IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 June 2022, accepted 29 June 2022, date of publication 5 July 2022, date of current version 14 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188644

RESEARCH ARTICLE

Deep Learning-Based Hybrid Analog-Digital Signal Processing in mmWave Massive-MIMO Systems

ALIREZA MORSALI[1], (Member, IEEE), AFSHIN HAGHIGHAT[2], (Senior Member, IEEE), AND BENOIT CHAMPAGNE[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0G4, Canada
[2]InterDigital Canada, Montreal, QC H3A 3G4, Canada

Corresponding author: Alireza Morsali (alireza.morsali@mail.mcgill.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), and in part by InterDigital Canada.

ABSTRACT Hybrid analog-digital signal processing (HSP) is an enabling technology to harvest the potential of millimeter-wave (mmWave) massive-MIMO communications. In this paper, we present a general deep learning (DL) framework for efficient design and implementation of HSP-based massive-MIMO systems. Exploiting the fact that any complex matrix can be written as a scaled sum of two matrices with unit-modulus entries, a novel *analog* deep neural network (ADNN) structure is first developed which can be implemented with common radio frequency (RF) components. This structure is then embedded into an extended hybrid analog-digital deep neural network (HDNN) architecture which facilitates the implementation of mmWave massive-MIMO systems while improving their performance. In particular, the proposed HDNN architecture enables HSP-based massive-MIMO transceivers to approximate any desired transmitter and receiver mapping with arbitrary precision. To demonstrate the capabilities of the proposed DL framework, we present a new HDNN-based beamformer design that can achieve the same performance as fully-digital beamforming, with reduced number of RF chains. Finally, simulation results are presented confirming the advantages of the proposed HDNN design over existing hybrid beamforming schemes.

INDEX TERMS Hybrid beamforming, deep learning, deep neural networks, hybrid analog-digital beamforming, massive-MIMO, mmWave, beyond 5G (B5G), 6G.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology has revolutionized modern wireless communications by unveiling its potential to increase transmission capacity through the deployment of multiple antennas at the transmitter and receiver sides of a communication link [1]. In recent years, asymptotic analysis has revealed that *massive*-MIMO systems employing large scale antenna arrays, exhibit a linear increase in capacity with the minimum number of antennas employed at either the transmitter or receiver, even in sparse scattering environments [2], [3]. This property is of crucial importance for extending the applications of mmWave communications, which until recently had been only considered for short-range indoor and fixed outdoor scenarios, and enabling multi-Gbps data rates in future wireless networks [4]–[6].

Indeed, mmWave signals experience severe path loss (due to atmospheric absorption) and high penetration loss compared with microwave signals, which has hindered their use in wireless cellular and local area networks. However, recent advances in mmWave hardware, combined with the capabilities of massive-MIMO and the availability of spectrum above 6 GHz have revived mmWave communications.

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

Especially, the highly selective beam steering capabilities provided by large-scale antenna arrays and sophisticated beamforming[1] algorithms can mitigate the intrinsic limitations of mmWave channels [7].

In the conventional fully-digital (FD) implementation of MIMO systems, each antenna element is connected to a dedicated radio frequency (RF) chain. While this approach is suitable for commonly used small scale MIMO systems, it is not suitable for mmWave massive-MIMO systems equipped with large number of antenna elements due to the high production costs and power consumption of the associated RF circuitry. Therefore, although mmWave massive-MIMO is a prime technology for future generations of wireless networks (e.g., beyond 5G (B5G) and 6G), the implementation of such systems still faces many technical challenges, and to date remains a topic of ongoing research [8]–[11].

Hybrid analog-digital signal processing (HSP) is an ingenious and effective approach to facilitate the implementation of mmWave massive-MIMO transceivers [12]–[26]. In an HSP transmitter, which includes hybrid beamforming (HBF) as a special case, a low-dimensional baseband signal (e.g., precoder output) is converted to RF and then mapped into a higher-dimensional signal for transmission by the antennas, where the mapping is achieved by an analog processing network comprised of basic RF components such as phase-shifters, combiners and dividers; in an HSP receiver, the dual operations are performed in reverse order. Consequently, the HSP transmitter/receiver structure requires a smaller number of RF chains for conversion between the digital baseband and analog RF domains, compared to its FD counterpart. In this work, our aim is to develop and validate a novel deep learning (DL) framework for the efficient design and implementation of HSP-based mmWave massive-MIMO systems.

A. RELATED WORKS

One of the most prominent techniques for designing HSP systems consists in minimizing the Euclidean distance between the desired FD processor and its hybrid counterpart, which is the objective function used for HBF design in [15]–[22]. Particularly, in [15], [16], compressed sensing techniques are employed to exploit sparse characteristics of the mmWave channels while in [17], [18], a manifold optimization algorithm and a simultaneous matrix diagonalization technique are introduced, respectively, to solve the design problem. Channel sparsity is also considered in [14], [27] where iterative orthogonalization algorithms are proposed for designing spectrally efficient HBF transceivers. Gram–Schmidt orthogonalization is used in [21] to design a robust hybrid combiner with low complexity for an uplink multi-user scenario. The mean square error (MSE) is considered as the performance metric in [10], where an alternating minimization technique is used to design the HBF matrices. Considering that closed-form expressions with fixed amount of calculation are often more attractive in applications, non-iterative design algorithms exploiting this type of solutions are proposed in [28], [29]. The authors in [30] investigate the design and implementation (using CMOS process technology) of a low-complexity HBF based on orthogonal beamforming codebooks and a local search scheme.

Recently, in light of the huge success of machine learning and particularly deep learning (DL) in various fields of engineering, deep neural networks (DNNs) have attracted considerable attention among researchers for designing communication systems [31]–[40]. In [34], a DL model is proposed for predicting the beamforming vectors at several distributed and coordinated base stations (BSs) by using received pilot signals. In particular, the signatures of the signals jointly received at the BSs with omni/quasi-omni directional beam patterns are used to learn and predict the RF beamforming vectors. Another DL-based HBF approach for mmWave massive-MIMO is presented in [35], where an autoencoder is used to design the analog and digital precoders based on geometric mean decomposition. In [36], the problem of maximizing spectral efficiency with hardware limitation and imperfect channel state information (CSI) is tackled by training a DNN to learn the optimum beamformers. Imperfect CSI is also considered in [37], where multi-user DNN-based HBF design using codebooks is developed. Moreover, in order to apply this scheme to situations where the CSI is unknown, the concept of a reference RF beamformer is introduced. DL is used in [41] to design a joint hybrid processing framework that allows end-to-end optimization using back propagation. In [40], an unsupervised deep learning method is proposed for designing the synchronization signal in initial access as well as a codebook for the analog precoding.

Convolutional neural networks (CNNs) have also been investigated for HSP system design under various conditions [42]–[46]. A CNN framework for the joint design of precoder and combiner is proposed in [42] where the network accepts channel matrices as input and produces analog and baseband beamformers as output. In [43], three CNN architectures with different complexities are proposed to obtain approximations to the singular value decomposition (SVD), which are used in turn in the design of HBFs. In [44], a simplified hybrid precoding scheme is developed by considering the equivalent channel from the transmitter RF chains to the receiver RF chains. Based on this precoding approach, a novel CNN-based combiner architecture is proposed which can be trained to optimize the spectral efficiency under hardware limitations and imperfect CSI. Full-duplex mmWave systems are considered in [46] where a learning schemes for designing HBF via extreme learning machine and convolutional neural networks.

B. MOTIVATIONS AND CONTRIBUTIONS

As mentioned above, the premise of HSP is achieving the performance of FD systems with limited number of

[1]In practice, beamforming can be employed at both the transmitter and the receiver ends of a wireless link, where it is referred to as precoding and combining, respectively.

RF chains. However, this is not a trivial task mainly because of the constant modulus constraint on the analog beamformer entries, which causes the ensuing optimization problems to be non-convex. Moreover, since in precursory HBF works [15]–[22], the focus was placed on designing linear transformation matrices for the analog and digital beamformers, the recent DNN-based designs have followed the same structural guideline [34]–[37], [42]–[44] which imposes fundamental limits on system performance. In particular, the minimum number of required RF chains must be greater or equal to the number of transmitted/received symbols or the rank of the FD matrix [13], [14]. However, some of these limitations could potentially be mitigated by considering non-linear transformations.

The so-called *expressive power* of DNNs makes it possible to approximate continuous functions with arbitrary precision [47]. In the context of HSP, this property can thus be exploited to obtain a non-linear transformation that ultimately requires a smaller number of RF chains compared to the linear case. This motivates us to propose a general DL framework for efficient design and implementation of HSP-based massive-MIMO systems. To this end, we first develop a novel *analog* deep neural network (ADNN) structure using conventional RF components as found in existing HBF systems, i.e., [15]–[17]. The ADNN is then embedded in an extended hybrid analog-digital deep neural network (HDNN) architecture with the goal of facilitating the implementation of mmWave massive-MIMO transceiver systems and improving their performance. The proposed HDNN architecture allows for accurate approximation of desired transmitter and receiver mappings in HSP-based massive-MIMO systems. While the proposed framework is quite general and can be used in different applications, our main focus in this work lies on the study of uplink and downlink beamforming in massive-MIMO transmissions, for which we develop and investigate new HSP-based transceiver designs. Specifically, our main contributions can be summarized as follows:

- Considering the unit-modulus constraint of analog beamformers and inspired by the seminal works [12]–[14], we present a technique for decomposing any given matrix with arbitrary complex entries into a scaled sum of two matrices with unit-modulus entries.
- This decomposition is exploited to conceive an ADNN structure comprised of common basic RF components, i.e.: phase-shifters, rectifiers, combiners/dividers and switches. This structure enables the efficient implementation of DNNs *directly in the analog domain* which to the best of our knowledge, has not been previously addressed in the literature.
- To reduce the number of required RF chains in HSP systems, a deep learning framework for HSP design is then presented by introducing a novel HDNN architecture, comprised of a baseband digital DNN and an ADNN interconnected by means of RF chains.
- To demonstrate the capabilities of the proposed framework, we present uplink and downlink beamformer designs for mmWave massive-MIMO systems which achieve FD beamforming performance with limited number of RF chains. In particular, it is shown through these designs that the proposed HDNN framework is capable of reducing this number below the limits of conventional and existing DNN-based HBF designs.
- Extensive simulation results are presented to demonstrate the advantages of the proposed HDNN-based beamfomer designs, which can achieve the same performance as their FD counterparts with reduced number of RF chains.

