also has much higher power consumption. Through hybrid beamforming, we are effectively trading off a small amount of SE for a large reduction in power consumption and hardware costs. To better understand this trade-off, let us consider the power consumption of fully digital beamforming and hybrid beamforming for $N_t = 256$ and $N_s = N_{RF} = 5$ in THz scenarios. For fully digital beamforming, each antenna must be connected to an RF chain followed by a DAC. Hence, 256 RF chains and 256 DACs must be used. On the other hand, for hybrid beamforming with $N_s = N_{RF} = 5$, only 5 RF chains and 5 DACs are needed. It is to be noted that each RF chain consists of a mixer and an oscillator which consume 22 mW [58] and 4 mW [59] power, respectively, and each DAC consumes 110 mW [60] at THz frequencies. This implies that in the above-mentioned scenario, the power consumption of digital forming would be almost 34 W (51 times) higher than that of hybrid beamforming. However, as we can see from Figure 10, the SE increases only by 2-3 bits/s/Hz (less than 5%) for such a massive increase in power consumption. This brief account shows the potential of hybrid beamforming for reducing power consumption in THz frequencies.

[3]We designate the fully digital beamformer as 'Optimal' in all relevant figures.

B. RELATIVE COMPUTATIONAL TIME

Our primary objective is to reduce the computation time required for real-time hybrid beamforming in order to support THz communication in 6G networks. In this sub-section, we compare the performance of our proposed DNN-based beamforming approach to that of the well-known Alt-Min algorithms in terms of computational time. In order to obtain infallible validity, the computational time for each algorithm is determined by simulating 10,000 channel realizations and then averaging the results. The results are presented in Figure 11. We observe that our proposed DNN-based beamformer is almost 100 times faster than the fastest Alt-Min algorithm, the PE-Alt-Min. As the number of transmitter antennas increases, the computational requirements of the Alt-Min algorithms rise quite steeply because these algorithms rely on numerical iterations, which become more computationally expensive as the number of input variables involved increases. However, for our proposed approach,

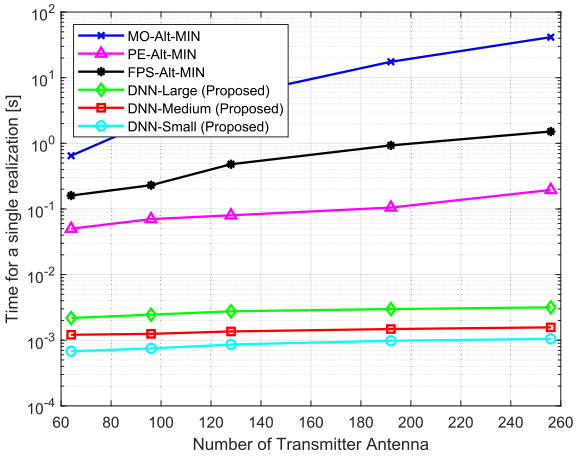

FIGURE 11. Computational time required by different beamforming algorithms with the number of transmitter antennas. The number of receiver antennae is always kept the same as the number of transmitter antennae ($N_t = N_r$).

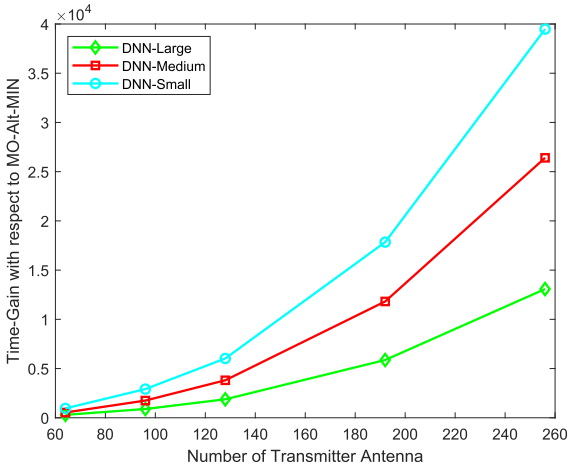

FIGURE 12. Relative computation gain of different model sizes with the number of transmitter antenna (Here, $N_r = N_t$).

