Deep Learning-Based Beamforming for Millimeter-Wave Systems Using Parametric ReLU Activation Function

Alshimaa Hamdy (✉ eng.alshimaahamdy@gmail.com)

Faculty of Artificial Intelligence, Delta Univeristy   https://orcid.org/0000-0003-4650-0484

Tarek Abed Soliman

Menofia university, Faculty of Electronics and Engineering

Mohamed Rihan

Menofia university, Faculty of Electronics and Engineering

Moawad I. Dessouky

Menofia university, Faculty of Electronics and Engineering

Deep Learning-Based Beamforming for Millimeter-Wave Systems Using Parametric ReLU Activation Function

Alshimaa H. Ismail[1], Tarek Abed Soliman[2], Mohamed Rihan[3] (Member IEEE), Moawad I. Dessouky[4]

[1]Alshimaa H. Ismail
Email: eng.alshimaahamdy@gmail.com
Affiliation: Faulty of Artificial Intelligence, Delta University
ORCID: https://orcid.org/0000-0003-4650-0484

[2]Tarek Abed Soliman
Email: t3abed@outlook.com
Affiliation: Department of Electronics and Communication Engineering, Faculty of Electronic Engineering, Menoufia University
ORCID: 0000-0003-0146-0116

[3]Mohamed Rihan
Email: mohamedrihan2005@gmail.com
Affiliation: Department of Electronics and Communication Engineering, Faculty of Electronic Engineering, Menoufia University, Menoufia 21974, Egypt
ORCID: https://orcid.org/0000-0003-4030-2559

[4]Moawad I. Dessouky
Email: dr_moawad@yahoo.com
Affiliation: Department of Electronics and Communication Engineering, Faculty of Electronic Engineering, Menoufia University, Menoufia 21974, Egypt

Abstract

Beamforming design is a crucial stage in millimeter-wave systems with massive antenna arrays. We propose a deep learning network for the design of the precoder and combiner in hybrid architectures. The proposed network employs a parametric rectified linear unit (PReLU) activation function which improves model accuracy with almost no complexity cost compared to other functions. The proposed network accepts practical channel estimation input and can be trained to enhance spectral efficiency considering the hardware limitation of the hybrid design. Simulation shows that the proposed network achieves small performance improvement when compared to the same network with the ReLU activation function.

Keywords: Deep learning, Millimeter-Wave, Beamforming, Massive MIMO

1. Introduction

5G wireless systems operate in mm-Wave bands to utilize their wider bandwidth. Large-scale antenna arrays are used to counterfeit propagation loss associated with mm-Wave bands. This configuration enables higher data rates and system capacity which both have growing demands in every aspect of our life. Mm-Waves have a short wavelength, which enables implementing more antennas into an array without increasing its dimension (enable massive MIMO) which increases the number of data streams in a cell. However, it is also expensive and power-consuming to implement a dedicated RF chain for each antenna. Recently, hybrid architecture has been receiving growing interest and consideration as a practical cost-effective solution to this problem. By splitting the beamforming process into analog and digital, hybrid beamforming enables the RF chain number to be less than the antenna number and solves the power and cost problem [1].

However, hybrid beamforming is also a complex design problem that has prominent hardware limitations and challenges. The analog beamformer architecture using phase shifters is limited by the constant modulus constraint [2,3,4]. The imperfect CSI requires a channel estimation technique and joint optimization of multiple variables that yield a non-convex problem. In this paper, we propose a using PReLU (Parametric Rectified Linear Unit) activation function in a deep learning technique to produce the optimal analog beamforming parameters while taking SNR and channel estimations as input. Our results show some enhancement over the same technique when non or other activation functions are used. In section 2, we will discuss existing approaches and their limitation in the proposed scenario followed by the problem statement and system model in section 3. Section 4 describes the proposed technique and algorithm and its simulation results are discussed in section 5. Section 6 concludes the paper.

2. Related Studies

Several studies attempted to overcome hybrid beamforming design challenges [2,3,4,5, and 6]. An orthogonal matching pursuit (OMP) and greedy-based method were proposed in [5], where a dictionary containing the array responses is constructed and used to select the precoders and combiners. However, to construct the dictionary, departure and arrival angles must be identified which limits the analog beamformer to a limited pre-defined codebook. The manifold optimization technique was proposed for the analog beamformer optimization in [2], [3]. The alternating minimization and phase extraction methods are applied to determine the analog beamformers. In [4], the authors considered a hybrid beamforming architecture in which digital (baseband) beamformer is concatenated to an RF (analog) beamformer that is implemented using phase shifters. The preceding studies implemented greedy or optimization-based techniques which raises the computation complexity and time. Besides, perfect CSI is considered in all of these algorithms.