The paper is organized as follows. Section II, introduces the extended HSP-based massive-MIMO system model under consideration in this work. Section III develops the novel ADNN structure which is based on a representation theorem for complex matrices with unit-modulus entries. Our proposed HDNN framework for HSP system design is presented in Section IV while a supervised beamformer design based on this structure is developed in Section V. Simulation results are presented in Section VI and finally, Section VII concludes the paper.

Notations: We use bold capital and lowercase letters to represent matrices and vectors, respectively. Superscripts $(.)^H$ and $(.)^T$ indicate Hermitian and transpose operations, respectively, while $\|\cdot\|_F$ denotes the Frobenious norm of a matrix. $\mathbf{I}_n$ and $\mathbf{1}_n$ denote an identity matrix of size $n \times n$ and a vector of all ones with size $n \times 1$, receptively. The element on the p^{th} row and q^{th} column of matrix $\mathbf{A}$ is denoted by $A(p, q)$. A complex $n \times 1$ Gaussian random vector $\mathbf{x}$ with mean vector $\mathbf{m} = \mathbb{E}\{\mathbf{x}\}$ and covariance matrix $\mathbf{R} = \mathbb{E}\{\mathbf{x}^H\mathbf{x}\}$ is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{R})$. The greatest integer less than or equal to x is denoted by $\lfloor x \rfloor$. The fields of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The magnitude and phase of a complex number $z \in \mathbb{C}$ are denoted by $|z|$ and $\angle z$, respectively. The function composition operation is denoted by $\circ$, i.e., $(f \circ g)(x) = f(g(x))$.

II. SYSTEM MODEL

In this section, we first review the conventional linear HBF massive-MIMO structure and then introduce an extended system formulation to support the development of our proposed DL framework for HSP.

A. CONVENTIONAL HBF STRUCTURE

Fig. 1 illustrates the typical downlink HBF configuration considered for the efficient realization of massive-MIMO BS transceivers in forthcoming wireless networks [15]–[22]. The transceiver system consists of a baseband processor with K-dimensional input stream, a number N_{RF} of RF chains, and an analog processor connected to N antennas. Unless otherwise specified, we consider a multi-user scenario wherein the BS transceiver provides services to $L \leq K$ user equipment (UE), each being allocated a fixed subset of the available data streams. For simplicity, we assume that the lth UE is equipped with M_l antennas and the same number of RF chains so that it is capable of FD processing, as our main focus lies

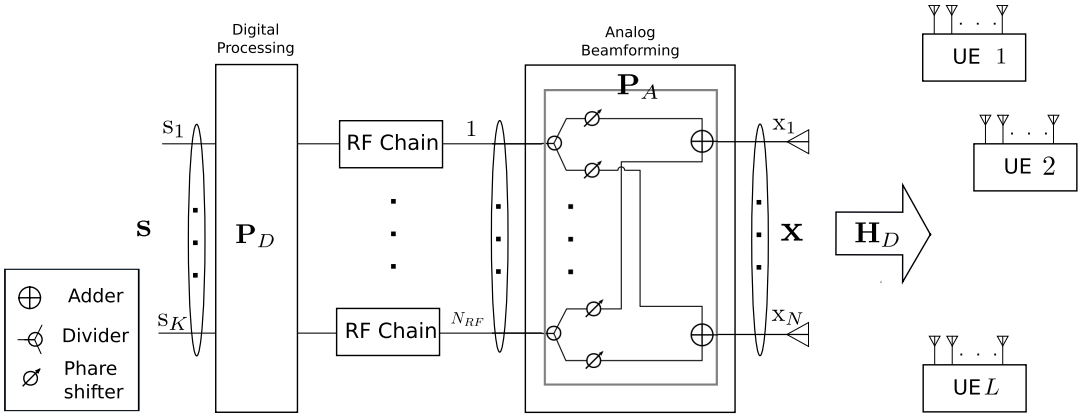

FIGURE 1. Conventional HBF structure for massive-MIMO transceiver serving L UEs. The transceiver is equipped with N_{RF} RF chains and N antennas; in downlink operation, $\mathbf{P}_D$ and $\mathbf{P}_A$ are the digital and analog precoders, respectively.

on the design and realization of HSP-based massive-MIMO transceivers at the BS. In the sequel, we let $M = \sum_{l=1}^{L} M_l$ denote the total number of antenna elements employed by the UEs. While Fig. 1 focuses on downlink transmission for simplicity, we herein consider both downlink and uplink connections, where the BS antennas are used for transmitting to and receiving from the UEs, respectively. Below, we further expand the corresponding signal models. For convenience, the key parameters and notations are summarized in Table 1.

TABLE 1. Parameters and notations of the system model.

Symbol	Description
$\mathbf{s} \in \mathbb{C}^K$	Input symbol vector
$\mathbf{P}_\mathrm{D} \in \mathbb{C}^{N_{RF} \times K}$	Digital precoders
$\mathbf{P}_\mathrm{A} \in \mathbb{U}^{N \times N_{RF}}$	Analog precoders
$\mathbf{x} \in \mathbb{C}^N$, $\mathbf{y} \in \mathbb{C}^M$	Downlink transmitted and received signal vectors
$\mathbf{H}_\mathrm{D} \in \mathbb{C}^{M \times N}$	Downlink channel matrix
$\mathbf{H}_\mathrm{U} \in \mathbb{C}^{N \times M}$	Uplink channel matix
$\mathbf{x} \in \mathbb{C}^M$, $\mathbf{y} \in \mathbb{C}^N$	Uplnik transmitted and received signal vectors
$\mathbf{C}_\mathrm{D} \in \mathbb{C}^{K \times N_{RF}}$	Digital combiner
$\mathbf{C}_\mathrm{A} \in \mathbb{U}^{N_{RF} \times N}$	Analog combiner
$\mathbf{n}$	White Gaussian noise vector
ρ	Average transmit power
K	Number of transmitted symbols
N_{RF}	Number of RF chains
N	Number of antennas at massive-MIMO BS
L	Number UEs
M_l	Number of antennas at lth UE
M	Total number of UE antennas $M = \sum_{l=1}^{L} M_l$

1) DOWNLINK

The signal vector transmitted by the BS over one symbol duration T_s, denoted as $\mathbf{x} \in \mathbb{C}^N$, is expressed as

$$\mathbf{x} = \sqrt{\rho}\,\mathbf{P}_\mathrm{A}\mathbf{P}_\mathrm{D}\mathbf{s}, \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T \in \mathcal{A}^K$ is the input symbol vector, with zero-mean uncorrelated random information symbols s_k taken from a discrete constellation $\mathcal{A} \subset \mathbb{C}$ (e.g., M-QAM or M-PSK), while matrices $\mathbf{P}_\mathrm{D} \in \mathbb{C}^{N_{RF} \times K}$ and $\mathbf{P}_\mathrm{A} \in \mathbb{U}^{N \times N_{RF}}$ represent the digital and analog precoders, respectively, where $\mathbb{U} = \{z \in \mathbb{C} \colon |z| = 1\}$. For normalization purposes, it is assumed that $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_K$ and $\|\mathbf{P}_\mathrm{A}\mathbf{P}_\mathrm{D}\|_F^2 = 1$, so that ρ is the average transmit power.

The received signals at the UEs are represented by a concatenated signal vector $\mathbf{y} \in \mathbb{C}^M$ which is written as

$$\mathbf{y} = \mathbf{H}_\mathrm{D}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{H}_\mathrm{D} \in \mathbb{C}^{M \times N}$ is a zero-mean random channel matrix representing flat fading transmission between the BS and the UE antennas, normalized such that $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = NM$ without loss of generality, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ is an additive noise vector of size M. Each UE applies optimal FD linear processing on its received signal prior to the decoding stage. The linear processing performed by the multiple UEs is expressed as

$$\hat{\mathbf{s}} = \mathbf{C}\mathbf{y}, \tag{3}$$

where $\mathbf{C} \in \mathbb{C}^{K \times M}$ is a block diagonal combiner matrix.

2) UPLINK

To simplify the presentation, we use similar notations as in the donwlink case, with trivial modifications in vector dimensions as needed. Specifically, the received signal vector at the BS, denoted by $\mathbf{y} \in \mathbb{C}^N$, is written as

$$\mathbf{y} = \mathbf{H}_\mathrm{U}\mathbf{x} + \mathbf{n}, \tag{4}$$

where $\mathbf{x} \in \mathbb{C}^M$ denotes the vector containing the concatenated transmitted signals from all UEs, $\mathbf{H}_\mathrm{U} \in \mathbb{C}^{N \times M}$ is the zero-mean transmission matrix between the UEs and the BS antennas, normalized such that $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = NM$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$ is an additive noise term of size N. In the conventional uplink HBF structure, a constrained form of linear processing is applied to the received signal, which is expressed as

$$\hat{\mathbf{s}} = \mathbf{C}_\mathrm{D}\mathbf{C}_\mathrm{A}\mathbf{y}, \tag{5}$$

where matrices $\mathbf{C}_\mathrm{D} \in \mathbb{C}^{K \times N_{RF}}$ and $\mathbf{C}_\mathrm{A} \in \mathbb{U}^{N_{RF} \times N}$ represent the digital and analog combiners, respectively. The final

combiner output $\hat{\mathbf{s}}$ is then passed to a decoding stage, whose details fall outside the scope of this work.

B. GENERALIZED HSP STRUCTURE

Here, we introduce an extended formulation for HSP[2] that generalizes the linear analog and digital transformations presented in Subsection II-A. This formulation provides an adequate basis for the subsequent development of our proposed DL-based HSP structures for massive MIMO transceivers.

1) DOWNLINK

In a more general setting, additional parameters are provided along with the symbol vector $\mathbf{s}$ as input to the digital component of the HSP-based massive MIMO transmitter. Depending on the chosen transmission scheme, this may include for instance the channel matrix $\mathbf{H}$ or the average transmit power ρ. The signal vector at the output of the baseband processor is therefore expressed as

$$\mathbf{x}_{\mathrm{BB}} = \mathscr{F}_T(\mathbf{z}_T) \in \mathbb{C}^{N_{RF}}, \tag{6}$$

where $\mathbf{z}_T$ is the input of appropriate size that includes the additional parameters, i.e., $\mathbf{z}_T = (\mathbf{s}, \mathbf{H}, \ldots, \rho)$, and $\mathscr{F}_T$ is a generic mapping from $\mathbf{z}_T$ to $\mathbf{x}_{\mathrm{BB}} \in \mathbb{C}^{N_{RF}}$. Then, N_{RF} parallel RF chains convert the baseband signal vector $\mathbf{x}_{\mathrm{BB}}$ into a bandpass modulated RF signal vector $\mathbf{x}_{\mathrm{RF}}$. Next, the latter vector is used along with additional parameters as input to the analog signal processing network whose output is the transmit signal vector, which is expressed as

$$\mathbf{x} = \sqrt{\rho}\, \mathscr{G}_T(\mathbf{z}_T^{RF}) \in \mathbb{C}^N, \tag{7}$$

where $\mathbf{z}_T^{RF}$ is the input of appropriate size, i.e., $\mathbf{z}_T^{RF} = (\mathbf{x}_{\mathrm{RF}}, \mathbf{H}, \ldots, \rho)$ and $\mathscr{G}_T$ is the corresponding mapping from $\mathbf{z}_{RF}$ to $\mathbf{x} \in \mathbb{C}^N$. The received signal at the UE is given by (2) and, in the case of FD linear combining, the demodulated symbols are obtained as in (3).

