IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 December 2022, accepted 2 February 2023, date of publication 14 February 2023, date of current version 23 February 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3244928

RESEARCH ARTICLE

Deep Learning Based Interference Exploitation in 1-Bit Massive MIMO Precoding

MOHSEN HOSSIENZADEH[1], HASSAN AGHAEINIA[1], AND MOHAMMAD KAZEMI[2], (Member, IEEE)
[1]Department of Electrical Engineering, Amirkabir University of Technology, Tehran 15875-4413, Iran
[2]Department of Electrical and Electronics Engineering, Bilknet University, 06800 Ankara, Turkey
Corresponding author: Hassan Aghaeinia (aghaeini@aut.ac.ir)

ABSTRACT In this paper, we focus on one-bit precoding approach for downlink massive multiple-input multiple-output (MIMO) systems, where we exploit the concept of constructive interference (CI) employing deep learning (DL) techniques. One of the main performance limiting factors in wireless communication systems is interference, which needs to be minimized or mitigated. By controlling the interference signals in order to add up constructively at the receiver side, there is a possibility to improve the system performance. This paper presents a DL-based one-bit precoding scheme that improves the massive MIMO performance via CI exploitation in the presence of one-bit digital to analog converters (DAC) as a hardware impairment. More precisely, for phase shift keying signaling, we first formulate the optimization problem in order to maximize the CI effects in the case of a base station equipped with one-bit DACs. Then, after solving the optimization problem and creating a large enough dataset, a DL network is trained to do the precoding. Numerical results show that the DL-based solution approaches the performance of the conventional interference exploitation one-bit precoding schemes in the massive MIMO systems while having an order of magnitude less complexity.

INDEX TERMS Massive MIMO, interference exploitation, one-bit DAC, precoding, deep learning.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology has been widely studied in the last twenty years and has been adapted by many wireless standards since it can provide significant gains in both throughput and reliability. Massive MIMO systems, where each base station (BS) is equipped with a large number of antennas, have been proposed in [1] and it has become a key enabling technology for the fifth-generation (5G), sixth-generation (6G) and future wireless communication systems.

Precoding has attracted significant interest in the development of 5G and 6G [2] since it makes it possible to simultaneously transmit data to multiple receivers in multi-antenna wireless communication systems. Dirty paper coding (DPC) precoding technique is capable of achieving the channel capacity theoretically [3], however, it is quite impractical due to the infinite source alphabet assumption and its high computational complexity. Therefore, less complex linear precoding methods such as maximum ratio transmission (MRT), zero-forcing (ZF), and regularized zero-forcing (RZF) have become appealing and attracted more research attention, however, these methods do not fully eliminate the multi-user interference [4], [5], [6]. In addition to all the above-mentioned precoding methods, optimization-based linear precoding methods which try to minimize power and signal-to-interference-plus-noise ratio (SINR) have received increasing research attention recently [7], [8].

One of the main challenges in designing systems with large-scale antenna array is that the implementation of the conventional digital beamforming strategies may not be practical [3], [4], [5], [6], [7], [8] since they require a dedicated high-resolution RF chain for each antenna element. When a transmitter is equipped with a large number of antennas, the fully digital beamforming architecture leads to high hardware complexity and excessive circuit power consumption. All these drawbacks make fully digital processing unappealing for a massive MIMO base station. In order to

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

address these challenges, some emerging techniques such as hybrid analog-digital (AD) precoding [9] and constant-envelope (CE) precoding assume either continuous-phase phase shifters [10] or discrete-phase ones [11] combined with low-resolution digital to analog converters (DAC) to reduce the hardware complexity and power consumption of a massive MIMO base station.

The use of low-resolution DACs, which is the focus of this paper, is a novel precoding scheme in which two low-resolution DACs are dedicated to each antenna element in order to reduce the hardware cost and power consumption per RF chain. Employing a high-resolution RF chain per antenna element potentially leads to high power consumption at the terminals with large-scale antenna arrays. Motivated by the fact that the power consumption of DACs employed in the RF chains grows exponentially with the number of quantization bits, in the massive MIMO where a large number of DACs are required, adopting the low-resolution DACs instead of high-resolution ones can greatly reduce the power consumption at the BS [12], [13]. In this paper, we design a precoder when one-bit DACs are employed at antenna elements of the BS, i.e., each DAC output can only have two distinct values. One-bit precoding is a novel precoding approach where BS employs two one-bit DACs for each antenna element, one for the real part and the other for the imaginary part [14]. The one-bit precoding has recently attracted lots of attention due to its ability to reduce complexity and power consumption significantly. There have been some works that consider precoding design in the presence of one-bit DACs. Some use linear precoding in which the quantization process is directly performed on the final linear precoded signal [15], [16], [17]. However, an error floor can still be observed in these schemes for the one-bit precoding. To address this problem, some non-linear precoding strategies which are more sophisticated are proposed in [18], [19], [20], and [21], where nonlinear one-bit precoding is designed to directly map the received symbol vector to a one-bit transmit signal vector through a symbol-level operation.

Constructive interference (CI) is the interference that pushes the received signals away from all of their corresponding decision boundaries of the modulated symbol constellation, which thus contributes to the useful signal power. It has been shown that precoding based on the CI improves bit error rate (BER) performance compared to the traditional methods in small-scale MIMO and large-scale MIMO systems with phase shift keying (PSK) modulation or quadrature amplitude modulation (QAM) [22], [23], [24].