Deep learning (DL) is extensively studied in massive MIMO and mm-Wave systems due to its ability to learn different models and solve optimization problems with less complexity and time. Deep learning is an effective approach to solve intractable problems [7]. Recently, deep-learning proved its capability to learn wireless channel's complex features [8]. Deep-learning is more robust to the practical imperfect CSI when compared to the methods assuming perfect CSI as in [2,3,4,5]. Most of the conventional beamforming algorithms require time-consuming, iterative, and high-complexity computations as in [3], [4]. On the other hand, deep neural networks (DNN) offer less complexity when solving the same optimization problems, and low computation time when deployed online after offline training [9].

DL-based techniques received a great deal of attention in communications society as a solution to several popular challenges such as coordinated beamforming for highly mobile mm-wave systems, estimating the channel parameters and direction of arrival (DOA), antenna selection, analog beam selection [10], [11], [12], [13], and [14]. DL-based techniques have also been considered for the hybrid beamforming design problem in [15], [16], [17], and [18].

The work in [15], introduced a convolutional neural network (CNN) for the hybrid joint design problem in mm-Wave and massive MIMO systems considering precoding and combining stages. The channel parameters are the input to the network and the optimal analog and digital beamformers are the output. The network was developed only for the single-user scenario. In [16], the authors proposed a DL-based beamforming technique, they introduced a beamforming neural network (BFNN) which after training, learns how to maximize the spectral efficiency considering practical CSI, the proposed network is deployed for a single-user setting and considered the precoder design only and strongly relies on the channel matrix perfectness. In [17], the authors also attempted a DL-enabled technique for the hybrid beamforming design problem, in which the hybrid precoder is selected through training based on the neural network for optimizing the precoding process of the mm-Wave massive MIMO system, such dense multiple fully connected layers can increase the complexity and time, also the study only optimizes the precoder with fixed combiners. In [18], a deep learning framework was provided for hybrid precoding showing enhancement in spectrum efficiency and bit error rates, but the design of the combiner was not considered and perfect CSI was assumed and the constant modulus constraint was not met. Moreover, some of these works also adopted the separation of channel estimation and precoding which doesn't suit the practical settings.
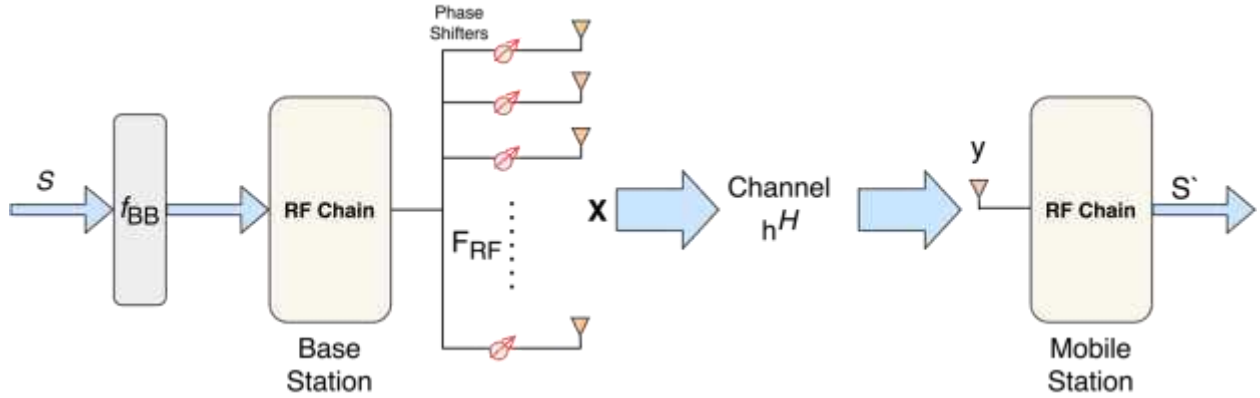
3. Problem Definition and System Model

Figure 1 The architecture of the proposed communication link

In this paper, we examine a single user downlink in a mm-Wave communication system with multiple antennas. We investigate the analog beamformer design with one RF chain. The optimal baseband beamformer has a closed-form solution and can easily be obtained [2], [4]. The output dimension can be increased from N_T to $N_{RF}N_T$ for the $N_T \times N_{RF}$ analog beamformer matrix to investigate the hybrid beamforming problem with multiple RF chains.