We emphasize that in the above formulation, the mappings $\mathscr{F}_T$ and $\mathscr{G}_T$ are no longer limited to the class of linear transformations and could exhibit non-linear features, which is necessary for the consideration of DNNs.

2) UPLINK

The received signal from the large scale antenna array at the BS, which is given by (4), is fed along with additional required parameters to the analog signal processing network, whose output is represented by

$$\mathbf{x}_{\mathrm{RF}} = \mathscr{G}_R(\mathbf{z}_R^{RF}) \in \mathbb{C}^{N_{RF}}, \tag{8}$$

where $\mathbf{z}_R^{RF} = (\mathbf{y}, \mathbf{H}, \ldots, \rho)$ is the input of appropriate size and $\mathscr{G}_R$ is the corresponding mapping. Then, N_{RF} parallel RF chains convert the RF signal vector $\mathbf{x}_{\mathrm{RF}}$ into a digital signal vector $\mathbf{x}_{\mathrm{BB}}$ which is subsequently passed through a baseband processor, as represented by

$$\hat{\mathbf{s}} = \mathscr{F}_R(\mathbf{z}_R) \in \mathbb{C}^K, \tag{9}$$

where $\mathbf{z}_R = (\mathbf{x}_{\mathrm{BB}}, \mathbf{H}, \ldots, \rho)$ is the input and $\mathscr{F}_R$ is the corresponding digital mapping. Here again, we envisage a general configuration where the mappings $\mathscr{G}_R$ and $\mathscr{F}_T$ could be non-linear.

III. ANALOG (RF) DEEP NEURAL NETWORKS

In this section, we present a novel ADNN structure for efficient implementation of DNNs directly in the analog domain. We start by introducing an essential notations for the compact description of DNN architectures. We then present the linear and non-linear RF modules which will be needed to carry out the necessary operations in DNNs. Finally, we present the proposed ADNN structure.

A. DEEP NEURAL NETWORKS

We consider complex-valued DNNs since in communication systems, complex numbers are used for the representation of the transmitted/received bandpass signals.[3] The mathematical expression of a complex-valued multi-layer perceptron (MLP), or artificial neural network, with one hidden layer, d neurons and a linear layer at its output, is given by

$$\Gamma(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{W}\, \psi_d(\mathscr{A}(\mathbf{x})), \tag{10}$$

where $\mathbf{x} \in \mathbb{C}^n$ is the input vector, $\mathscr{A} : \mathbb{C}^n \to \mathbb{C}^d$ is an affine transformation, ψ_d is a non-linear activation function and $\mathbf{W} \in \mathbb{C}^{m \times d}$ is the weight matrix of the output layer. Specifically, $\mathscr{A}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ where $\mathbf{A} \in \mathbb{C}^{d \times n}$, $\mathbf{b} \in \mathbb{C}^d$ are the weight matrix and bias vector of the hidden layer, respectively. Moreover, for a given scalar activation function $\psi : \mathbb{C} \to \mathbb{C}$, the element-wise activation function $\psi_d : \mathbb{C}^d \to \mathbb{C}^d$ is defined as $\psi_d(\boldsymbol{\xi}) = [\psi(\xi_1), \psi(\xi_2), \ldots, \psi(\xi_d)]^T$. where $\boldsymbol{\xi} \in \mathbb{C}^d$. Consequently, the set of parameters characterizing the MLP can be written as $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{A}, \mathbf{b}\}$. A common practice for simplifying the implementation of DNN algorithms is to remove the bias term by adding an additional unit element to $\mathbf{x}$ and a column to $\mathbf{A}$. Consequently, by defining $\mathbf{A}_b = [\mathbf{A}, \mathbf{b}]$ and $\mathbf{x}_b = [\mathbf{x},\ 1]^T$, we can rewrite (10) as

$$\Gamma(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{W}\, \psi_d(\mathbf{A}_b \mathbf{x}_b). \tag{11}$$

To simplify the presentation, we focus on feed-forward DNN constructed from MLP with multiple hidden layers, but extensions to other types of networks are possible. Let us consider a DNN comprised of $L \in \mathbb{N}$ hidden layers indexed by $l \in \{1, 2, \ldots, L\}$, $n_l \in \mathbb{N}$ neurons at the l-th hidden layer, and one output layer indexed by $l = L + 1$, with input vector $\mathbf{x} \in \mathbb{C}^{N_0}$ and output vector $\mathbf{y} \in \mathbb{C}^{n_{L+1}}$. The mathematical representation of such DNN is given by

$$\mathbf{y} = \Omega(\mathbf{x}; \boldsymbol{\theta}) = (\mathscr{A}_{L+1} \circ \psi_{n_L} \circ \mathscr{A}_L \circ \cdots \circ \psi_{n_1} \circ \mathscr{A}_1)(\mathbf{x}) \tag{12}$$

where $\mathscr{A}_l : \mathbb{C}^{n_{l-1}} \to \mathbb{C}^{n_l}$ is the affine transformation used at the l-th layer, i.e., $\mathscr{A}_l(\mathbf{x}) = \mathbf{A}_l\mathbf{x} + \mathbf{b}_l$, with $\mathbf{A}_l \in \mathbb{C}^{n_l \times n_{l-1}}$ and $\mathbf{b}_l \in \mathbb{C}^{n_l}$ representing the weights and

[2] Although we focus on HSP at the BS to simplify presentation, similar concepts can be applied to the UE transceivers.

[3] In a practical implementation of a transceiver system the required operations between complex signal values and parameters can be realized through manipulation of the in-phase (I) and quadrature (Q) components.

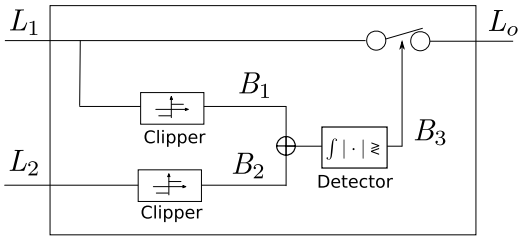

FIGURE 2. Block diagram of the analog ReLU activation function. The input is applied to L_1, the carrier is applied to L_2, and the output is obtained from L_o.

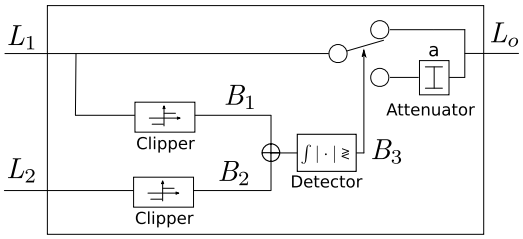

FIGURE 3. Block diagram of the analog PReLU activation function. The input is applied to L_1, the carrier is applied to L_2, and the output is obtained from L_o.

biases of the l-th layer. The ordered set of all parameters characterizing the DNN in (12) can be written as $\boldsymbol{\theta} = \{\mathbf{A}_1, \mathbf{b}_1, \mathbf{A}_2, \mathbf{b}_2, \ldots, \mathbf{A}_{L+1}, \mathbf{b}_{L+1}\}$.

B. LINEAR RF MODULES

As discussed above, two different types of operations are required for calculating the output of each DNN layer, i.e., for the l-th layer, a linear transformation $\mathscr{A}_l$ and an elementwise non-linear function ψ_{n_l}. Let us first focus on the implementation of the linear transformation, which is not an immediate task when using phase-shifters and power-combiners/dividers in RF domain. In fact, most of the literature on hybrid beamforming is dedicated to this issue, and especially the non-convexity of design problems involving transformation matrices with unit-modulus entries. Motivated by hybrid designs in [12]–[14], we present below a first proposition which paves the way for developing an analog structure consisting of phase-shifters and power-combiners/dividers, that can perform an arbitrary linear transformation.

Theorem 1: Any given complex matrix $\mathbf{A} \in \mathbb{C}^{\kappa \times n}$ can be written as a scaled sum of two matrices $\mathbf{R}^{(1)}, \mathbf{R}^{(2)} \in \mathbb{U}^{\kappa \times m}$, with unit-modulus entries, i.e.,

$$\mathbf{A} = c(\mathbf{R}^{(1)} + \mathbf{R}^{(2)}), \tag{13}$$

for some positive constant $c \geq \frac{1}{2}\|\mathrm{vec}(\mathbf{A})\|_\infty$.

Proof: It is sufficient to show that for all $p = 1, 2, \ldots, \kappa$ and $q = 1, 2, \ldots, n$, we have

$$A(p, q) = c\big(R^{(1)}(p, q) + R^{(2)}(p, q)\big). \tag{14}$$

Since $\mathbf{R}^{(1)}, \mathbf{R}^{(2)} \in \mathbb{U}^{N_R \times K}$, the above equation can be further written as

$$A(p, q) = c\big(e^{j\theta^{(1)}_{p,q}} + e^{j\theta^{(2)}_{p,q}}\big), \tag{15}$$

where $e^{j\theta^{(1)}_{p,q}}$ and $e^{j\theta^{(2)}_{p,q}}$ represent the entries on the p^{th} row and q^{th} column of $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$, respectively. Since $2c$ is greater than the absolute value of all the entries of $\mathbf{A}$, from Theorem 1 in [12], there exist non-unique $\theta^{(1)}_{p,q}$ and $\theta^{(2)}_{p,q}$ such that (15) holds which proves the theorem. ∎

Note that for a given matrix $\mathbf{A}$, the matrices $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$ such that $\mathbf{A} = c(\mathbf{R}^{(1)} + \mathbf{R}^{(2)})$ holds are non-unique; below, we present a simple way to construct $\mathbf{R}^{(1)}, \mathbf{R}^{(2)}$. By expressing the elements of $\mathbf{A}$ in polar form as $A(p, q) = |A_{p,q}| \exp(j\vartheta_{p,q})$, we first compute

$$c = \frac{1}{2} \max_{p,q} |A_{p,q}|. \tag{16}$$

We then calculate the entries of $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$ as

$$R^{(1)}(p, q) = e^{j\left(\vartheta_{p,q} + \cos^{-1}\left(\frac{|A_{p,q}|}{2c}\right)\right)} \tag{17a}$$

$$R^{(2)}(p, q) = e^{j\left(\vartheta_{p,q} - \cos^{-1}\left(\frac{|A_{p,q}|}{2c}\right)\right)}. \tag{17b}$$

By simple mathematical manipulations one can easily verify the validity of the presented solutions in (16) and (17) (see also [14]).