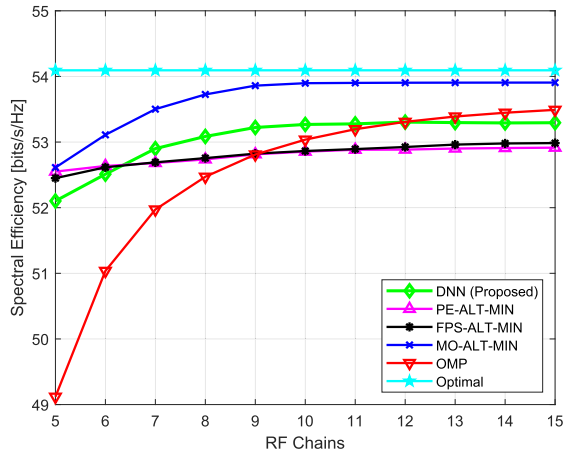

FIGURE 13. SE of different algorithms with respect to the number of RF chains.

increasing the number of antennas only requires the input layer to be changed, while the rest of the network architecture remains exactly as predefined. Hence, only a slight increase in inference time is observed. This is a major advantage of the proposed neural network-based beamforming scheme. We also observe that the computational time of the DNN-Small network is the lowest, which is due to the fact that it has the least number of parameters. However, we shall see in the subsequent discussion that this reduction in computational cost comes at the price of lower SE. For a more comprehensive comparison between the three neural network architectures, we examine their relative time gain with respect to the MO-Alt-Min algorithm. Figure 12 reveals that for 256 transmitter antennas, the largest DNN model is approximately 10,000 times faster than the MO-Alt-Min, and the smallest one is around 40,000 times faster. We have also evaluated the number of parameters and computational time for the CNN model proposed in [39]. It is observed that the CNN model has 136 million parameters for $N_t = N_r = 64$ with an inference time of 4.88 ms and 304 million parameters for $N_t = N_r = 96$ with an inference time of 7.15 ms, compared to 2.18 ms and 2.45 ms taken by our proposed DNN-Large model for $N_t = N_r = 64$ and $N_t = N_r = 96$, respectively. Additionally, the CNN model only achieves approximately 72% of the SE achieved by our model. To ensure a fair and accurate comparison, all the simulation setups were kept exactly the same for both models.

C. PERFORMANCE WITH VARYING NUMBERS OF RF CHAINS

One of the vital considerations in hybrid beamforming design is the number of RF chains at the transmitter and the receiver. Reducing the number of RF chains as compared to fully digital beamforming has been the primary reason hybrid beamforming research has blossomed. With that in mind, one of the major deliberations in developing our proposed approach is to maintain high SE even when the number of RF chains available varies. Here, we demonstrate the results achieved by our proposed scheme even when the number of RF chains varies substantially. Simulation results in Figure 13 show that our DNN model outperforms all other algorithms except the MO-Alt-Min in most scenarios. Here, the transmit power is kept constant at 0 dBm, and the number of RF chains is varied from 5 to 15. Here, we have kept the minimum number of RF chains at 5 since $N_{\mathrm{RF}} < N_s$ is not possible. As the data for the DNN training is generated using the MO-Alt-Min, it is not possible for our supervised learning method to outperform its own data-generating algorithm. Nevertheless, our proposed neural network achieves almost the same SE as that of the MO-Alt-Min, even when the number of RF chains varies significantly. It has been proven that when the number of RF chains reaches twice the number of data streams, the SE of hybrid beamforming saturates and becomes almost the same as that of fully digital beamforming [5]. In our simulations, there are five independent data streams. Unsurprisingly, we observe that for 10 or more RF chains, the sum rate of