Let s be the transmitted symbol vector from the base station (BS) equipped with N_T antennas to a mobile station (MS) equipped with one receive antenna, and let $\mathbb{E}\{|s|^2\} = 1$. Because only one RF chain is considered, the baseband precoder f_{BB} is a scalar. At the transmitter, the symbol vector is multiplied by baseband precoder f_{BB} and then by an $Nt \times 1$ analog precoder F_{RF}. The final transmitted signal can be written as

$x = F_{RF}f_{BB}s,$ (1)

At the receiver, the received signal is given by

$$y = h^H F_{RF} f_{BB} s + n, \quad (2)$$

where n denotes the additive white Gaussian noise (AWGS) and $h^H \in \mathbb{C}^{1 \times N_T}$ is the channel vector between the base station and the user with $\mathbb{E}\{\|h^H\|_F\} = 1 \times N_T$. The widely used Saleh-Valenzuela mm-Wave model is considered in this architecture where N_c clusters of N_{ray} paths are assumed in the channel model [19] [20] [21]. For h^H with one LoS path and (L−1) NLoS paths, which is given by

$$h^H = \sqrt{\frac{N_t}{L}} \sum_{l=1}^{L} \alpha_l a_t^H (\phi_t^l), \quad (3)$$

Where α_l denotes the complex channel gain of the lth path, and the N_T x 1 response vector of the antenna array at the BS is denoted by $a_t(\phi_t^l)$. The departure angle of the the lth path is ϕ_t^l., The optimization objective is the spectral efficiency, for the proposed system, spectral efficiency is

$$R = \log_2 \left(1 + \frac{1}{\sigma^2} \|h^H F_{RF} f_{BB}\|^2 \right), \quad (4)$$

When the constant modulus constraint, $|[F_{RF}]_i|^2 = 1$, for $i = 1, \ldots, N_t$, is considered, in addition to the power constraint $\|F_{RF} f_{BB}\|^2 \leq P$, the optimum f_{BB} for maximizing R is given by $\sqrt{P/N_t}$. Then, the beamforming optimization problem for F_{RF} is given by

$$\begin{aligned} \text{minimize } F_{RF.} \quad & \log_2 \left(1 + \frac{\gamma}{N_t} \|h^H F_{RF}\|^2 \right) \\ \text{subject to} \quad & |[F_{RF}]_i|^2 = 1, \quad for \ i = 1, \ldots, Nt, \quad (5) \end{aligned}$$

where $\gamma = \frac{P}{\sigma^2}$ expresses the signal-to-noise ratio (SNR). It is assumed that $\gamma_{est} = \gamma$, where γ_{est} expresses the estimated SNR since the SNR can be accurately estimated compared to the CSI.

4. Deep Learning-Based Beamforming with Parametric ReLU

The conventional method of replacing the analog beamformer with a multi-layer neural network is not suitable when considering the analog phase shifters architecture [11], [22], and [23]. As specified in Eq (2), the acquired signal is correlated to the analog precoder, thus it can't be used as the input as some existing studies did [11], [22], and [10]. Moreover, since most of the deep learning frameworks don't support complex output, it is difficult to guarantee that the output (F_{RF}) meets the constant modulus constraint. To this end, we introduce a deep learning-based network that takes the above-mentioned issues into consideration. The proposed network takes the SNR estimation γ_{est} and the channel estimation h_{est} as inputs and outputs the optimum analog beamforming F_{RF}. A Lambda layer is implemented to guarantee that the output (F_{RF}) is satisfying the constant modulus constraint. A loss function that is correlated to the objective in Eq (5) is used to train the proposed network. The loss functions can be given by

$$Loss = -\frac{1}{N} \sum_{n=1}^{N} \log_2 \left(1 + \frac{\gamma_n}{N_t} \|h_n^H F_{RF,n}\|^2 \right), \quad (6)$$

Where N expresses the training samples, and h_n, γ_n, and $F_{RF,n}$ expresses the CSI, SNR, and analog beamformer output correlated to the nth sample. Minimizing the loss resembles an increase in spectral efficiency.

In the training stage and based on the system model, channel samples and noise in addition to the transmit pilot symbols are simulated. The channel estimation method in [6] is then applied to estimate CSI. Ultimately, the channel estimation h_{est} and the SNR estimation γ_{est} are fed to the network as the input.