C. NONLINEAR RF MODULES

Next, we discuss possible structures for the realization of non-linear activation functions in the analog domain with common RF modules.

In [48], it is shown that an arbitrarily wide MLP with one hidden layer is a universal approximator due to the presence of the non-linear activation function. Different conditions for the activation function $\psi : \mathbb{R} \to \mathbb{R}$ have been presented under which the universality property of DNNs is satisfied [48]–[52]. In [48], the required conditions for ψ are given as being continuous, bounded and non-constant, while these conditions are simplified to non-polynomiality in [50]. Universality of DNNs is studied in [53], [54] for the rectified linear unit (ReLU) activation function [55], i.e.,

$$\mathrm{ReLU}(x) = x^+ = \max(0, x), \quad x \in \mathbb{R}. \tag{18}$$

Since the introduction of ReLU, several variations have been proposed among witch the leaky-ReLU and parametric ReLU (PReLU) have attracted considerable attention [56], [57]. In PReLU the vanishing part of the response is replaced by a nonzero linear section, i.e.,

$$\mathrm{PReLU}(x) = \begin{cases} ax & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \tag{19}$$

where the linear slope $a > 0$ can be learned along with other DNN parameters.

In complex-valued DNNs, the activation function must also be complex-valued. Based on [58], the definition of ReLU can be extended to complex numbers as

$$\mathrm{ReLU}(z) = \mathrm{ReLU}(\Re(z)) + j\mathrm{ReLU}(\Im(z)), \quad z \in \mathbb{C}, \tag{20}$$

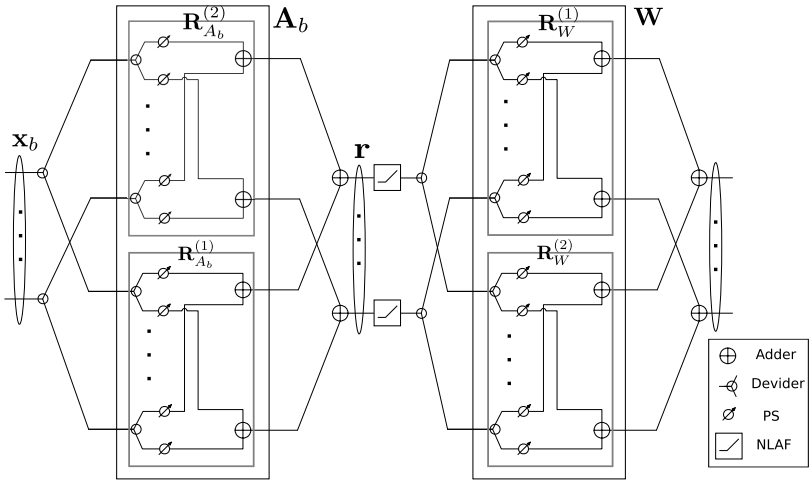

FIGURE 4. Analog MLP (NLAF = non-linear activation function, i.e. ReLU or PReLU).

with similar extension for PReLU. Since the baseband signals (i.e., the individual elements of the signal vector $\mathbf{x}_{\text{BB}}$) are modulated on a sinusoidal carrier, both the in-phase (I) and quadrature (Q) components of the RF signals oscillate between a negative and positive peak value, regardless of the sign of the baseband signals. Consequently, unlike the baseband case, ReLU cannot be directly implemented with a single rectifier (e.g., diode). Herein, we therefore present a novel analog ReLU module which can realize (20) in the RF domain using basic circuit components.

The conceptual block diagram of the proposed analog ReLU module is shown in Fig. 2. For complex-valued signals, this module is in effect applied to both the I and Q components. The modulated signal is fed to L_1 and the carrier is fed to L_2 as a reference. The output L_o is equal to the input signal L_1 when the baseband signal is positive and zero otherwise. Indeed, when the modulated signal L_1 and the carrier L_2 are in phase, corresponding to a positive value of the baseband component, B_1 and B_2 are equal and the energy of their sum is non-zero. Consequently, the output of the energy detector $B_3 = 1$ which can be used to activate (i.e., close) the switch. When the modulated signal L_1 and the carrier L_2 are out of phase (corresponding to a negative value of the baseband component), we have $B_2 = -B_1$. Consequently, the energy of their sum is zero and the output of energy detector $B_3 = 0$, which opens the switch. Similarly, an analog PReLU module can be designed as shown in Fig. 3 where the attenuator is a passive component which is designed based on the learned value of a in (19).

D. ANALOG DEEP NEURAL NETWORK (ADNN)

The RF modules conceived in the previous two subsections can now be used to achieve our main objective, which is to design an analog DNN structure, i.e. ADNN, such that

$$\begin{aligned} \mathscr{G}_T(\mathbf{x}_{\text{RF}}) &= \Omega_T(\mathbf{x}_{\text{RF}}; \boldsymbol{\theta}^A_T) \\ \mathscr{G}_R(\mathbf{y}) &= \Omega_R(\mathbf{y}; \boldsymbol{\theta}^A_R), \end{aligned} \tag{21}$$

where Ω_T and Ω_R represent DNNs with respective parameter sets $\boldsymbol{\theta}^A_T$ and $\boldsymbol{\theta}^A_R$. As discussed in Subsection III-A, DNNs are formed by concatenating several MLPs. Consequently, we first focus on the RF implementation of an MLP with single hidden layer as in (11). For developing the ADNN, we consider commonly used RF components in existing HSP, i.e., phase-shifters, and power-dividers (which also work as combiners), along with the simple analog ReLU or PReLuU modules introduced previously.

Theorem 2: A given MLP $\Gamma(\mathbf{x}; \boldsymbol{\theta})$ with one hidden layer and non-linear activation function ψ_d (either ReLU or PReLU) can be realized by the RF structure shown in Fig. 4.

Proof: As discussed earlier an MPL is defined by matrices $\mathbf{A}_b$ and $\mathbf{W}$ which are used to calculate $\Gamma(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{W}\,\psi_d(\mathbf{A}_b\mathbf{x}_b)$ where here, $\psi_d(\mathbf{x}) = \text{ReLU}(\mathbf{x})$ or $\text{PReLU}(\mathbf{x})$. Using Theorem 1, there exist matrices $\mathbf{R}^{(1)}_{\mathbf{A}_b}$, $\mathbf{R}^{(2)}_{\mathbf{A}_b}$, $\mathbf{R}^{(1)}_{\mathbf{W}}$ and $\mathbf{R}^{(2)}_{\mathbf{W}}$ with constant-modulus entries, and positive scalars c_A, c_W, such that

$$\mathbf{A}_b = c_A(\mathbf{R}^{(1)}_{\mathbf{A}_b} + \mathbf{R}^{(2)}_{\mathbf{A}_b}) \tag{22a}$$

$$\mathbf{W} = c_W(\mathbf{R}^{(1)}_{\mathbf{W}} + \mathbf{R}^{(2)}_{\mathbf{W}}). \tag{22b}$$

As can be seen from Fig. 4, $\mathbf{x}_b$ first goes through dividers which feed the analog subnetworks $\mathbf{R}^{(1)}_{\mathbf{A}_b}$, and $\mathbf{R}^{(2)}_{\mathbf{A}_b}$, whose output are finally combined to produce vector $\mathbf{r}$. Since we operate in the analog domain, the power of the signal is split between each branch, so that only $\frac{\mathbf{x}_b}{\sqrt{2}}$ is effectively applied on each branch; Similar, a factor of $\frac{1}{\sqrt{2}}$ must be included in the ouput of the combiner to account for properties of the corresponding S-matrix [59]. Consequently, the vector of input signals to the ReLUs can be written as

$$\mathbf{r} = \frac{1}{\sqrt{2}}(\mathbf{R}^{(1)}_{\mathbf{A}_b}\frac{\mathbf{x}_b}{\sqrt{2}} + \mathbf{R}^{(2)}_{\mathbf{A}_b}\frac{\mathbf{x}_b}{\sqrt{2}}) = \frac{1}{2}(\mathbf{R}^{(1)}_{\mathbf{A}_b} + \mathbf{R}^{(2)}_{\mathbf{A}_b})\mathbf{x}_b. \tag{23}$$

By substituting (22a) in (23), we arrive at

$$\mathbf{r} = \frac{1}{c_A}\mathbf{A}_b\mathbf{x}_b. \tag{24}$$

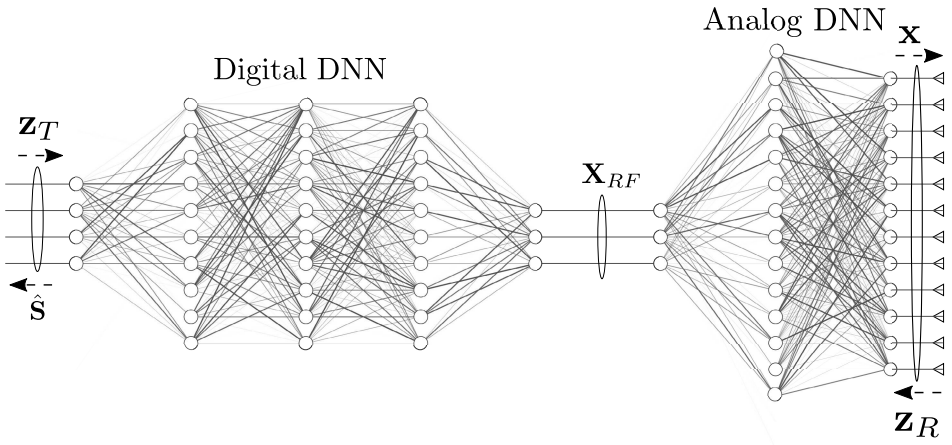

FIGURE 5. HDNN-based massive-MIMO transceiver.