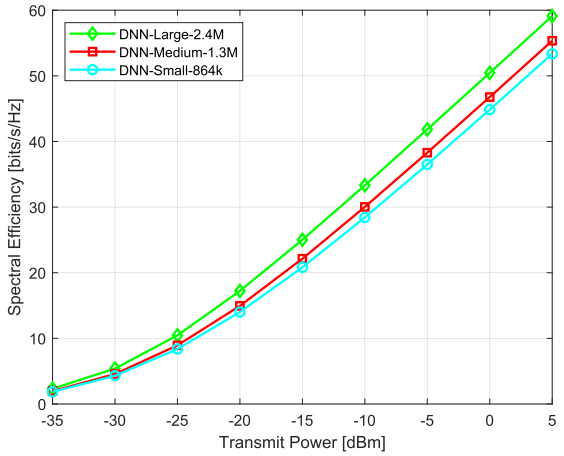

FIGURE 14. SE of different model sizes for varying transmit power with $N_t = N_r = 256$ and $N_s = N_{RF} = 5$.

all the algorithms saturates and is almost similar to that of the optimal fully digital beamformer. As the training data of our model comes from MO-Alt-Min, it also achieves an unfluctuating SE when the number of RF chains is increased beyond 10.

D. TRADE-OFF BETWEEN SPECTRAL EFFICIENCY AND MODEL-SIZE

As described briefly in the earlier discourse, we propose three slightly different neural networks with a distinct number of total parameters to explore the potential trade-off that arises between the achieved SE and the computational gain. As shown in Figure 12, the smallest DNN model (DNN-Small) is almost four times as fast as the largest one (DNN-Large) due to its significantly fewer number of parameters, which generates a much lower computational overhead. However, this benefit comes at the expense of reduced SE, which becomes evident from Figure 14. The number of parameters of the model is decreased simply by reducing the number of neurons in each layer. Hence, the smaller models rely on fewer features to make predictions about the beamforming matrices. This makes the predictions less accurate than those of the largest one and consequently results in a lower SE. This trade-off can be well utilized when there are different QoS requirements and in various distinct network scenarios. For instance, if a high data rate is required with no stringent constraints on latency, then the largest model can be deployed to enhance the SE. On the other hand, if latency constraints are more rigorous, such as in ultra-reliable low-latency communications (URLLC), we may opt for the lightest DNN-Small model, which has only 864,000 parameters.

VII. CONCLUSION AND FUTURE WORK

This paper has developed a novel 1D-CNN-LSTM-based fusion-separation DNN for fully connected hybrid beamforming, and contextual mathematical modelling has also been shown. Specifically, we have devised a beamforming method to enable UM-MIMO communications at THz frequencies by employing some of the most influential and cutting-edge deep-learning techniques. To validate our claim, we have undertaken extensive simulations to illustrate the efficacy and performance of our approach. We have analyzed the numerical results in terms of SE and computational complexity by varying pertinent parameters such as the number of antennas, the number of RF chains, and the transmit power. The simulation results have proven that the proposed approach can achieve almost the same SE as achieved by the well-known MO-based beamforming algorithms while significantly outperforming them in computational gain. In the future, we will pursue research into partially connected beamforming systems, such as dynamic arrays of sub-arrays, considering imperfect CSI and energy efficiency.

ACKNOWLEDGMENT

This research is in collaboration with Cisco Innovation Central Canberra.