It is assumed $\gamma_{est} = \gamma$ as mentioned in Section 2. Then, the proposed network attempts to minimize the loss function defined in Eq (6) to generate the analog beamforming $F_{RF,n}$. The channel parameters and SNR values (perfect CSI) are simulation-generated, they are used in the computation of the loss function, as shown in Fig. 2. Using the perfect CSI in the computation of the loss function and using the channel estimation as the input of the proposed network enforces it to learn how to achieve the optimum spectral efficiency with perfect CSI and stability against the estimation errors.

The channel estimation is applied again in the deployment stage then fed to the proposed network to generate the optimized beamformer. The perfect CSI is not required in the deployment stage and is only used during the offline stage to calculate the loss function. In the deployment stage, the proposed network parameters are updated and fixed after training.

Since the channel estimation (the input) is complex-valued while the proposed network is a real-value, the imaginary and real domains of the channel estimation are concatenated together and further with γ_{est} to generate a $(2N_T + 1) \times 1$ real-value input vector followed by three hidden layers with 256, 128, 64 neurons in each. Algorithm 1 explains the proposed network workflow.
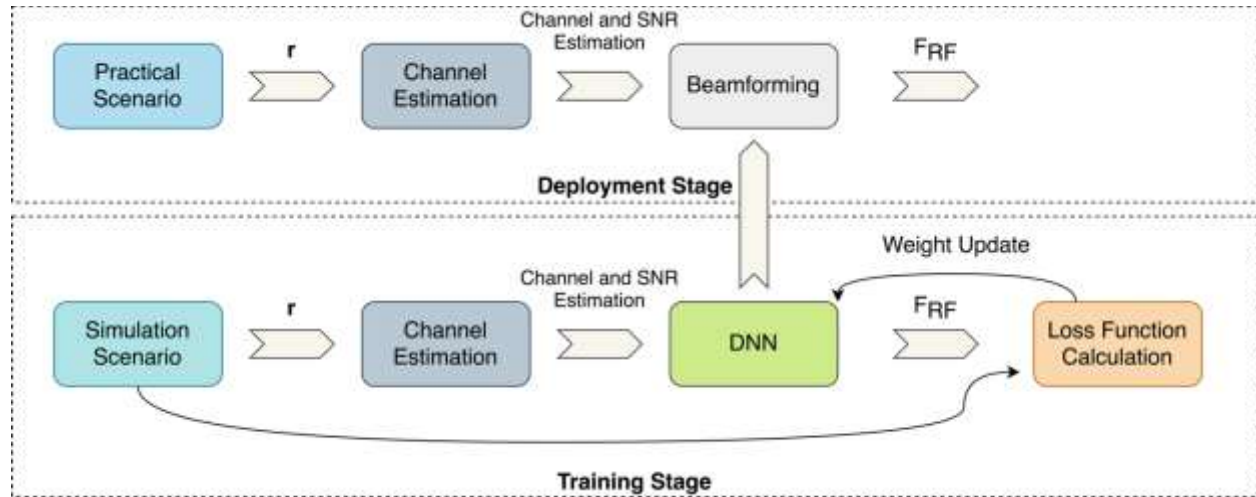

Figure 2 The proposed network stages

Activation functions have a significant role in deep learning networks, it affects its ability to converge and the convergence speed. The neural network's computational efficiency, accuracy, and final output are also determined by the activation function. The activation function is assigned to every neuron and the function decides whether the neuron is activated or not. It is crucial to have computationally efficient activation functions since they are processed for each neuron in the network for every data sample. Furthermore, backpropagation is the most used method for model training which puts more computational effort on the activation function.

The linear activation function derivative is constant and irrelevant to the input which prevents the use of backpropagation to train the model. Furthermore, if linear activation functions are used, the last layer will always be a linear function of the first layer regardless of how many layers are implemented, therefore, when linear activation functions are used, the neural network operates as one layer.

Recent deep learning networks employ non-linear activation functions, which enable the model to create the desired mapping between the inputs and outputs of the network, which are essential for the model to learn complex data. Non-linear functions have a derivative that is relevant to the inputs and therefore allows backpropagation. They allow multiple layers to be stacked to create a deep neural network.

In this paper we used the parametric rectified linear unit (PReLU) function [24], this function overcomes several disadvantages in previous functions. Sigmoid and TanH functions suffered from the vanishing gradient problem where input's very high and low values yield no change to the prediction, outputs not zero centered, and computationally expensive. Rectified linear unit (ReLU) function suffered from the dying ReLU problem, the backpropagation cannot be performed and the network stops learning when its inputs are negative or reach zero. The dying ReLU problem is tackled by the Leaky ReLU function but, still, cannot give consistent predictions for negative inputs.