In communication systems, machine learning (ML), especially deep learning (DL), techniques have recently been used to learn transmitter and receiver component functionalities such as finding a low dimensional representation of network input [25], [26], [27], [28]. Consider an infinite resolution zero-forcing precoding with high-resolution DACs in the BS. In the ideal case of having no noise, in order to have the same received symbol as the transmitted one, the precoding matrix and the transmit precoded vector should be $\boldsymbol{P} = \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}$ and $\boldsymbol{x} = \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{s}$, respectively. In this case, the elements of $\boldsymbol{x}$ have infinite resolution. In order to have $\boldsymbol{x}$ with just two distinct phases (-1 and 1) for real and imaginary parts of its elements, the precoder block needs to have some important specifications as follows:

1) It should be able to extract the main features of $\boldsymbol{H}$ and $\boldsymbol{s}$ in order to send almost the same information as high-dimension representation but with lower resolution.
2) It should generate the output in such a way that it is possible to reconstruct the data $\boldsymbol{s}$ after passing through the channel.
3) It should be able to learn the channel uncertainty.

These items are important and confirmed capabilities of DL networks and show that the concept of low-dimensional representation based on deep learning is well suited for the one-bit precoding problem. This makes us eager to use the DL concept in the interference exploitation in a one-bit massive MIMO precoding scenario.

In this paper, we point out that the optimal precoding design using one-bit DACs is the crucial component of transmitters in massive MIMO systems. W show that by using machine learning techniques both precoder and target constellation can be learned. This paper leverages a novel DL structure that simultaneously uses convolutional layer (CL) and fully connected (FC) layer concepts for a downlink precoder in order to solve the CI-based one-bit precoding problem. Numerical results show that the proposed approach is robust to channel state information (CSI) uncertainty. The proposed DL-based solution has almost the same performance as the conventional interference exploitation one-bit precoding algorithms in massive MIMO systems while having an order of magnitude less complexity and being more robust to channel uncertainty and vastly changing propagation conditions. In this paper, we consider the near-optimal precoding strategy for one-bit massive MIMO systems proposed in [24], then we train a deep learning network using a large enough dataset, which reduces the computational complexity, and makes the proposed scheme suitable for practical scenarios with high performance and precision. The main contributions of the paper are as follows:

- In the downlink massive MIMO scenario, a DL-based scheme is developed for one-bit precoding design. The DL network is used to capture the information from CSI and data symbol vector, which can be considered as a black box with multiple CL and FC layers in order to realize the end-to-end precoding design. A precoding design architecture based on DL is proposed by redesigning the interference exploitation approach based on partial branch and bound technique in one-bit massive MIMO systems.
- The proposed DL-based precoding design, which is learned by a rich dataset, has low complexity and is resistant to imperfect CSI.

The remainder of this paper is organized as follows. In Section II, we present the system model and the concept of CI. Section III introduces the proposed one-bit massive

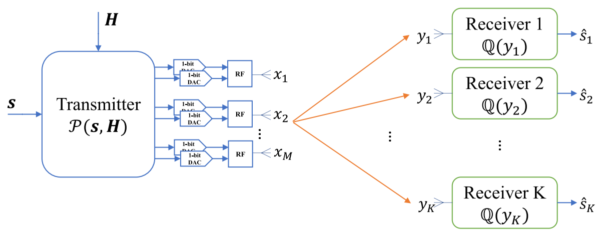


FIGURE 1. **System model.**

MIMO precoding based on the DL approach for PSK signaling. Simulation results are presented in Section IV, and the paper is concluded in Section V.

II. SYSTEM MODEL

We consider the downlink of a multi-user MIMO system, in which a BS with a large number of antennas, M, serves K single-antenna users in the same time-frequency resources. It is assumed that the BS has a perfect knowledge of CSI. The data symbol of different users can be represented in a vector $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ whose elements come from a unit-norm PSK constellation. We denote the vector of channel gains between the BS and k-th user by $\boldsymbol{h}_k \in \mathbb{C}^{M\times 1}$ which is the k-th column of the instantaneous CSI matrix, $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k]^H \in \mathbb{C}^{K\times M}$, between the BS and all users. $\boldsymbol{h}_k$ is a flat-fading Rayleigh channel vector whose elements are from a standard complex gaussian distribution $\mathcal{CN}(0, 1)$. The M-dimensional transmitted one-bit signal vector $\boldsymbol{x}$ can be written as a function of the perfectly-known CSI $\boldsymbol{H}$ and the data symbol vector $\boldsymbol{s}$ as

$$\boldsymbol{x} = \mathcal{P}(\boldsymbol{s}, \boldsymbol{H}), \tag{1}$$

where $\mathcal{P} : \mathbb{C}^{K\times 1} \times \mathbb{C}^{M\times K} \to \mathbb{C}^{M}$ is the precoder. The system model is depicted in Fig.1, where $\mathbb{Q}$ is a function that maps the received symbol y_k to the nearest 8PSK constellation point and $\hat{s}_k$ is an estimate of the transmitted symbol s_k.