REFERENCES

[1] S. Zhang, C. Xiang, and S. Xu, "6G: Connecting everything by 1000 times price reduction," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 107–115, 2020.
[2] C.-X. Wang, J. Wang, S. Hu, Z. Jiang, J. Tao, and F. Yan, "Key technologies in 6G terahertz wireless communication systems: A survey," *IEEE Veh. Technol. Mag.*, vol. 16, no. 4, pp. 27–37, Dec. 2021.
[3] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "Terahertz-band ultra-massive spatial modulation MIMO," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2040–2052, Jul. 2019.
[4] J. Zhang, X. Yu, and K. B. Letaief, "Hybrid beamforming for 5G and beyond millimeter-wave systems: A holistic view," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 77–91, 2020.
[5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
[6] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
[7] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[8] D. Dampahalage, K. B. S. Manosha, N. Rajatheva, and M. Latva-Aho, "Supervised learning based sparse channel estimation for RIS aided communications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 8827–8831.
[9] Y. Chen, L. Yan, C. Han, and M. Tao, "Millidegree-level direction-of-arrival estimation and tracking for terahertz ultra-massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 869–883, Feb. 2022.
[10] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "AI models for green communications towards 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 210–247, 1st Quart., 2022.
[11] H. Zhou, M. Erol-Kantarci, and H. V. Poor, "Learning from peers: Deep transfer reinforcement learning for joint radio and cache resource allocation in 5G RAN slicing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 4, pp. 1925–1941, Dec. 2022.
[12] R. U. Murshed, A. H. Hridhon, and M. F. Hossain, "Deep learning based power allocation in 6G URLLC for jointly optimizing latency and reliability," in *Proc. 5th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2021, pp. 1–6, doi: 10.1109/EICT54103.2021.9733558.
[13] V. Venkateswaran and A. Van Der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[14] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60 GHz WPAN," in *Proc. 6th Int. ICST Conf. Commun. Netw. China (CHINACOM)*, Aug. 2011, pp. 361–365.

[15] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6, doi: 10.1109/GLOCOM.2011.6134486.

[16] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[17] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[18] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Flexible unsupervised learning for massive MIMO subarray hybrid beamforming," in *Proc. GLOBECOM - IEEE Global Commun. Conf.*, Dec. 2022, pp. 3833–3838.

[19] K. Chen, J. Yang, Q. Li, and X. Ge, "Sub-array hybrid precoding for massive MIMO systems: A CNN-based approach," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 191–195, Jan. 2021.

[20] C. Lin and G. Y. Li, "Energy-efficient design of indoor mmWave and sub-THz systems with antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4660–4672, Jul. 2016.

[21] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Feb. 2016.

[22] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3403–3418, Jul. 2018.

[23] N. Hassan, K.-L. A. Yau, and C. Wu, "Edge computing in 5G: A review," *IEEE Access*, vol. 7, pp. 127276–127289, 2019.

[24] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter, and S. Rangan, "A case for digital beamforming at mmWave," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 756–770, Feb. 2020.

[25] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Unsupervised deep learning for massive MIMO hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7086–7099, Nov. 2021.

[26] D. Zhu, B. Li, and P. Liang, "A novel hybrid beamforming algorithm with unified analog beamforming by subspace construction based on partial CSI for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 594–607, Feb. 2017.

[27] S. Montori, C. Fritzsch, L. Marcaccioli, R. V. Gatti, R. Jakoby, and R. Sorrentino, "Design and measurements of a 1-bit reconfigurable elementary cell for large electronic steerable reflectarrays," in *Proc. 40th Eur. Microw. Conf.*, 2010, pp. 918–921.

[28] Z. El Jaoussi and A. Haddi, "Large MIMO systems and hybrid beamforming in 5G or future communication," *Int. J. Eng. Res. Technol.*, vol. 13, no. 5, p. 1039, May 2020.

[29] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[30] F. Sohrabi and W. Yu, "Hybrid beamforming with finite-resolution phase shifters for large-scale MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 136–140.

[31] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[32] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[33] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, Jul. 2005.

[34] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501–3511, Dec. 2006.

[35] A. F. Molisch, M. Z. Win, and J. H. Winters, "Reduced-complexity transmit/receive-diversity systems," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2729–2738, Nov. 2003.