PReLU doesn't suffer from the dying ReLU problem and can perform well with negative inputs and allow backpropagation, it also presents the negative part as an argument to provide consistent predictions for negative inputs as in Eq (7).

$$f(x_i) = \begin{cases} x_i, & \text{if } x_i > 0 \\ \alpha_i x_i, & \text{otherwise} \end{cases}, \; Eq\ (7)$$

Notice that x_i is only one entry from the vector **x** and the subscript i in α_i allow the nonlinear activation to change on different channels which is known as channel-wise PReLU, if we decide to learn the same α for all features, channel-shared PReLU is used. Accordingly, α defines how the function will perform. We tested different activation functions and found that PReLU achieves the best results. PReLU improves model fitting with no extra complexity as well. The results from [24] show that PReLU enhances the performance of small and large models without adding computational cost.

Considering the online stage only, the proposed network complexity can be calculated as the number of floating-point operations (FLOPs) in all layers given by $(2N_I - 1) N_O$ [25], where N_I and N_O refers to the input and output dimensions respectively. The FLOPs number of the proposed network is about 0.14 million when $N_T = 64$. For comparison, the complexity of conventional HBF algorithms in [3][4] and [5] is about 0.26 million considering the same number of N_T and the complexity of the same network with traditional ReLU is also 0.14 million.

The PReLU activation function is associated with an initializer, regularizer, a constraint, and shared axes option. The initializer allows us to determine how the α weights are initialized, we tested with the default $\alpha = 0$ and $\alpha = 0.25$, the value that authors in [24] recommended. We found that output is better with the default value and considered this value in this study. the regularizer controls weight fluctuations by applying penalties to outliers, authors in [24] suggested that some regulators may force α to certain values and sways PReLU towards ReLU or Leaky ReLU, we didn't use a regularizer. Also, network parameters can be constrained to a fixed limit during training, we didn't use any constraints so the activation function doesn't be monotonic. Shared axes enable sharing axes over space and set to none in the proposed network.

Algorithm 1 DL-based hybrid precoding algorithm using PReLU for massive mm-Wave system
Input: Estimated SNR γ_{est} and Channel estimation h_{est}

Output:
The optimum analog beamforming F_{RF}

---

1. **Initialization:** N_T=64 (number of transmit antennas), P=1(normalized transmit power), a function to transferer the phase to complex-valued analog beamformer, and Lambda layer to compute the rate.
2. Simulate channel samples, noise, and transmit pilot symbols.
3. Produce a series of training sequences.
4. Construct the DL network layers with PReLU activation function.
5. Estimated CSI is used as input and SNRs associated with different samples are generated.
6. Use perfect CSI to calculate the Loss according to Eq (6) to achieve the objective in Eq (5).
7. Train the network to minimize the loss to achieve $F_{RF,n}$ as in Eq (6) (corresponds to the highest SE).
8. Update the weights
9. **Return** The optimum analog beamforming F_{RF}

---

1. Simulation and Results

In our simulation, a uniform array of 64 transmit antennas N_T are deployed at the BS spaced with half-wave. The channel model in [5] is used and the same parameters as those in [18] are applied. The channel estimation algorithm in [6] is employed for obtaining h_{est}. The Adam optimizer and PReLU activation function are used. The network layers, the output shape (dimension), the number of trainable parameters in every layer, and the activation function are listed in Table 1. A batch normalization layer is implemented before every layer to enhance the convergence. Eventually, the Lambda layer ensures that the constant modulus constraint is met at the output. The number of layers and neurons in each layer and the samples used in the offline training stage are set as in [16].

Table 1 The proposed network architecture and hyperparameters

Layer (type)	Output Shap	Activation	# Parameters	Connected to
imerfect_CSI (Input Layer)	[(None, 1, 2, 64)]	N/A	0	
batch_normalization_1	(None, 1, 2, 64)	PReLU	256	imerfect_CSI
flatten_1	(None, 128)	N/A	0	batch_normalization_1
batch_normalization_2	(None, 128)	PReLU	512	flatten_1
dense_1	(None, 256)	N/A	33280	batch_normalization_2
batch_normalization_3	(None, 256)	N/A	1024	dense_1
dense_2	(None, 128)	N/A	33024	batch_normalization_3
dense_3	(None, 64)	N/A	8256	dense_2
perfect_CSI (Input Layer)	[(None, 64)]	N/A	0	

lambda_1	(None, 64)	N/A	0	dense_3
SNR_input (Input Layer)	[(None, 1)]	N/A	0	
Lambda_2	(None, 1)	N/A	0	perfect_CSI, lambda_1, SNR_input