After passing through the ReLU and performing similar steps as above for $\mathbf{W}$, the output of the analog MLP can be written as

$$\Gamma(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{c_A c_W}\mathbf{W}\,\psi_d(\mathbf{A}_b\mathbf{x}_b) = \mathbf{W}\,\psi_d(\mathbf{A}_b\frac{1}{c_A c_W}\mathbf{x}_b), \quad (25)$$

where the second equality is due to the positive homogeneity of ReLU i.e., $c\,\mathrm{ReLU}(\mathbf{x}) = \mathrm{ReLU}(c\mathbf{x})$ for any positive c ($c \in \mathbb{R}^+$). ■

Corollary 1: Any given DNN $\Omega(\mathbf{x};\boldsymbol{\theta})$ (12) can be realized in the RF domain by concatenating several analog MLPs, as presented in Theorem 2.

IV. DEEP LEARNING FRAMEWORK FOR HSP

In this section, by taking advantage of the proposed ADNN structure, we present a general DL framework for HSP which enables hybrid systems to exhibit similar performance as FD systems with limited number of RF chains. These results are in contrast with conventional hybrid systems where, due to non-convex constraints imposed on the RF beamfomer, degraded performance must be accepted for reducing the number of RF chains. Moreover, by introducing ADNN, our approach enables further improvements over existing DNN-based HSP designs, which face the same limitations as the conventional hybrid design, i.e., the minimum number of required RF chains is equal to the number of symbol streams. The proposed HDNN, however, is not bound to this constraint: indeed, as will be demonstrated for selected cases, the number of RF chains can be further reduced while maintaining the same performance as the FD systems.

A. PROPOSED DL FRAMEWORK FOR HSP

1) DOWNLINK

Let us consider the downlink signal model introduced in Section II-B for the generalized HSP structure. In the proposed HDNN framework, the non-linear digital processor in (6) indeed corresponds to a first DNN, whose output is the baseband signal vector $\mathbf{x}_{\mathrm{BB}}$ i.e.,

$$\mathbf{x}_{\mathrm{BB}} = \mathscr{F}_T(\mathbf{z}_T) = \Omega_T^D(\mathbf{z}_T;\boldsymbol{\theta}_T^D), \quad (26)$$

where $\boldsymbol{\theta}_T^D$ is the parameter set of this *digital* DNN. The baseband output of this DNN is then converted to the RF domain where it is represented by $\mathbf{x}_{\mathrm{RF}}$. Subsequently, the non-linear operator in (7) is realized by means of a second ADNN structure, whose output (up to a scale factor) is the desired transmit signal vector $\mathbf{x}$, i.e.,

$$\mathbf{x} = \sqrt{\rho}\,\mathscr{G}_T(\mathbf{z}_T^{RF}) = \Omega_T^A(\mathbf{z}_T^{RF};\boldsymbol{\theta}_T^A), \quad (27)$$

where $\boldsymbol{\theta}_T^A$ is the parameter set of the ADNN.

To simplify the training process, we can consider the ADNN as additional layers to the digital DNN, i.e.,

$$\mathbf{x} = \Omega_T^A\Big(\big(\Omega_T^D(\mathbf{z}_T;\boldsymbol{\theta}_T^D)\big);\boldsymbol{\theta}_T^A\Big). \quad (28)$$

In light of (12) and (10), we can write the above equation as

$$\mathbf{x} = \Omega_T^H(\mathbf{z}_T;\boldsymbol{\theta}_T), \quad (29)$$

where $\boldsymbol{\theta}_T = \{\boldsymbol{\theta}_T^A, \boldsymbol{\theta}_T^D\}$ is the composite parameter set of the resulting downlink HDNN.

2) UPLINK

In this case, the non-linear digital processor in (8) corresponds to a first ADNN, whose output is the RF signal vector $\mathbf{x}_{\mathrm{RF}}$, i.e.,

$$\mathbf{x}_{\mathrm{RF}} = \mathscr{G}_R(\mathbf{z}_R) = \Omega_R^A(\mathbf{z}_R;\boldsymbol{\theta}_R^A), \quad (30)$$

where $\boldsymbol{\theta}_R^A$ is the parameter set of the ADNN. The RF output of this ADNN is then converted to the baseband where it is represented by $\mathbf{x}_{\mathrm{BB}}$. Subsequently, the non-linear operator in (9) is realized by means of a digital DNN to produce the desired output vector $\hat{\mathbf{s}}$, i.e.,

$$\hat{\mathbf{s}} = \mathscr{F}_R(\mathbf{z}_R^{RF}) = \Omega_R^D(\mathbf{z}_R^{RF};\boldsymbol{\theta}_R^D), \quad (31)$$

where $\boldsymbol{\theta}_R^D = \{\boldsymbol{\theta}_R^A, \boldsymbol{\theta}_R^D\}$ is the complete parameter set of the DNN. Hence, proceeding as in the downlink case, we can write

$$\hat{\mathbf{s}} = \Omega_R^H(\mathbf{z}_R; \boldsymbol{\theta}_R) = \Omega_R^A\Big(\big(\Omega_R^D(\mathbf{z}_R; \boldsymbol{\theta}_R^D)\big); \boldsymbol{\theta}_R^A\Big), \tag{32}$$

where $\boldsymbol{\theta}_R$ is the parameter set of the resulting uplink HDNN.

The proposed HDNN architecture is depicted in Fig. 5 where the forward propagation is shown for both downlink and uplink scenarios. Here we considered regular DNNs, although different network structures and DL techniques such as CNNs, RNNs, etc. can also be employed in the proposed framework. Due to production cost and power consumption, it is desirable to use as few RF chains as possible. In conventional HSP systems the minimum number of RF chains is equal to the number of transmitted symbols because digital and particularly analog beamformers are linear transformations.

B. DL-BASED TRANSCEIVER DESIGN

In the following, we briefly discuss how the proposed framework can be used for HSP-based massive-MIMO transceiver design. For the sake of brevity, we only discuss the downlink problem formulation but the uplink problem can be also formulated in a similar way.

The HDNN can be trained to learn an arbitrary transmission scheme, as defined by the mappings $\mathscr{F}_T(\cdot)$ and $\mathscr{G}_T(\cdot)$ in (6) and (7). This can be done by training the HDNN using the following loss function,

$$L(\boldsymbol{\theta}_T) = \Big\| \Omega_T^H(\mathbf{z}_T; \boldsymbol{\theta}_T) - \mathscr{G}_T \circ \mathscr{F}_T(\mathbf{s}) \Big\|_2 . \tag{33}$$

This approach can be applied in principle to a variety of non-linear beamforming techniques such as vector perturbation [60], Tomlinson-Harashima precoding [61], [62], robust beamformer design under imperfect CSI [20], and space-time coding [63]. In the next section, we design and train HDNNs with $\mathscr{F}_T(\mathbf{s})$ selected as the eigen-mode beamforming.

When the HDNN is trained using (33), the output of the HDNN ideally appropriates the selected transmission scheme, i.e., $\mathscr{F}_T(\mathbf{s})$. However, taking a step further and exploiting the power of the machine learning paradigm, it is possible to let the transmitter learn the optimal transmission scheme without explicitly specifying its structure *a priori*. To be more specific, the loss function for training the HDNN can be a performance measure of the overall communication link, i.e., taking the transmitter and receiver into account, as well as other constraints and system requirements. For instance, in the case of a massive-MIMO BS with a single UE employing a fixed combiner matrix $\mathbf{C}$ known at the transmitter side, the HDNN can be trained using,

$$L(\boldsymbol{\theta}_T) = \Big\| \mathbf{C}\,\mathbf{H}_{\mathrm{D}}\,\Omega_T^H(\mathbf{z}_T; \boldsymbol{\theta}_T) - \mathbf{s} \Big\|_2 . \tag{34}$$

For a multi-user scenario, the signal-to-interference-plus-noise (SINR) can be used instead of the above symbol error.

C. DISCUSSION

In this subsection, we provide further discussion regarding the implementation aspects and computational complexity of the proposed design.

1) IMPLEMENTATION

Except for the nonlinear RF modules, the proposed scheme is characterized by an HADP architecture similar to [8]–[22]. Still, certain important considerations should be noted. In particular, since the linear modules in ADNN are comprised of phase-shifters, as the number of layers (and/or width of the layers) grow larger, the number of required phase-shifters also increases. However, recent advances in RF semiconductor and mmWave hardware will allow for deployment of a larger number of RF components [64], [65]. Direct implementation of DNNs at higher frequencies as well as in the optical domain has been recently attracting considerable attention [66]–[68]; these works pave the way towards design and fabrication of low-cost RF components for HDNNs.

2) COMPUTATIONAL COMPLEXITY

At the inference level, the proposed HSP framework design does not introduce any major additional complexity compared to the existing learning-based ones. In particular, the computational complexity is similar (if not lower) as the HDNN is directly used to generate the transmitted signal and decode the received signal, whereas in most existing works (e.g. [31]–[40]), DNN are used to calculate the coefficients of the beamformer matrices which in turn are applied in subsequent beamforming computations. At the training level, complexity of for the proposed HDNN depends on the design and constraints of the system. In general, since in the proposed framework the HDNN directly generates the beamformed signals, complexity of the training is comparable to the existing DL-based techniques. In the context of HBF, one common issue faced by existing DNN-based techniques is how to handle channel variations in the inference and training processes. One possible approach, suitable for static or slowly time-varying channels, consists in training the DNN for each channel coherence time interval. This simplified approach, employed in our simulation study, has the advantage of reducing the on-line training time. For rapidly varying channel environments however, the short coherence time makes it necessary to constantly retrain the network, which eventually become impractical. In such scenarios, more efficient techniques could be used by leveraging transfer learning to facilitate the training process [69]. Alternatively, one could enrich the training data with channel information and training the DNN off-line to allow generalization over an ensemble of channel realizations. How to optimally consider rapidly time-varying channels in the design and implementation of DL-based HBF remains an open avenue for research.

V. SUPERVISED LEARNING-BASED HBF WITH FD PERFORMANCE

In this section, to illustrate the potential advantages of the proposed DL framework for HSP, we apply it to design supervised learning-based hybrid beamformers that achieve the performance of FD beamforming with limited number of RF chains. We focus on the single user scenario although extension to multi-user is possible.

A. BEAMFORMING WITH $N_{RF} \geq K$

As discussed above, by using the proposed framework, it is possible to closely approximate any mapping form symbol vector $\mathbf{s}$ to the desired transmitted signal $\mathbf{x}$. Consequently, we can use supervised learning to train the DNN such that the hybrid system generates the output of the FD beamforming with limited number of RF chains.