[36] A. F. Molisch and X. Zhang, "FFT-based hybrid antenna selection schemes for spatially correlated MIMO channels," *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 36–38, Jan. 2004.

[37] A. M. Elbir, K. V. Mishra, M. R. B. Shankar, and B. Ottersten, "A family of deep learning architectures for channel estimation and hybrid beamforming in multi-carrier mm-Wave massive MIMO," 2019, *arXiv:1912.10036*.

[38] Q. Wang, K. Feng, X. Li, and S. Jin, "PrecoderNet: Hybrid beamforming for millimeter wave systems with deep reinforcement learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1677–1681, Oct. 2020.

[39] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, Jul. 2019.

[40] A. Morsali, A. Haghighat, and B. Champagne, "Deep learning-based hybrid analog-digital signal processing in mmWave massive-MIMO systems," *IEEE Access*, vol. 10, pp. 72348–72362, 2022.

[41] L.-H. Shen, T.-W. Chang, K.-T. Feng, and P.-T. Huang, "Design and implementation for deep learning based adjustable beamforming training for millimeter wave communication systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2413–2427, Mar. 2021.

[42] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. London, U.K.: Pearson Education, 2015.

[43] A. Molisch, V. Ratnam, S. Han, Z. Li, S. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[44] *NR; Physical Layer Procedures for Data*, 3rd Generation Partnership Project (3GPP), document TS 38.214, version 15.4.0, Jan. 2019, [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3216

[45] S. Priebe, M. Jacob, and T. Kuerner, "AoA, AoD and ToA characteristics of scattered multipath clusters for THz indoor channel modeling," in *Proc. 17th Eur. Wireless Sustainable Wireless Technol.* Vienna, Austria: VDE, 2011, pp. 1–9.

[46] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 347–360, Mar. 2000.

[47] C. Lin and G. Y. Li, "Indoor terahertz communications: How many antenna arrays are needed?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3097–3107, Jun. 2015.

[48] H. Kasai, "Fast optimization algorithm on complex oblique manifold for hybrid precoding in millimeter wave MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 1266–1270.

[49] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[50] T. Peken, S. Adiga, R. Tandon, and T. Bose, "Deep learning for SVD and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6621–6642, Oct. 2020.

[51] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: A self-gated activation function," 2017, *arXiv:1710.05941*.

[52] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[53] A. Y. Ng, "Feature selection, L_1 vs. L_2 regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*. New York, NY, USA: Association for Computing Machinery, 2004, p. 78, doi: 10.1145/1015330.1015435.

[54] T. Carneiro, R. V. M. Da NóBrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance analysis of Google colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018.

[55] M. Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[56] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.

[57] K. Gokcesu and H. Gokcesu, "Generalized Huber loss for robust learning and its efficient minimization for a robust statistics," 2021, *arXiv:2108.12627*.

[58] H.-J. Song, J.-Y. Kim, K. Ajito, N. Kukutsu, and M. Yaita, "50-Gb/s direct conversion QPSK modulator and demodulator MMICs for terahertz communications at 300 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 3, pp. 600–609, Mar. 2014.

[59] B. Razavi, "A 300-GHz fundamental oscillator in 65-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 46, no. 4, pp. 894–903, Apr. 2011.
[60] E. Olieman, A. J. Annema, and B. Nauta, "An interleaved full Nyquist high-speed DAC technique," *IEEE J. Solid-State Circuits*, vol. 50, no. 3, pp. 704–713, Jan. 2015.

RAFID UMAYER MURSHED was born in Sylhet, Bangladesh. He received the B.Sc. degree in electrical and electronics engineering (EEE) from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, in May 2022. Currently, he is a Research Assistant with the BUET-Japan Institute of Disaster Prevention and Urban Safety (JIDPUS), BUET, where he is working on developing a disaster-resilient delay-sensitive telecommunication network for earthquake early warning broadcast. His research interests include the applications of deep learning and reinforcement learning in wireless communication technologies, such as UM-MIMO, beamforming, resource allocation, reconfigurable intelligent surfaces, as well as in signal processing, and the Internet of Things.