Assuming the number of estimated paths is correct (L_{est}=3), Fig. 3, 4, and 5 show the spectral efficiency (SE) versus signal-to-noise ratio (SNR) with three different estimation levels, defined by three pilot-to-noise power ratios (PNRs), i.e., −20 dB, 0 dB, and 20 dB. The simulation results show that the proposed network utilizing PReLU slightly outperforms the same network using the ReLU activation function. Fig. 6 illustrates the enhancement represented by the value of the difference in spectral efficiency values for both functions with different PNRs.
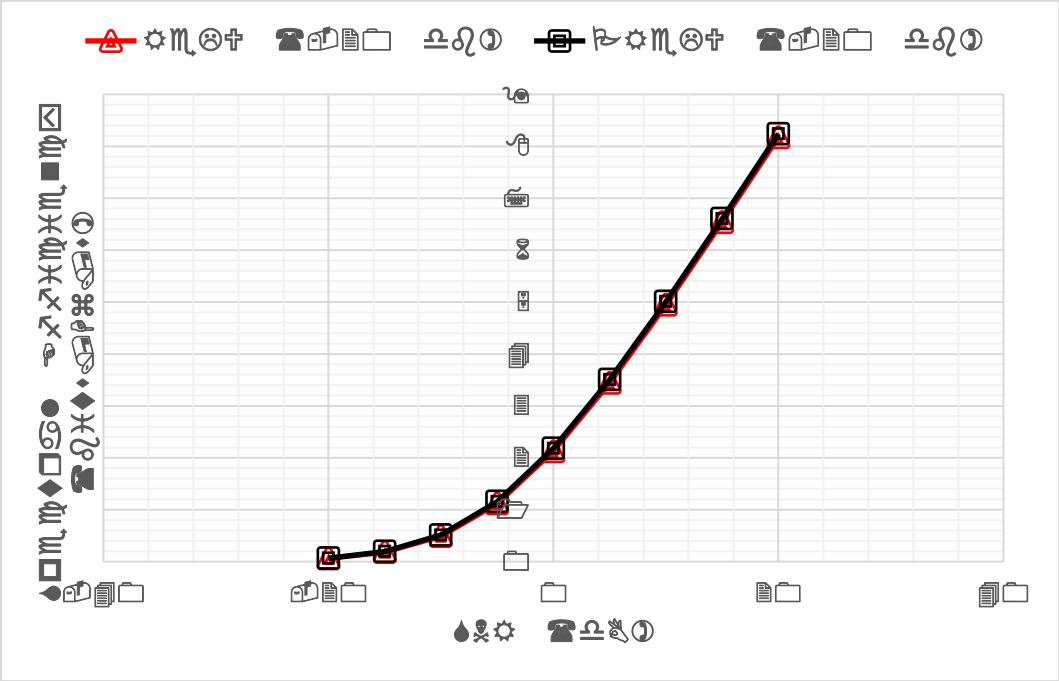

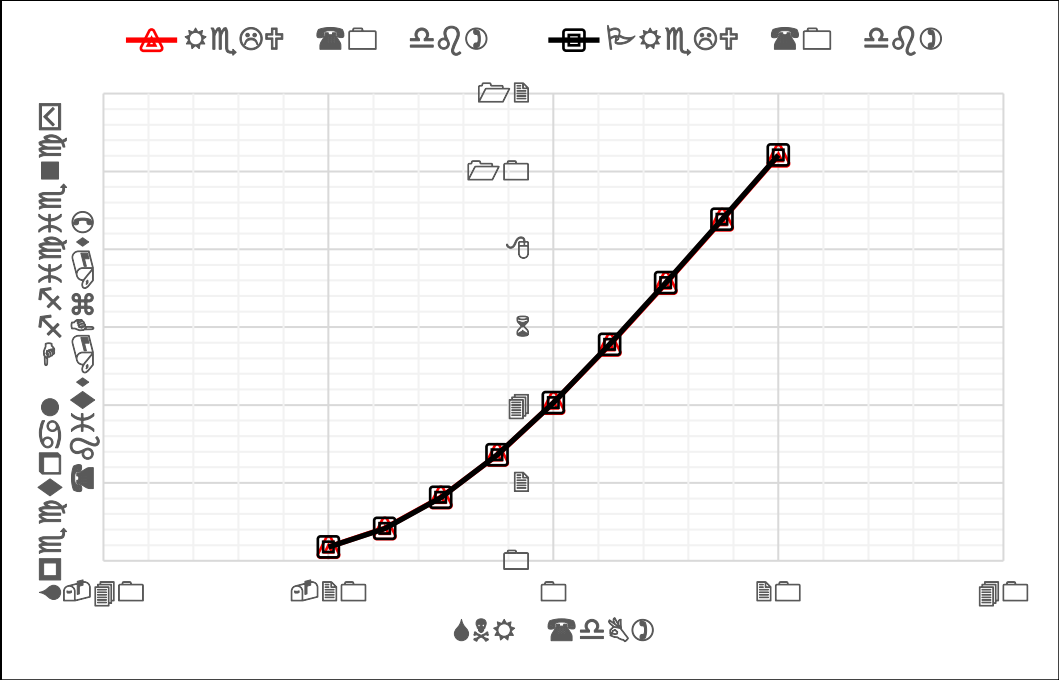*Figure 3 SE vs SNR for PReLU and ReLU when PNR = -20 dB*