In the downlink, the optimal eigen-mode precoding is obtained by solving the following problem:

$$\max_{\mathbf{P}} \quad \log_2 \det(\mathbf{I}_N + \mathbf{H}_\mathrm{D}\mathbf{P}\mathbf{P}^H\mathbf{H}_\mathrm{D}^H), \tag{35a}$$

$$\text{s.t. } \mathrm{Tr}(\mathbf{P}\mathbf{P}^H) \leq 1. \tag{35b}$$

The solution to the above problem is given by $\mathbf{P} = \mathbf{V}_D\Upsilon_D$, where the diagonal weight matrix Υ_D is calculated via water filling [6] and $\mathbf{V}_D$ is a unitary matrix obtained from the singular value decomposition of the channel matrix, represented in standard form as, $\mathbf{H}_\mathrm{D} = \mathbf{U_D}\Sigma_\mathbf{D}{\mathbf{V}_D}^H$. Thus, the desired output of the HDNN for symbol vector $\mathbf{s}$ is given by $\mathbf{x} = \mathbf{P}\mathbf{s}$ and using (29), we wish to find $\boldsymbol{\theta}_T$ such that $\Omega_T^H(\mathbf{z}_T;\boldsymbol{\theta}_T) = \mathbf{P}\mathbf{s}$.

We use an HDNN where the digital DNN has $L_D^{(1)}$ hidden layers of size Kc_1 and the ADNN has $L_A^{(1)}$ hidden layer of size Nc_2. These hyperparameters (which control the depth of the network and width of the hidden layers) may vary based on the number of antennas and the number of RF chains and can be obtained by cross referencing. In practice, as will be illustrated in Section VI, we find that a single analog layer, i.e. $L_A^{(1)} = 1$, is adequate to emulate the performance of the FD optimal eigen-mode HBF system in the case $N_{RF} \geq K$. For each channel instance $\mathbf{H}_\mathrm{D}$, the network is trained by generating random symbol vectors and the corresponding desired transmitted signals, as represented by the dataset $\mathcal{D}_\mathbf{H} = \{(\mathbf{s}^{(i)}, \mathbf{P}\mathbf{s}^{(i)})\}_{i=1}^{N_D}$. Since the DNNs must be trained for regression, selecting the right loss function is crucial, especially in mmWave HBF where a slight deviation from optimality can cause large performance degradation. From an optimization perspective, the ℓ_1 norm results in a sparser error vector which compared to the ℓ_2 norm, which is more suitable for our purpose. Consequently, we propose to use the mean absolute error loss function, expressed as

$$L(\boldsymbol{\theta}) = \frac{1}{N_D}\sum_{\mathbf{s}\in\mathcal{B}_j} |\Omega_T^H(\mathbf{z}_T;\boldsymbol{\theta}_T) - \mathbf{P}\mathbf{s}|, \tag{36}$$

for training within each mini-batch $\mathcal{B}_j$. Finally, using an optimization method such as adaptive moment estimation (Adam) the weights and biases of the network are updated during the backpropagation phase.

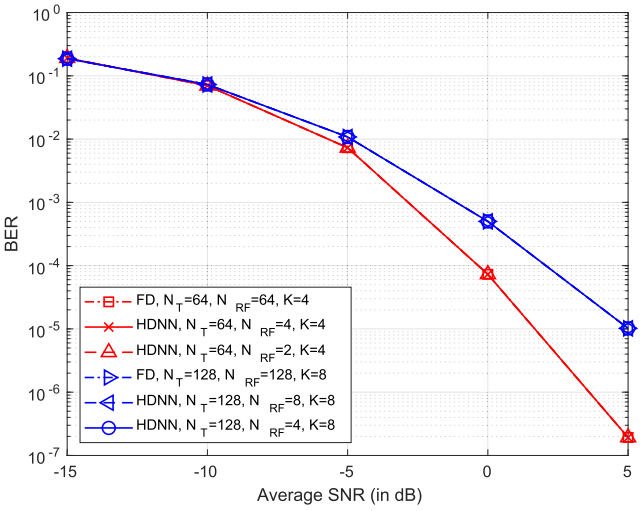

FIGURE 6. BER versus SNR of proposed HDNN designs and FD beamforming for downlink connection in 64 × 4 and 128 × 8 MIMO systems.

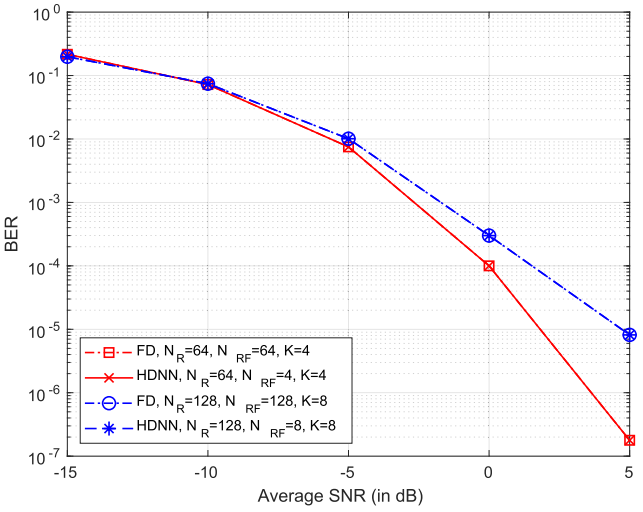

FIGURE 7. BER versus SNR of proposed HDNN designs and FD beamforming for uplink connection in 64 × 4 and 128 × 8 MIMO systems.

For the uplink beamfomer (combiner), a similar network structure with the same hyperparameters as the downlink case can be used, as previously shown in Fig. 5. The optimal FD combiner matrix $\mathbf{C}$ can be obtained from

$$\max_{\mathbf{C}} \quad \log_2 \det\left(\mathbf{I}_K + \rho(\mathbf{C}\mathbf{C}^H)^{-1}\mathbf{C}\mathbf{H}_\mathrm{U}\mathbf{H}_\mathrm{U}^H\mathbf{C}^H\right). \tag{37}$$

By writing the singular value decomposition of the uplink channel as $\mathbf{H}_\mathrm{U} = \mathbf{U_U}\Sigma_\mathbf{U}{\mathbf{V}_U}^H$, we find, $\mathbf{C}^H = \mathbf{U}_U^a$, where $\mathbf{U}_U = [\mathbf{U}_U^a, \mathbf{U}_U^b]$ and $\mathbf{U}_U^a$ contains the first K columns of $\mathbf{U}_U$, corresponding to the K dominant singular values. Training the DNN for uplink is not as straightforward as for the downlink because the received signal is contaminated by noise, which must be accounted for. Consequently, we let

$$\mathbf{y} = \rho\mathbf{z} + \mathbf{n}, \tag{38}$$

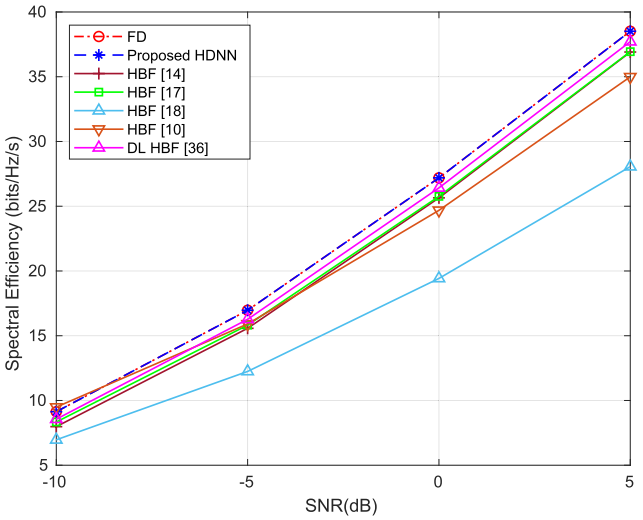

FIGURE 8. Spectral efficiency versus SNR of different methods for downlink connection from a $N = 128$ massive-MIMO BS to a UE with $M = 8$ antennas.

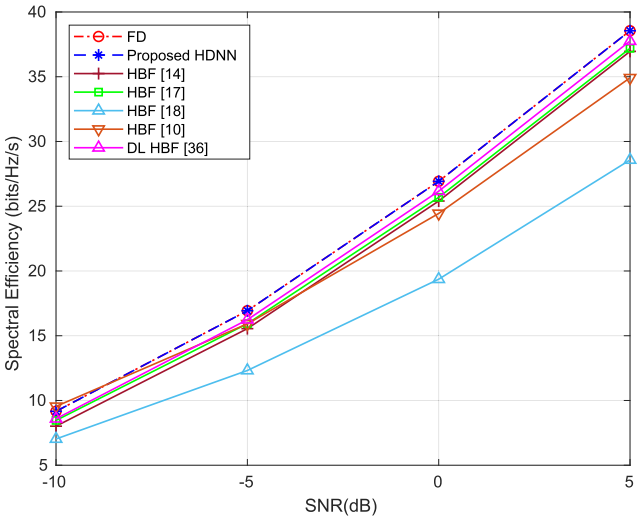

FIGURE 9. Spectral efficiency versus SNR of different methods for uplink connection from a UE with $M = 8$ to a massive-MIMO BS with $N = 128$.

where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is a Gaussian random vector. Our goal is to find $\boldsymbol{\theta}_R$ such that $\Omega_R^H(\mathbf{y}; \boldsymbol{\theta}_R) = \mathbf{Cy}$, i.e., the network is trained for the dataset $\mathcal{D}_\mathbf{H} = \{(\mathbf{y}^{(i)}, \mathbf{Cy}^{(i)})\}_{i=1}^{N_D}$. The same loss function and optimizer as in the downlink can be used for learning the network parameters.

B. DOWNLINK BEAMFORMING WITH $N_{RF} < K$

To illustrate the potential of the proposed framework, in this subsection we present a hybrid precoder which can achieve the same performance as the FD beamforming with less than K RF chains. From (1), we can see that to achieve the same performance as the FD beamforming in conventional HSP, i.e., $\mathbf{P}_{\text{FD}} = \mathbf{P}_\text{A}\mathbf{P}_\text{D}$, the minimum number of required RF chains must be equal to rank($\mathbf{P}_{\text{FD}}$), where $\mathbf{P}_{\text{FD}} \in \mathbb{C}^{N\times K}$. In practice, we usually have rank($\mathbf{P}_{\text{FD}}$) $= K$, which is why in the HBF literature it is assumed that $N_{RF} \geq K$. Even with this assumption achieving the same performance as the FD beamforming is generally not possible [15]–[22], except under certain conditions [23]–[25].

Using the proposed framework, we design an HDNN in which a digital DNN with $L_D^{(2)}$ layers of size Kc_3 is connected via $K/2$ RF chains to an ADNN with $L_A^{(2)}$ hidden layer of size Nc_4, where c_3 and c_4 are network parameters that control the width of hidden layers in baseband and analog neural networks, respectively.