ZULQARNAIN BIN ASHRAF was born in Dhaka, Bangladesh, in October 1996. He received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, in May 2022. His research interests include 5G and 6G wireless technologies, such as beamforming, terahertz communications, massive MIMO, and deep learning.

ABU HORAIRA HRIDHON received the B.Sc. degree in electrical and electronic engineering (EEE) from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, in 2022. His research interests include wireless communication, computer vision, signal processing, and machine learning.

KUMUDU MUNASINGHE (Member, IEEE) received the Ph.D. degree in telecommunications engineering from The University of Sydney. He is currently an Associate Professor of network engineering and the Leader of the IoT Research Group, Human Centred Research Centre, University of Canberra. He has more than 100 refereed publications with more than 1250 citations (H-index: 21) in highly prestigious journals, conference proceedings, and two books to his credit. He has secured more than U.S. $1.7 million in competitive research funding by winning grants from the Australian Research Council (ARC), the Commonwealth and State Governments, and the Department of Defense, and Industry. His research interests include next generation mobile and wireless networks, the Internet of Things, green communications, smart grid communications, and cyber-physical-security. He has also won the highly prestigious ARC Australian Postdoctoral Fellowship, served as the co-chair for many international conferences, and served as an editorial board member for a number of journals. His research has been highly commended through many research awards, including two VC's research awards and three IEEE best paper awards. He is currently a Charted Professional Engineer, an Engineering Executive, and a Companion (fellow status) of Engineers Australia.

ABBAS JAMALIPOUR (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Nagoya University. He is currently a Professor of ubiquitous mobile networking with The University of Sydney. He has authored nine technical books, 11 book chapters, more than 550 technical articles, and five patents, all in the area of wireless communications. He is a fellow of the Institute of Electrical, Information, and Communication Engineers (IEICE) and the Institution of Engineers Australia, an ACM Professional Member, and an IEEE Distinguished Speaker. He is also the President of the IEEE Vehicular Technology Society. He was a recipient of a number of prestigious awards, such as the 2019 IEEE ComSoc Distinguished Technical Achievement Award in Green Communications, the 2016 IEEE ComSoc Distinguished Technical Achievement Award in Communications Switching and Routing, the 2010 IEEE ComSoc Harold Sobol Award, the 2006 IEEE ComSoc Best Tutorial Paper Award, and 15 best paper awards. He has been the General Chair or the Technical Program Chair for a number of conferences, including IEEE ICC, GLOBECOM, WCNC, and PIMRC. Previously, he was the Executive Vice-President and the Editor-in-Chief of the *VTS Mobile World* and has been an Elected Member of the Board of Governors of the IEEE Vehicular Technology Society, since 2014. He was the Editor-in-Chief of IEEE Wireless Communications, the vice resident of various conferences, and a member of the Board of Governors of the IEEE Communications Society. He serves as an Editor of IEEE Access, IEEE Transactions on Vehicular Technology, and several other journals.

MD. FARHAD HOSSAIN (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and electronic engineering (EEE) from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2003 and 2005, respectively, and the Ph.D. degree from the School of Electrical and Information Engineering, The University of Sydney, Australia, in 2014. Currently, he is a Professor with the Department of Electrical and Electronic Engineering, BUET. He also has vast experience as an electrical and electronic engineering consultant. He has published more than 85 refereed articles in highly prestigious journals and conference proceedings. His research interests include 6G cellular mobile networks, smart grid communications, the IoT and sensor networks, underwater communications, deep learning in wireless networking, network architectures, and protocols designs. He was a recipient of best paper awards at three international conferences and the Student Travel Grant at the IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA, in 2012. He has been serving as an editor, the track chair, the TPC member, and a reviewer for many international journals and conferences.