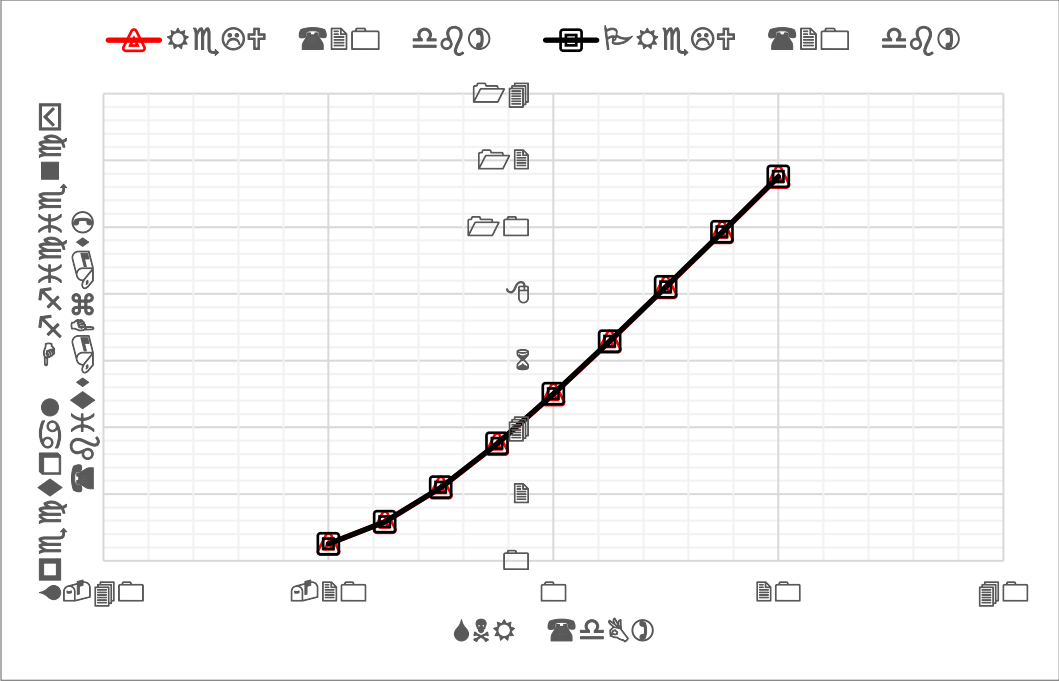*Figure 4 SE vs SNR for PReLU and ReLU when PNR = 0 dB*