The received signal at the k-th user can then be modeled as

$$y_k = \boldsymbol{h}_k^H \boldsymbol{x} + z_k, \tag{2}$$

where $\boldsymbol{x} = [x_1, \ldots, x_M]^T$ is a normalized vector such that $\|\boldsymbol{x}\|_2^2 = 1$ with its entries picked from the set $\left\{\pm\frac{1}{\sqrt{2M}} \pm \frac{1}{\sqrt{2M}}j\right\}$, and $z_k \sim \mathcal{CN}\left(0, 2\sigma^2\right)$ is the circularly symmetric additive white gaussian noise (AWGN).

III. ONE-BIT MASSIVE MIMO PRECODING BASED ON DEEP LEARNING

In this section, first, the one-bit precoding problem with the assumption of constructive interference exploitation is formulated and then after some problem manipulation, the objective function and its constraints are made convex which is then solved using CVX toolbox [24]. Finally, a deep learning network is trained by a dataset that has already been generated by the results of convex optimization.

A. PROBLEM FORMULATION

The interference that pushes the received symbol away from the detection boundaries and makes them to be more concentrated on the constellation points is called CI. CI exploitation is an appealing strategy in the physical layer of wireless communication since it transforms the power of the interfering signal into a useful signal. The most significant advantage of CI compared to conventional precoding is error rate performance improvement and power saving. On the other hand, the high complexity of designing the precoding matrix can be mentioned as a drawback since the CI precoding has to update the precoding matrix on a symbol level.

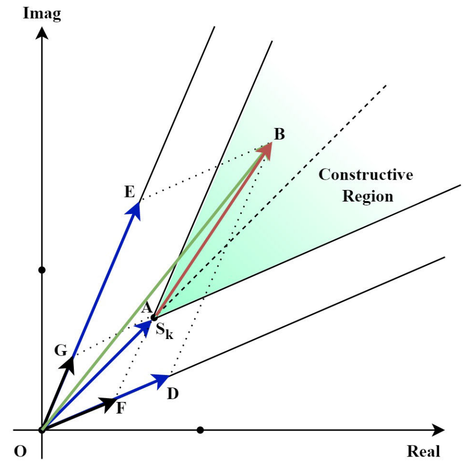


FIGURE 2. **CI condition for PSK signaling. $\overrightarrow{OA}$: The ideal transmitted symbol. $\overrightarrow{OG}$ and $\overrightarrow{OF}$ are two decomposed vectors of $\overrightarrow{OA}$ which are parallel to the detection boundaries. $\overrightarrow{OB}$: The received symbol. $\overrightarrow{AB}$: The channel effect. $\overrightarrow{OD}$ and $\overrightarrow{OE}$ are two decomposed vectors of $\overrightarrow{OB}$ which are parallel to the detection boundaries [24].**

In this section, the mathematical formulation of the CI condition for PSK modulation is presented. As shown in Fig.2, s_k is an ideal transmitted symbol denoted by $\overrightarrow{OA}$ and shown in blue, which can be decomposed into two vectors $\overrightarrow{OG} = s_k^{\mathcal{A}}$ and $\overrightarrow{OF} = s_k^{\mathcal{B}}$ which are parallel to the two detection boundaries of s_k. $\overrightarrow{OB}$ corresponds to the transmitted signal after the channel effect decomposing into vectors $\overrightarrow{OA}$ and $\overrightarrow{AB}$, equivalently it is decomposed into two vectors $\overrightarrow{OD}$ and $\overrightarrow{OE}$ which are parallel to the two detection boundaries; i.e.,

$$\begin{aligned}\overrightarrow{OB} &= \overrightarrow{OA} + \overrightarrow{AB}\\ &= \boldsymbol{h}_k^T \boldsymbol{x}\\ &= \alpha_k^{\mathcal{A}} s_k^{\mathcal{A}} + \alpha_k^{\mathcal{B}} s_k^{\mathcal{B}},\end{aligned} \tag{3}$$

where $\alpha_k^{\mathcal{A}}$ and $\alpha_k^{\mathcal{B}}$ are real variables. The error probability decreases by maximizing these two variables over all possible transmitted symbols. The following optimization problem expresses this issue.

$$\begin{aligned}\boldsymbol{x} &= \arg\max_{\boldsymbol{x}} \min_{k,u} \alpha_k^{U}\\ \boldsymbol{h}_k^T \boldsymbol{x} &= \alpha_k^{\mathcal{A}} s_k^{\mathcal{A}} + \alpha_k^{\mathcal{B}} s_k^{\mathcal{B}},\\ \forall k &\in \{1, 2, \ldots, K\},\\ \forall u &\in \{\mathcal{A}, \mathcal{B}\},\\ \forall x_n &\in \left\{\pm\frac{1}{\sqrt{2M}} \pm \frac{1}{\sqrt{2M}}j\right\}.\end{aligned} \tag{4}$$