The same training procedure as discussed in the previous subsection is used to update the network parameters. The proposed HBF with $K/2$ RF chains can achieve the same performance as a FD system as will be shown in the next section. In this case, however, a larger number of analog layers is needed to achieve the desired degree of approximation with the HDNN, i.e. $L_A^{(2)} > 1$.

It is possible to design a similar HDNN for the uplink beamforming, however, in order to achieve the same performance as the FD combiner, due to presence of noise, a more intricate DNN architecture and training process is required which remains a topic for future studies.

TABLE 2. Simulation parameters.

Parameter	Symbol	Values
Number of clusters	N_c	5
Rays per cluster	N_ray	10
Path gain	α_{ij}	$\mathcal{CN}(0,1)$ for $i \in [1, N_\text{c}]$, $j \in [1, N_\text{ray}]$
AoA	θ_{ij}^r	Laplacian distribution with mean $\bar{\theta}_{ij}^\text{r}$
AoD	θ_{ij}^t	Laplacian distribution with angle $\bar{\theta}_{ij}^\text{t}$
Mean cluster angles	$\bar{\theta}_{ij}^\text{r}$, $\bar{\theta}_{ij}^\text{t}$	Uniformly distributed in $[0, 2\pi]$
Angular spread	θ_AS	10°
Antenna spacing	d	$\frac{\lambda}{2}$

C. DISCUSSION

Here, we briefly discuss how the proposed framework can be leveraged for achieving performance of the FD systems.

The proposed HDNN takes advantage of the universality of DNNs [48]–[51] which allows for reduction in the number of RF chains. In effect, this is similar to the ability of autoencoders to outperform principal component analysis (PCA) in dimensionality reduction [70]. However, depending on the specific task the HDNN may suffer from a bottle necking effect in the flow of information from the digital DNN to the ADNN and vice versa. This problem can be remedied by increasing the width and depth of the ADNN or the digital DNN, as it has been shown that ReLU deep neural networks are universal approximators [51]–[54]. Consequently, by properly designing the HDNN structure, any continuous function from $\mathbf{s}$ to $\mathbf{x}$ and from $\mathbf{y}$ to $\hat{\mathbf{s}}$ can be approximated by the HDNN, thus, driving the loss function (36) arbitrarily close to zero. Simulation results in subsequent section verifies that the presented HDNN in this section can in fact achieve FD performance both in uplink and downlink connections.

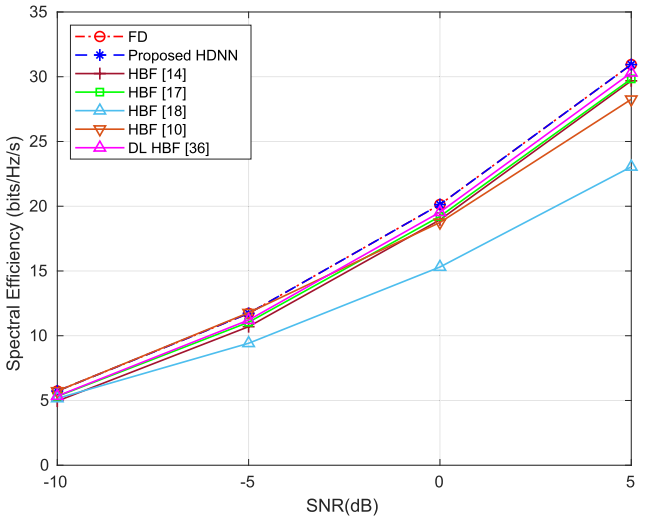

FIGURE 10. Spectral efficiency versus SNR of different methods for downlink connection of $N = 64$ massive-MIMO BS and a UE with $M = 4$ antennas.

VI. SIMULATION RESULTS

In this section, simulation results are presented to illustrate the merits of the proposed HDNN introduced in Section V. Comparisons with FD beamforming, as well as benchmark hybrid designs from the literature are included.

A. METHODOLOGY

We use the following mmWave massive-MIMO channel model [10], [14],

$$\mathbf{H} = \sqrt{\frac{NM}{N_{\rm c}N_{\rm ray}}} \sum_{i=1}^{N_{\rm c}} \sum_{j=1}^{N_{\rm ray}} \alpha_{ij} \mathbf{a}_{\rm r}(\theta_{ij}^{\rm r}) \mathbf{a}_{\rm t}(\theta_{ij}^{\rm t})^H, \tag{39}$$

where $N_{\rm c} = 5$ is the number of clusters, and $N_{\rm ray} = 10$ is the number of rays in each cluster. Similar to [10], [14], the path gains are independently generated as $\alpha_{ij} \sim \mathcal{CN}(0, 1)$. The transmit and receive antenna responses are denoted by $\mathbf{a}_{\rm r}(\theta_{ij}^{\rm r})$ and $\mathbf{a}_{\rm t}(\theta_{ij}^{\rm t})$ respectively. For simplicity, a uniform linear array of size N with half-wavelength spacing is employed,[4] where,

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}}[1, e^{j\pi \sin(\phi)}, \ldots, e^{j(N-1)\pi \sin(\phi)}]. \tag{40}$$

The angles of arrival $\theta_{ij}^{\rm r}$ and departure $\theta_{ij}^{\rm t}$ are independently generated according to the Laplacian distribution with mean cluster angles $\bar{\theta}_{ij}^{\rm r}$ and $\bar{\theta}_{ij}^{\rm t}$, uniformly distributed in $[0, 2\pi]$, while the angular spread is set to 10 degrees within each cluster [14]. We further assume that the channel estimation and system synchronization are perfect. For convenience, the simulation parameters are summarized in Table 2.

[4]Other structures such as uniform rectangular array can be also deployed as the proposed method does not depend on the antenna configuration of the system.

TABLE 3. Implementation details of the HDNN with N antennas for the two cases $N_{RF} = K$ and $N_{RF} = K/2$.

N_{RF}=K	Hidden layers	Neurons per layer
Digital DNN	5	$2N$
ADNN	1	$3N$
N_{RF}=$K/2$	Hidden layers	Neurons per layer
Digital DNN	4	$2N$
ADNN	4	$6N$

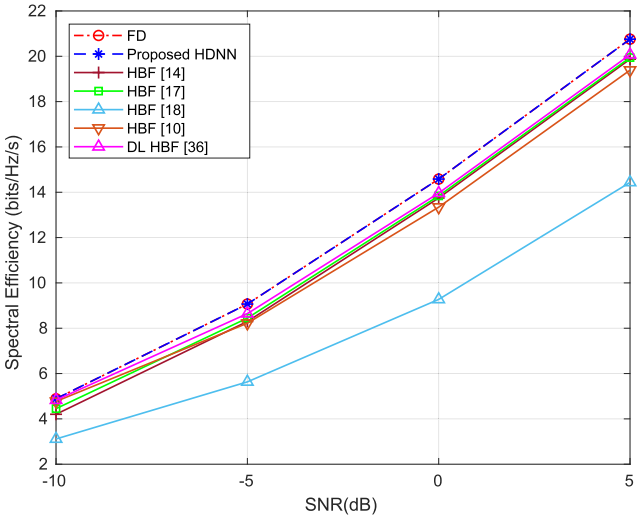

FIGURE 11. Spectral efficiency versus SNR of different methods for uplink connection a UE with $M = 4$ and a massive-MIMO BS with $N = 64$.

The parameters of the proposed HDNN are set as follows: For the case $N_{RF} = K$, the layer configuration parameters are set to $L_D^{(1)} = 5$, $L_A^{(1)} = 1$ and $c_1 = 2$, $c_2 = 3$. For the case $N_{RF} = K/2$, the model parameters are set to $L_D^{(2)} = 4$, $L_A^{(2)} = 4$, $c_3 = 2$ and $c_4 = 6$. The HDNN implementation details are summarized in Table 3.

For each channel coherence time interval T_C, the downlink HDNN is trained with $\mathcal{D}_{\mathbf{H}} = \{(\mathbf{s}^{(i)}, \mathbf{Ps}^{(i)})\}_{i=1}^{N_D}$ where $\mathbf{s}^{(i)}$ are the actual transmitted symbol vectors with duration T_S and integer N_D is such that $N_D T_S \leq T_C$. We find that for the downlink, the network can be trained with a small of symbol vectors. Specifically, for the presented results below, we set $N_D = 10$. Training is performed using the Adam optimizer with learning rate 0.001 for 500 epochs with mini-batch size of 5. For uplink however, since the goal is to decode the transmitted signal vectors in the presence of noise, a larger data set must be employed. Fortunately, this data can be easily generated under the Gaussian assumption for the signal and noise using (38). In our experiments, 5000 random vectors $\mathbf{y}^{(i)}$ are hence generated as synthetic combiner inputs and used along the corresponding combiner outputs $\mathbf{Cy}^{(i)}$ in the training phase. Training is performed using the Adam optimizer with learning rate 0.001 for 5 epochs with mini-batch size of 50.

B. RESULTS AND DISCUSSION

In this part, we investigate the spectral efficiency (SE) and bit error rate (BER) performance for the proposed HDNN design and the optimal FD beamforming. Since both schemes achieve the same performance, as will be shown, further comparisons between our proposed approach and existing hybrid designs follow directly by consulting the corresponding studies where FD beamforming is used as a benchmark, e.g., [10], [14], [17], [19], [34]–[37].

The BER performance versus signal-to-noise ratio (SNR) (which corresponds to parameter ρ in (1) under normalization of the additive noise) for 64×4 and 128×8 point-to-point MIMO setups is illustrated for uplink and downlink connections in Fig. 6 and 7, respectively. In both cases the multi-antenna UE performs FD beamforming, 4-QAM constellation is used and the number of transmitted symbols is equal to the number of UE's antennas, i.e., $K = M$. It can be observed that the proposed HDNN design matches the performance of the FD system with only K RF chains. Moreover, for downlink beamforming, our proposed design in Subsection V-B achieves the same performance with $N_{RF} = K/2$ RF chains which to the best of our knowledge has not been reported in the literature before. This possibility appears to be a consequence of the universal approximation property of DNNs [48]–[54].

Next, we compare the SE of our proposed HDNN design in Subsection V-A to FD beamforming as well as benchmark HBF designs in, [10], [14], [17], [19] and [36]. The massive-MIMO BS, which is equipped with $N = 128$ antennas, serves a single UE equipped with $M = 8$ antennas; $K = 8$ data streams are employed. While the FD system requires $N_{RF} = 128$ RF chains, hybrid beamforming techniques only use $N_{RF} = 8$ RF chains. The SE of the downlink and uplink connections for the various schemes are plotted in Fig. 8 and Fig. 9, respectively. As shown in the figures, for HBF designs [10], [14], [17], [19] and [36] the cost of using less RF chains is degraded performance compared to FD systems, whereas, the proposed HDNN matches the rate of the FD beamforming with the same number of RF chains as the existing hybrid designs.