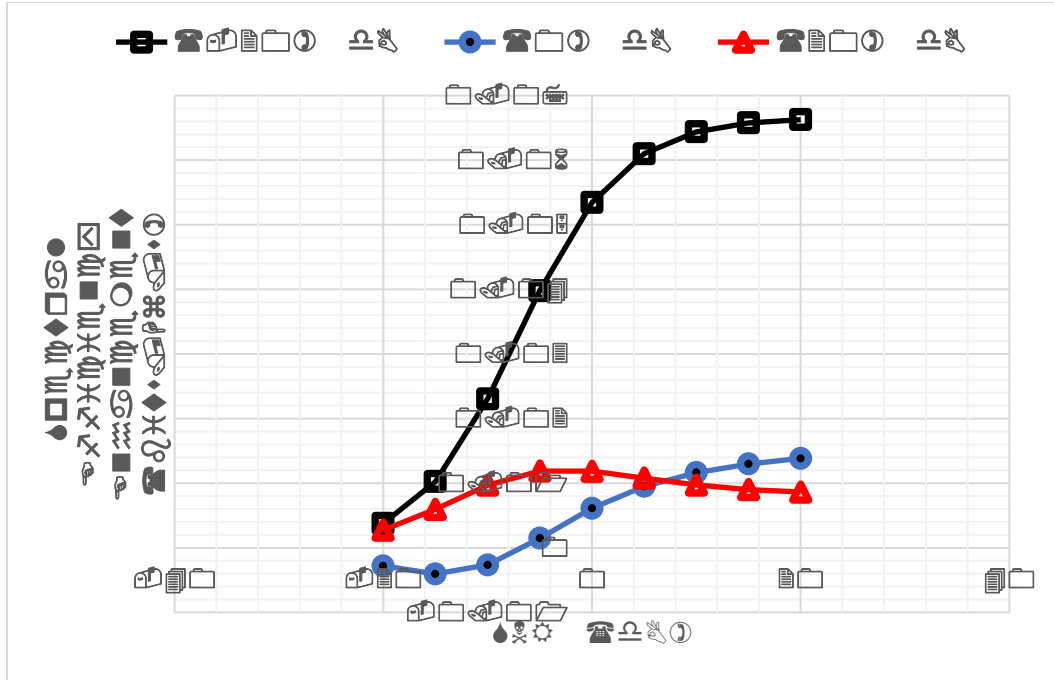*Figure 5 SE vs SNR for PReLU and ReLU when PNR = 20 dB*

Figure 6 SE enhancement of PReLU over ReLU with different PNRs

2. Conclusion

In this paper, it has been shown that the PReLU activation function can enhance spectral efficiency with small margins over the traditional ReLU function. The same hyperparameters, datasets, and architecture were used in both scenarios and only the activation function changed with $\alpha = 0$ and 0.25 and opted for zero value as it showed better results. PReLU seems to be more significant for large datasets than small datasets where the effects of the vanishing gradient and dying ReLU problems can be significant. Therefore, PReLU can be considered for large and deep networks to gain marginal performance improvements over traditional ReLU at no additional complexity.

References

1- A. F. Molisch *et al*., "Hybrid Beamforming for Massive MIMO: A Survey," in *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134-141, Sept. 2017, doi: 10.1109/MCOM.2017.1600400.

2- T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," IEEE Trans. on Commun., vol. 67, no. 5, pp. 3693-3708, May 2019.

3- 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 in Signal Process., vol. 10, no. 3, pp. 485–500, Apr. 2016.

4- 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.

5- O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath et al., "Spatially sparse precoding in millimeter wave MIMO systems," IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

6- A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 831–846, Oct. 2014.

7- I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press Cambridge, 2016.

8- H. Ye, G. Y. Li, and B. Juang, " Power of deep learning for channel estimation and signal detection in OFDM systems," IEEE Wireless Commun. Lett., vol. 7, no. 1, pp. 114–117, Feb. 2018.

9- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.

10- A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," Apr. 2018, arXiv:1804.10334. [Online]. Available: https://arxiv.org/abs/1804.10334

11- H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive MIMO system," IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

12- A. M. Elbir, K. V. Mishra, and Y. C. Eldar, "Cognitive radar antenna selection via deep learning," IET Radar, Sonar Navigat., to be published

13- Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," IEEE J. Sel. Topics Signal Process., vol. 12, no. 2, pp. 340–352, May 2018.

14- D. Yu and L. Deng, "Deep learning and its applications to signal and information processing [exploratory DSP]

15- A. M. Elbir, "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems," in IEEE Communications Letters, vol. 23, no. 7, pp. 1240-1243, July 2019, doi: 10.1109/LCOMM.2019.2915977.

16- Lin, T., & Zhu, Y. (2019). Beamforming design for large-scale antenna arrays using deep learning. IEEE Wireless Communications Letters, 9(1), 103-107.

17- 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.

18- X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6010-6021, Sep. 2017.

19- H. He, C. Wen, S. Jin and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 852-855, Oct. 2018.

20- R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcicy, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in Proc. IEEE Inf. Theory Appl. Workshop, Feb. 2015, pp. 90–97.

21- S. Dörner, S. Cammerer, J. Hoydis, and S. Brink, "Deep learning based communication over the air," IEEE J. Sel. Topics in Signal Process., vol. 12, no. 1, pp. 132–143, Feb. 2018.

22- X. Gao, S. Jin, C. Wen and G. Y. Li, "ComNet: combination of deep Learning and expert knowledge in OFDM receivers," IEEE Commun. Lett., vol. 22, no. 12, pp. 2627-2630, Dec. 2018.

23- C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 748–751, Oct. 2018.

24- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

25- P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," [Online]. Available: https://arxiv.org/abs/1611.06440.