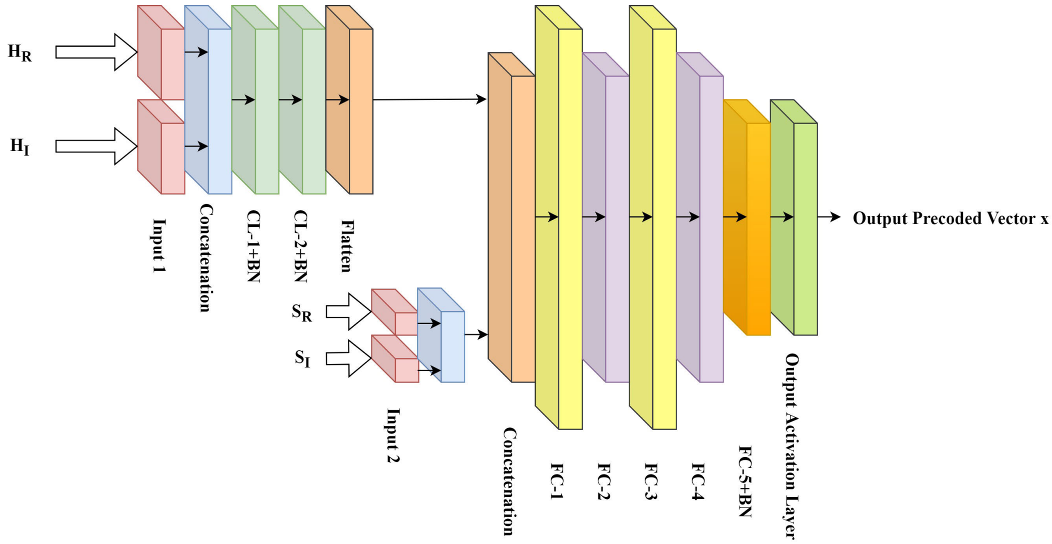


FIGURE 3. **The structure of the proposed DL based precoder.**

In order to transform this optimization problem into a convex one, we first transform the complex equations into equivalent real ones and then relax the non-convex constraints as follows. Further details about this transformation are provided in Appendix A.

$$\boldsymbol{x} = \arg\max_{\boldsymbol{x}} \min \Lambda$$
$$|\tilde{x}_m| \leq \frac{1}{\sqrt{2M}}, \quad \forall m \in \{1, 2, \ldots, 2M\}, \tag{5}$$

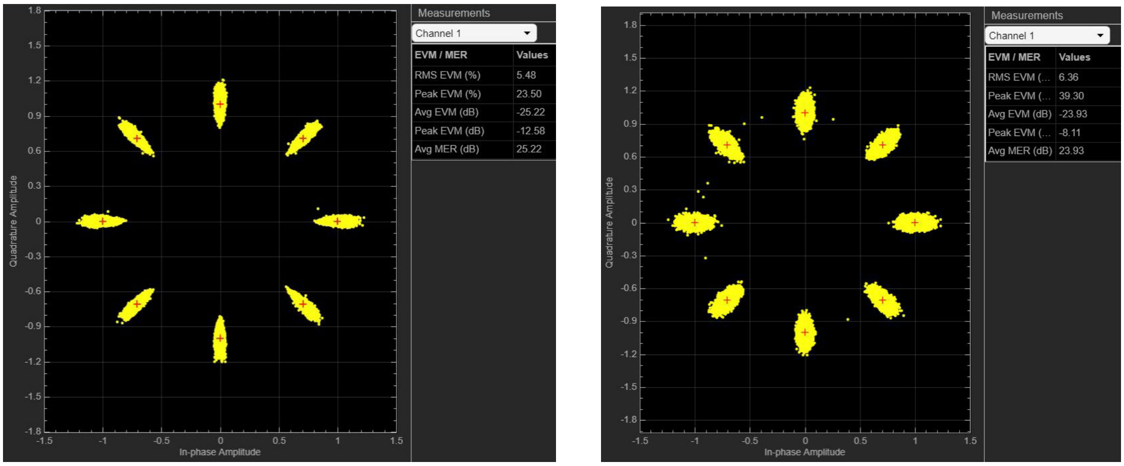
where,

$$\Lambda = \left[\alpha_1^{\mathcal{A}}, \alpha_2^{\mathcal{A}}, \ldots, \alpha_K^{\mathcal{A}}, \alpha_1^{\mathcal{B}}, \alpha_2^{\mathcal{B}}, \ldots, \alpha_K^{\mathcal{B}}\right],$$
$$\alpha_k^{\mathcal{A}} = \frac{\boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)}{\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)},$$
$$\alpha_k^{\mathcal{B}} = \frac{\boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Re}\left(s_k^{\mathcal{A}}\right) - \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)}{\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)},$$
$$\tilde{\boldsymbol{x}} = \left[\boldsymbol{Re}(\boldsymbol{x})^T, \boldsymbol{Im}(\boldsymbol{x})^T\right]^T. \tag{6}$$

It has been proved in [24] that the number of $\tilde{\boldsymbol{x}}$ entries whose amplitudes are smaller than $1/\sqrt{2M}$ is smaller than $(2K-1)$, and that at least $(2M-2K+1)$ entries already satisfy the one-bit constraints. In order to find the best choice for the other $(2K-1)$ elements, a well-known optimization method, the branch and bound (B&B) algorithm, is used. See [24] for further details. Since, searching the complete space of solutions for this optimization problem is time-consuming, in [24] a recursive procedure is proposed to solve the optimization problem in (5) using B&B algorithm.