Fig. 10 illustrates the downlink SE of a massive-MIMO BS equipped with $N = 64$ antennas and serving a UE with $M = 4$ antennas by means of $K = 4$ substreams. It can be observed that the proposed HDNN matches the SE of the FD beamforming while outperforming the benchmark HBF designs. The SE of the uplink connection for the same system configuration used in reverse (i.e., uplink) direction is depicted in Fig. 11. As in the downlink case, the proposed HDNN design achieves the same rate as the FD system, and hence a higher rate than the existing hybrid designs.

VII. CONCLUSION

In this paper, we presented a general DL framework for efficient design and implementation of HSP in massive-MIMO systems. By exploiting the fact that any complex matrix can be written as a scaled sum of two matrices with unit-modulus entries, we first presented a novel ADNN which can be implemented with common RF. This structure was subsequently embedded into an extended HDNN architecture to facilitate the implementation of mmWave massive-MIMO systems and improving their performance. Since the proposed HDNN architecture enables HSP-based massive-MIMO transceivers to approximate any desired transmitter and receiver mapping with arbitrary precision, we were able to present a new HDNN-based beamformer design that can achieve the same performance as FD beamforming, with reduced number of RF chains. Simulation results were finally presented which confirmed that our design can achieve the same performance as the FD systems.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[4] A. Schumacher, R. Merz, and A. Burg, "Adding indoor capacity without fiber backhaul: An mmWave bridge prototype," *IEEE Commun. Mag.*, vol. 59, no. 4, pp. 110–115, Apr. 2021.
[5] Y. Xing, T. S. Rappaport, and A. Amitava, "Millimeter wave and sub-THz indoor radio propagation channel measurements, models, and comparisons in an office environment," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3151–3155, Oct. 2021.
[6] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.
[7] 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.
[8] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
[9] L. Yang and W. Zhang, "Hybrid precoding design achieving fully digital performance for millimeter wave communications," *J. Commun. Inf. Netw.*, vol. 3, no. 4, pp. 74–84, Dec. 2018.
[10] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. Ben Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019.
[11] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.
[12] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.
[13] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog–digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.
[14] 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.
[15] 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.
[16] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[17] 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.

[18] R. Mai, D. H. N. Nguyen, and T. Le-Ngoc, "MMSE hybrid precoder design for millimeter-wave massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

[19] D. H. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, Jr., "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, 2017.

[20] A. Morsali and B. Champagne, "Robust hybrid analog/digital beamforming for uplink massive-MIMO with imperfect CSI," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[21] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, Jun. 2016.

[22] L. Jiang and H. Jafarkhani, "MmWave amplify-and-forward MIMO relay networks with hybrid precoding/combining design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1333–1346, Feb. 2020.

[23] A. Morsali, A. Haghighat, and B. Champagne, "Realizing fully digital precoders in hybrid A/D architecture with minimum number of RF chains," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2310–2313, Oct. 2017.

[24] A. Morsali, S. Norouzi, and B. Champagne, "Single RF chain hybrid analog/digital beamforming for mmwave massive-MIMO," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.

[25] A. Morsali and B. Champagne, "Achieving fully-digital performance by hybrid analog/digital beamforming in wide-band massive-MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 5125–5129.

[26] A. Morsali, A. Haghighat, and B. Champagne, "Generalized framework for hybrid analog/digital signal processing in massive and ultra-massive-MIMO systems," *IEEE Access*, vol. 8, pp. 100262–100279, 2020.

[27] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[28] F. Khalid, "Hybrid beamforming for millimeter wave massive multiuser MIMO systems using regularized channel diagonalization," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 705–708, Jun. 2019.

[29] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1445–1459, Mar. 2018.

[30] C.-K. Ho, H.-Y. Cheng, and Y.-H. Huang, "Hybrid precoding processor for millimeter wave MIMO communications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 12, pp. 1992–1996, Dec. 2019.

[31] B. Zhu, J. Wang, L. He, and J. Song, "Joint transceiver optimization for wireless communication PHY using neural network," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1364–1373, Jun. 2019.

[32] Z. Qin, H. Ye, G. Y. Li, and B. H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Mar. 2019.

[33] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, Feb. 2020.

[34] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[35] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[36] T. Lin and Y. Zhu, "Beamforming design for large-scale antenna arrays using deep learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 103–107, Jan. 2020.

[37] L. Sung and D.-H. Cho, "Multi-user hybrid beamforming system based on deep neural network in millimeter-wave communication," *IEEE Access*, vol. 8, pp. 91616–91623, 2020.

[38] J. Tao, Q. Wang, S. Luo, and J. Chen, "Constrained deep neural network based hybrid beamforming for millimeter wave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[39] H. Iimori, G. T. F. De Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "A stochastic gradient descent approach for hybrid mmWave beamforming with blockage and CSI-error robustness," *IEEE Access*, vol. 9, pp. 74471–74487, 2021.

[40] 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, May 2021.

[41] P. Dong, H. Zhang, and G. Y. Li, "Framework on deep learning-based joint hybrid processing for mmWave massive MIMO systems," *IEEE Access*, vol. 8, pp. 106023–106035, 2020.

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

[43] 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.

[44] X. Bao, W. Feng, J. Zheng, and J. Li, "Deep CNN and equivalent channel based hybrid precoding for mmWave massive MIMO systems," *IEEE Access*, vol. 8, pp. 19327–19335, 2020.

[45] S. Huang, Y. Ye, and M. Xiao, "Learning-based hybrid beamforming design for full-duplex millimeter wave systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 120–132, Mar. 2021.

[46] 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," *IEEE Trans. Cogn. Commun. Netw.*, p. 1, 2021.

[47] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, "The expressive power of neural networks: A view from the width," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, Dec. 2017, pp. 6231–6239.

[48] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.

[49] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Comput.*, vol. 3, no. 2, pp. 246–267, 1991.

[50] M. Lensho, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Netw.*, vol. 6, no. 6, pp. 861–876, 1993.

[51] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.

[52] H. N. Mhaskar and T. Poggio, "Deep vs. shallow networks: An approximation theory perspective," *Anal. Appl.*, vol. 14, no. 6, pp. 829–848, 2016.

[53] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations," *Mathematics*, vol. 7, p. 992, Oct. 2019.

[54] B. Hanin and M. Sellke, "Approximating continuous functions by ReLU nets of minimal width," 2017, *arXiv:1710.11278*.

[55] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[56] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, p. 3.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[58] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. Pal, "Deep complex networks," 2018, *arXiv:1705.09792*.

[59] D. M. Pozar, *Microwave Engineering*. Hoboken, NJ, USA: Wiley, 2005.

[60] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[61] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, nos. 5–6, pp. 138–139, 1971.

[62] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. C-20, no. 4, pp. 774–780, Aug. 1972.

[63] A. Morsali, S. S. Hosseini, B. Champagne, and X.-W. Chang, "Design criteria for omnidirectional STBC in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1435–1439, Oct. 2019.

[64] Y.-H. Lin and Z.-M. Tsai, "A wideband compact 5-bit phase shifter with low loss and RMS errors for 5G applications," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 10, pp. 1134–1137, Oct. 2021.

[65] T. Singh and R. R. Mansour, "Loss compensated PCM GeTe-based latching wideband 3-bit switched true-time-delay phase shifters for mmWave phased arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 3745–3755, Sep. 2020.
[66] H. Zhang, M. Gu, X. D. Jiang, J. Thompson, H. Cai, S. Paesani, R. Santagati, A. Laing, Y. Zhang, M. H. Yung, Y. Z. Shi, F. K. Muhammad, G. Q. Lo, X. S. Luo, B. Dong, D. L. Kwong, L. C. Kwek, and A. Q. Liu, "An optical neural chip for implementing complex-valued neural network," *Nature Commun.*, vol. 12, no. 1, pp. 1–11, Dec. 2021.
[67] T. Talaśka, "Components of artificial neural networks realized in CMOS technology to be used in intelligent sensors in wireless sensor networks," *Sensors*, vol. 18, no. 12, p. 4499, Dec. 2018.
[68] N. Leroux, A. De Riz, D. Sanz-Hernández, D. Marković, A. Mizrahi, and J. Grollier, "Convolutional neural networks with radio-frequency spintronic nano-devices," 2021, *arXiv:2111.04961*.
[69] Y. Yang, F. Gao, Z. Zhong, B. Ai, and A. Alkhateeb, "Deep transfer learning-based downlink channel prediction for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7485–7497, Dec. 2020.
[70] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

ALIREZA MORSALI (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Shahid Bahonar University of Kerman, Kerman, Iran, in 2009 and 2011, respectively, and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2021. Since then, he has been a Research Associate with McGill University. His research interests include coding theory, signal processing, wireless communications, and artificial intelligence. He was a recipient of several awards and scholarships, including the McGill Engineering Doctoral Award, the STARaCom Collaborative Grant, and the McGill BLUE Fellowship.

AFSHIN HAGHIGHAT (Senior Member, IEEE) received the B.S. degree from the K. N. Toosi University of Technology, Tehran, Iran, in 1992, and the M.A.Sc. and Ph.D. degrees from Concordia University, Montreal, QC, Canada, in 1998 and 2005, respectively, all in electrical engineering. He is a Principal Engineer with InterDigital Inc., where he has been involved in advanced research and standardization of 5G NR and 4G LTE cellular systems. Prior to InterDigital, he was with Harris Corporation, from 1998 to 2005, where he contributed and led the development of advanced software-defined modems for high data rate backhaul point-to-point microwave systems. Before joining Harris, he worked with SR Telecom, from 1997 to 1998, where he was responsible for design and implementation of various microwave integrated transceiver platforms for point-to-multipoint radios. His research interests include RF systems, communications, and signal processing.

BENOIT CHAMPAGNE (Senior Member, IEEE) received the B.Eng. degree in engineering physics from Polytechnique Montréal, in 1983, the M.Sc. degree in physics from the University of Montreal, in 1985, and the Ph.D. degree in electrical engineering from the University of Toronto, in 1990. From 1990 to 1999, he was an Assistant Professor and then an Associate Professor with INRS Telecommunications, Montreal. In 1999, he joined McGill University, where he is currently a Full Professor with the ECE Department. His research interests include statistical signal processing and wireless communications, where he has coauthored more than 300 publications. He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.