B. DL-BASED ONE-BIT PRECODING

In this section, we consider a massive MIMO set-up and investigate the downlink precoding design using deep learning networks in the presence of one-bit DACs. We propose a supervised learning strategy using the dataset which is generated by solving the optimization problem introduced in the previous subsection. As stated before, using the classic optimization strategies in order to find the best solution for the problem in (4) is very time-consuming. There are two main reasons why deep learning is used to deal with the problem of interference exploitation in one-bit massive MIMO precoding. The first and most important reason is to reduce computational complexity. To be more specific, various algorithms such as partial branch and bound (P-BB) have been proposed for solving the problem of interference exploitation in one-bit massive MIMO precoding based on linear programming (LP) relaxation model. This approach consists of two stages: first, solving the LP relaxation model, and second, utilizing some optimization techniques in order to determine the values of elements of the LP relaxation solution that do not satisfy the one-bit constraint. The P-BB algorithm solves the subproblem in the second stage using branch and bound procedure which is not suitable for practical implementation due to its high computational complexity. To be more clear, for instance, in some cases where the user antenna ratio (K/M) is high, a large fraction of elements cannot be determined at the first stage by solving the LP relaxation model, so, the dimension of the problem of determining the values of those infeasible components at the second stage will be large and grows exponentially with respect to the number of users, which results in serious performance degradation, while the computational complexity of our DL-based solution grows linearly with respect to the number of users. The computational complexity of the DL network depends on a number of network variables such as the number of layers, the number of nodes in each layer, and the filters in each layer. In our DL network, the number of output nodes is independent of the number of users and if we double the number of users, the input feature dimension (the number of input nodes) of the DL network



(a) Partial branch and bound method

(b) Proposed deep learning-based solution

FIGURE 4. **The received constellation ($SNR = 30dB$).**

is doubled as well, so, the computational complexity of the DL network is almost doubled. So, training an efficient deep neural network in order to solve this non-linear problem can be very effective.

The structure of the proposed deep neural network is shown in Fig.3, where the channel state information matrix $\boldsymbol{H}$ and data vector $\boldsymbol{s}$ are the inputs and the precoded vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_{2M}]^T = \left[\boldsymbol{x}_R^T, \boldsymbol{x}_I^T\right]^T$ is the output with $\boldsymbol{x}_R$ and $\boldsymbol{x}_I$ being its real and imaginary parts, respectively. Note that the real and imaginary parts of the matrix $\boldsymbol{H}$ and vector $\boldsymbol{s}$ are given as separate inputs. The network consists of two CL and five fully connected (FC) layers. In order to extract the main features of matrix $\boldsymbol{H}$, a concatenation of its real and imaginary parts is processed by CL layers and then vectorized by the flattening layer. First, the CL layer processes the input by a $(3 \times 3 \times 1)$ kernel with stride (2×1) in order to extract important features between each antenna and all users. The other CL layer is used to capture more features from channel state information matrix $\boldsymbol{H}$ with two channels each with a (1×5) kernel. Then, the output data from CL-2 is vectorized by a flattening layer. It must be noted that the batch normalization layers (BN) are used after each CL to speed up the training process and prevent overfitting. FC layers are applied on two concatenated vectors, CL output and $\left[\boldsymbol{Re}(\boldsymbol{s})^T, \boldsymbol{Im}(\boldsymbol{s})^T\right]^T$, in order to do regression on these features. The leaky Relu is adopted in this structure as the activation function before BN which is represented by $\sigma_{\text{LeakyRelu}}(x) = \max(\alpha x, x)$ with α being a small positive value less than one. The output data has a dimension of $2N_t \times 1$ which contains both real and imaginary parts of the precoded vector $\boldsymbol{x}$. In order to satisfy the one-bit constraint of precoded vector $\boldsymbol{x}$, the softmax activation function is adopted as the last activation function. The parameters of the proposed network are summarized in Table 1.

The training strategy in the proposed network which is based on supervised learning is that all labeled data are first gathered by solving the B&B optimization algorithm in [24] which is shown to be a near-optimal solution,

TABLE 1. **Parameters of the proposed DL network.**

Type of layer	Parameter
CL-1	$(3 \times 3 \times 1)$
CL-2	$(1 \times 5 \times 2)$
FC-1	1024
FC-2	512
FC-3	1024
FC-4	512
FC-5	256
Output activation layer	Softmax

then after some preprocessing on data, the training is done in a supervised manner. Since the one-bit precoded massive MIMO target output just has two possible states, the problem can be considered as a multilabel classification problem. Furthermore, the loss function is binary crossentropy which is a common one for multilabel classification problems. The binary crossentropy loss function calculates the predicted precoding vector by computing the following average:

$$\text{Loss} = -\left(\frac{1}{2M}\right)\sum_{i=1}^{2M} x_i \log \hat{x}_i + (1 - x_i)\log\left(1 - \hat{x}_i\right), \quad (7)$$

where $\hat{x}_i$ is the i-th scaler value in the output layer from the sigmoid activation function, and x_i is the corresponding target value. Sigmoid activation function is a good choice since it is possible to predict two possible classes with target probabilities x_i and $(1 - x_i)$.

As shown before in (1), interference exploitation one-bit massive MIMO precoding can be written as a non-linear function of channel state information and symbol vector. The trained deep neural network using the rich datasets, which are already generated, can be employed to efficiently design the optimal output signal vector on the antenna elements and build a direct mapping from $\boldsymbol{H}$ and $\boldsymbol{s}$ to $\boldsymbol{x}$. It is worth mentioning that, based on our extensive simulations, the designed neural network is resistant to imperfect channel state information.

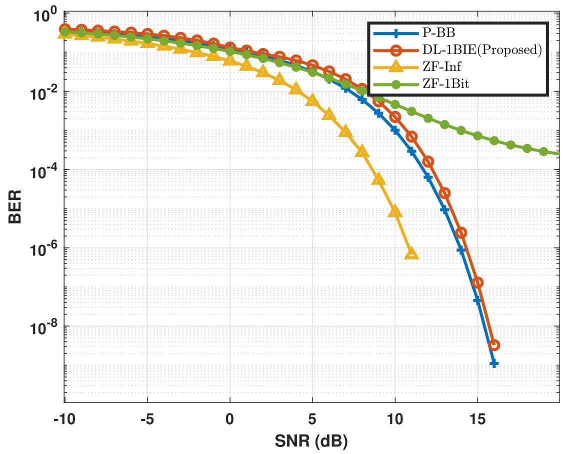


FIGURE 5. BER vs. transmit SNR for $M = 128$ and $K = 5$.

IV. SIMULATION RESULTS

In this section, in order to show the effectiveness of the proposed DL-based precoding, the simulation results are presented. In our simulations, we assume 8PSK modulation, while other PSK and QAM modulations can also be used. We consider $M = 128$ antennas and $K = 5$ users. In the first step, known perfect CSI is assumed, and then, we also investigate the effect of imperfect CSI. For the sake of efficient implementation, the proposed deep neural network is implemented by TensorFlow library. All the simulations are run on a computer with NVIDIA GeForce RTX 3090 graphical processing unit (GPU) and Intel Core i9 − 9980XE CPU. For the learning process two different datasets are considered, first a "Simple" dataset, which consists of 100,000 samples of the propagation channel $\boldsymbol{H}$ and a specific symbol vector $\boldsymbol{s}$. In this step, some DL network parameters are tuned. Next, using the tuned parameters in the first step and an "Extensive" dataset, which contains 1,000,000 realizations of $\boldsymbol{H}$, $\boldsymbol{s}$ and labeled output $\boldsymbol{x}$, the DL network is fine-tuned. $\boldsymbol{H} \in \mathbb{C}^{K\times M}$ whose elements are from a standard complex Gaussian distribution $\mathcal{CN}(0, 1)$ and $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ whose elements come from a unit-norm 8-PSK constellation are randomly generated and the corresponding outputs (the transmitted one-bit signal vectors $\boldsymbol{x}$) are labeled using P-BB algorithm. Adam optimizer is employed with the initial learning rate $\mu = 0.001$, which is decreased by a decay factor of 0.1 every 20 epoch. The mini-batch size is 256. In order to show the effectiveness of the proposed solution, the simulation results are compared with three other one-bit massive MIMO precoding approaches, namely, P-BB, infinite-resolution zero-forcing, and one-bit resolution zero-forcing.

We first illustrate some measurements on the received constellation such as RMS, EVM, Peak EVM, Avg EVM and Avg MER for the non-DL and the proposed deep learning solutions. A detailed explanation of these measurement values is provided in Appendix B. As shown in Fig.4, the received constellation points in the proposed deep neural network are slightly less concentrated compared to the non-DL solution. The mentioned measurement values are shown

TABLE 2. Running time for $M = 128$ and $K = 5$.

Environment	CPU		GPU
	DL-1BIE	P-BB	DL-1BIE
MATLAB	1.7	9.5	0.23
Python	1.1	*	0.18

on the right side of these figures. One reason for the difference between the non-DL results and deep learning-based results is the quantization which is done in the last layer of the DL network after the softmax activation layer in order to get just two separate values.

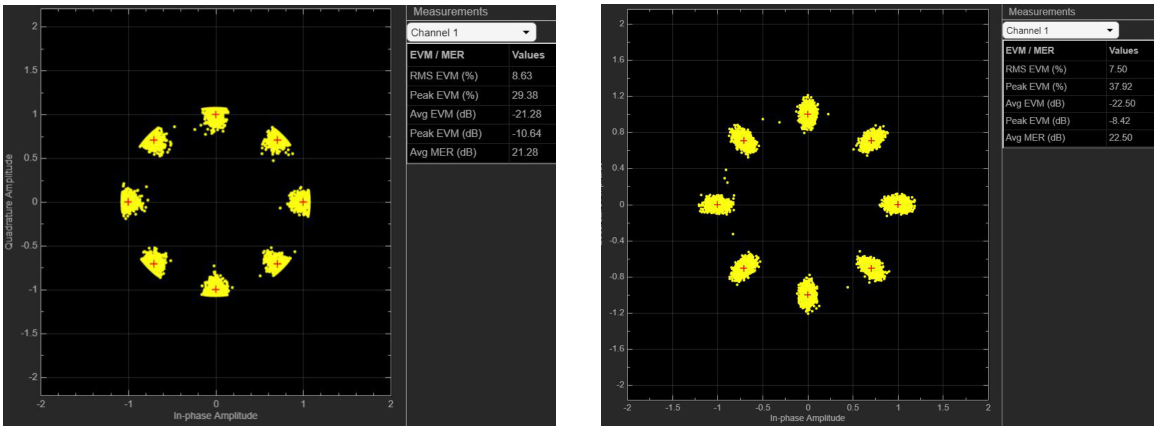
The BER for the P-BB precoding method [24], the proposed deep learning-based precoding method and zero-forcing precoding strategy with infinite and quantized resolutions in DACs are shown in Fig.5 for $M = 128$ and $K = 5$. We see that the proposed deep learning-based precoding has almost the same performance as the P-BB algorithm while having lower computational complexity (as shown in Table 2). It is also illustrated that the proposed method has a better performance than the quantized ZF as a linear precoding method. In order to provide a more clear comparison, the average running times are provided in Table 2 for $M = 128$ and $K = 5$, where DL-based 1-bit interference exploitation (DL-1BIE) is the proposed solution.

The P-BB method is run on CPU in MATLAB environment while the proposed deep learning precoding is run on both CPU and GPU in MATLAB and Python environments, respectively, for a fair comparison. It must be noted that the training process for the proposed deep learning-based precoding is done in offline mode, so the training run time is not considered in Table 2. As shown in Table 2, the running time for the proposed method is dramatically less in both CPU and GPU compared to the non-DL solution. Based on our extensive simulations, it is observed that the processing time for the CVX-based method exponentially increases with the number of users while the processing time for the proposed solution is almost linear with the number of users.

Finally, the simulation results for the case of imperfect CSI are investigated. The imperfect CSI is modeled as an additive term to the perfect channel state information as follows.

$$\widehat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{H_e}, \tag{8}$$

where $\boldsymbol{H_e}$ is the channel estimation error whose elements come from a zero-mean complex Gaussian distribution $\mathcal{CN}\left(0, \sigma_e^2\right)$. We take $\sigma_e^2 = 0.1$. As illustrated in Fig.6, the received constellation for the proposed deep learning solution and the partial branch and bound method are affected by the imperfect channel estimation. From the received constellation scatterplot of these two methods in the same simulation conditions, it is clear that the partial branch and bound method is more affected by this destructive factor. Moreover, the BER versus signal-to-noise ratio (SNR) is presented for the case of imperfect CSI in Fig.7, which shows that the proposed deep learning method has a lower bit error rate, demonstrating that it is more robust against imperfect CSI.



(a) Partial branch and bound method (b) Proposed DL-based solution

FIGURE 6. The received constellation (SNR 30*dB*) for the imperfect CSI case with $\sigma_e^2 = 0.1$.

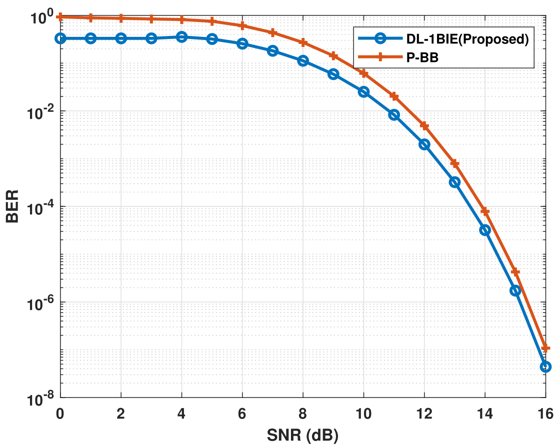


FIGURE 7. BER vs. transmit SNR with $M = 128$ and $K = 5$ for the imperfect CSI case with $\sigma_e^2 = 0.1$.

V. CONCLUSION

In this paper, exploiting the constructive interference, we provided a one-bit massive MIMO precoding scheme based on the deep learning technique. In order to avoid the high computational complexity of the existing solutions, an efficient precoder based on the deep neural network is trained. Simulation results show that the proposed method has almost the same performance as the state-of-the-art while having lower computational complexity. It is also shown that the proposed method is more robust against the imperfect CSI compared to the state-of-the-art.

APPENDIX A
EQUIVALENT REAL VERSION OF THE OPTIMIZATION PROBLEM IN (4)

The signal received by the k-t user can be written as the sum of the vectors $\vec{OG} = s_k^{\mathcal{A}}$ and $\vec{OF} = s_k^{\mathcal{B}}$,

$$\begin{aligned}
\boldsymbol{h}_k^T\boldsymbol{x} &= \alpha_k^{\mathcal{A}} s_k^{\mathcal{A}} + \alpha_k^{\mathcal{B}} s_k^{\mathcal{B}}, \\
&= \alpha_k^{\mathcal{A}}\left(\boldsymbol{Re}\left(s_k^{\mathcal{A}}\right) + j\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)\right) \\
&\quad + \alpha_k^{\mathcal{B}}\left(\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) + j\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)\right) \\
&= \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right) + j\boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)
\end{aligned}$$

where,

$$\begin{aligned}
&\alpha_k^{\mathcal{A}}\boldsymbol{Re}\left(s_k^{\mathcal{A}}\right) + \alpha_k^{\mathcal{B}}\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) \\
&\quad = \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right) \\
&\alpha_k^{\mathcal{A}}\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right) + \alpha_k^{\mathcal{B}}\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right) \\
&\quad = \boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right) \\
&\alpha_k^{\mathcal{A}} = \frac{\boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)}{\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)} \\
&\alpha_k^{\mathcal{B}} = \frac{\boldsymbol{Im}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Re}\left(s_k^{\mathcal{A}}\right) - \boldsymbol{Re}\left(\boldsymbol{h}_k^T\boldsymbol{x}\right)\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)}{\boldsymbol{Im}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Re}\left(s_k^{\mathcal{B}}\right) - \boldsymbol{Re}\left(s_k^{\mathcal{A}}\right)\boldsymbol{Im}\left(s_k^{\mathcal{B}}\right)}
\end{aligned}$$

APPENDIX B
RECEIVED CONSTELLATION MEASUREMENTS

- MER (Modulation Error Ratio) is a measure of the SNR in digital modulation applications. The MER formulation, over N symbols, is as follow:

$$MER = 10\log_{10}\left(\frac{\sum_{n=1}^{N}\left(I_k^2 + Q_k^2\right)}{\sum_{n=1}^{N}\left(e_k\right)}\right)\text{dB}, \tag{9}$$

  where,
  - $e_k = \left(I_k - \tilde{I}_k\right)^2 + \left(Q_k - \tilde{Q}_k\right)^2$,
  - I_k: In-phase measurement of the kth symbol.
  - Q_k: Quadrature phase measurement of the kth symbol.
  - I_k and Q_k represent ideal (reference) values, while $\tilde{I}_k$ and $\tilde{Q}_k$ represent measured (received) symbols.
  - N is the input vector length.
- EVM (Error Vector Magnitude) is the Root Mean Square (RMS) of the error vectors, With the following formulation:

$$\%EVM = \sqrt{\frac{\frac{1}{N}\sum_{k=1}^{N}\left(e_k\right)}{\text{EVM normalization reference}}} \times 100\%, \tag{10}$$

where we have used average constellation power and peak constellation power as two possible normalization references for EVM measurement.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[3] M. H. M. Costa, "Writing on dirty paper (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 439–441, May 1983.

[4] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[5] T. Haustein, C. von Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, "Performance of MIMO systems with channel inversion," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 35–39, Jun. 2002.

[6] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[7] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[8] F. Wang, X. Wang, and Y. Zhu, "Transmit beamforming for multiuser downlink with per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 4692–4697.

[9] 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.

[10] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An efficient manifold algorithm for constructive interference based constant envelope precoding," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1542–1546, Oct. 2017.

[11] M. Kazemi, H. Aghaeinia, and T. M. Duman, "Discrete-phase constant envelope precoding for massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2011–2021, May 2017.

[12] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[13] C. Svensson, S. Andersson, and P. Bogner, "On the power consumption of analog to digital converters," in *Proc. NORCHIP*, Linköping, Sweden, Nov. 2006, pp. 49–52.

[14] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.

[15] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.

[16] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *Proc. 16th IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, Yasmine Hammamet, Tunisia, Dec. 2009, pp. 683–686.

[17] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3381–3385.

[18] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

[19] M. Shao, Q. Li, W.-K. Ma, and A. M.-C. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, Oct. 2019.

[20] F. Sohrabi, Y.-F. Liu, and W. Yu, "One-bit precoding and constellation range design for massive MIMO with QAM signaling," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 557–570, Jun. 2018.

[21] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7559–7575, Sep. 2018.

[22] L. T. N. Landau and R. C. de Lamare, "Branch-and-bound precoding for multiuser MIMO systems with 1-bit quantization," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 770–773, Dec. 2017.

[23] S. Jacobsson, W. Xu, G. Durisi, and C. Studer, "MSE-optimal 1-Bit precoding for multiuser MIMO via branch and bound," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3589–3593.

[24] A. Li, F. Liu, C. Masouros, Y. Li, and B. Vucetic, "Interference exploitation 1-bit massive mimo precoding: A partial branch-and-bound solution with near-optimal performance," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3474–3489, Feb. 2020.

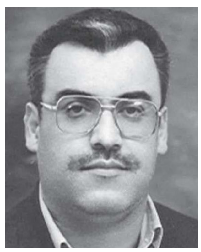
[25] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[26] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2017.

[27] C. Zhang, Y.-L. Ueng, C. Studer, and A. Burg, "Artificial intelligence for 5G and beyond 5G: Implementations, algorithms, and optimizations," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 149–163, Jun. 2020.

[28] H. Dahrouj, R. Alghamdi, H. Alwazani, S. Bahanshal, A. A. Ahmad, A. Faisal, R. Shalabi, R. Alhadrami, A. Subasi, M. T. Al-Nory, O. Kittaneh, and J. S. Shamma, "An overview of machine learning-based techniques for solving optimization problems in communications and signal processing," *IEEE Access*, vol. 9, pp. 74908–74938, 2021.

MOHSEN HOSSIENZADEH received the B.S. degree from Urmia University, and the M.S. degree from the Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran, where he is currently pursuing the Ph.D. degree with the Electrical Engineering Department. His research interests include wireless communications, signal processing, and machine learning with particular focus on m-MIMO systems.



HASSAN AGHAEINIA received the B.Sc. and M.Sc. degrees in electronic engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 1987 and 1989, respectively, and the M.Sc. and Ph.D. degrees in electronic engineering from Valenciennes University (UVHC), Valenciennes, France, in 1992 and 1996, respectively. Since 1996, he has been a Faculty Member with the Amirkabir University of Technology, where he is currently an Associate Professor with the Communication Engineering Group. His research interests include work in advanced communication systems and digital image processing.

MOHAMMAD KAZEMI (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2007 and 2010, respectively, and the Ph.D. degree in electrical engineering from the Amirkabir University of Technology, Tehran, in 2017. He was a Visiting Student with Bilkent University, Turkey, in 2015. He was a Researcher in MMWCL with the Amirkabir University of Technology, from 2017 to 2019. He is currently a Postdoctoral Fellow with the EE Department, Bilkent University. His research interests include wireless communications and signal processing, with a particular focus on massive MIMO and massive random access systems. He is also the Editorial Assistant to the Editor-in-Chief of IEEE TRANSACTIONS ON COMMUNICATIONS.