Deep Neural Network based Stable Digital Predistortion with Augmented terms for Switchless Class-G Power Amplifier

Leave Author List blank for your initial paper submission. IWS will be rigorously enforcing the double-blind reviewing requirements. (style: *IWS Author List*)
Leave Affiliation List blank for your initial paper submission. (style: *IWS Affiliation List*)

***Abstract*—This paper proposes a deep neural network based digital predistortion (DPD), which introduces augmented real-valued time-delay terms into the input layer. We use the Exponential Linear Unit (ELU) activation in the hidden layer neurons instead of the Sigmoid, aiming to mitigate the gradient vanishing problem. The proposed model undergoes extensive experiments on a highly nonlinear SLCG power amplifier driven by a wideband signal ranging from 80MHz to 200MHz. The results demonstrate the nonlinear compensation capabilities of ARVTD-DNN are superior to DDR and single-hidden-layer RVTDFNN by approximately 4-5 dB for the adjacent channel leakage ratio and by about 3 dB for the normalized mean square error. Moreover, the proposed model exhibits better robustness with the widening of the driving signal bandwidth.**

***Index Terms*—5G mobile communication, Digital predistortion, artificial neural network, ARVTDNN, exponential linear unit, switchless Class-G power amplifier.**

I. Introduction

In wireless transmitter systems, linearization of power amplifiers has been a crucial problem. In fifth-generation communication systems, the wideband signals often exhibit high Peak-to-Average Power Ratio (PAPR) characteristics. Such signals are more sensitive to nonlinearities of PA, leading to severe distortions and efficiency degradations. Numerous Polynomial Model-based DPDs have been proposed, such as Volterra, MP, DDR[1], etc. In recent years, with the development of Artificial Neural Networks (ANNs), many researchers have turned their attention to incorporating neural networks into DPDs, achieving rich outcomes.

The RVTDFNN[2][3], focusing on the input I/Q signals and their past values, is currently the mainstream foundation for neural network-based DPD. Introducing the envelope of the input signal into the neural network's input, forming the so-called ARVTDNN[4], enhances the generation of richer basis functions compared to RVTDFNN, leading to higher modeling accuracy. Recently, the outstanding performance of deep neural networks (DNNs) in fields like image recognition and classification has attracted the interests of PA modeling researchers. [3]-[5] discussed the effects of using ANNs in DPD, but they mostly utilized shallow NNs, and there are fewer researches on the use of DNNs in DPD. Simultaneously, envelope terms are introduced at the input layer to enhance the modeling accuracy. Experimental results demonstrate that compared to DDR-DPD and single-hidden-layer RVTDFNN-DPD, the proposed model exhibits superior linearization performance and robustness, particularly under wideband signals.

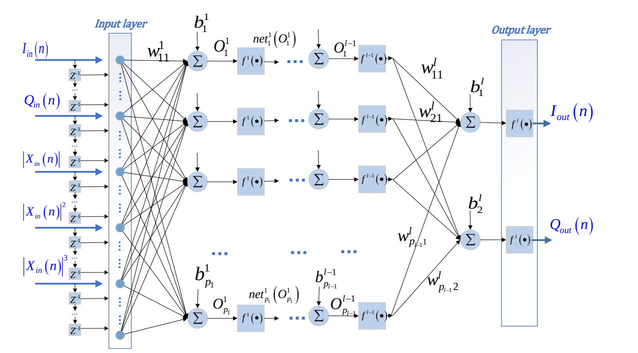

Fig. 1. The proposed augmented real-valued time-delay deep neural network.

II. The Proposed DPD

RVTDFNN is a dynamic NNs proposed to consider memory effects on the basis of a fully connected feedforward NNs. Such a network can effectively simulate power amplifiers (PA) with strong memory effects[3]. ARVTDNN, based on RVTDFNN, introduces the nonlinear version of the envelope term of the input signal at its input. It enables the generation of a richer set of basis functions, better representing the complete behavioral characteristics of the PA, thereby enhancing modeling accuracy[4]. Fig. 2. illustrates the proposed ARVTD-DNN. It uses the Exponential Linear Unit (ELU) activation instead of the Sigmoid activation. The input to this model consists of the present and past values of real-valued orthogonal signals and their envelope terms. The output consists of two real-valued orthogonal signals.

A. ARVTD-DNN

In the proposed model, at any moment in the training of the neural network, the sample sequence at the input is a vector of the order $(2+k)\times(m+1)-by-1$

$$Input=[I_{in}(n),I_{in}(n-1),...,I_{in}(n-m),Q_{in}(n),Q_{in}(n-1),...,Q_{in}(n-m),$$
$$|X_{in}(n)|,|X_{in}(n-1)|,...,|X_{in}(n-m)|,...|X_{in}(n)|^{k},|X_{in}(n-1)|^{k} \tag{1}$$
$$,...,|X_{in}(n-m)|^{k}]$$

TABLE I

BW	Model	Num of Neurons	ACLR (L/U)(dBc)	ACPR (L/U)(dBc)	NMSE (dB)	EVM (%)
100M Hz	w/o DPD		-31.1 / -32.8	-32.6 / -33.9		3.5
	MP		-38.5 / -38.3	-39.3 / -39.5	-30.3	1.19
	RVTDF NN	17	-45.3 / -45.0	-45.2 / -45.6	-35.97	0.85
	DDR		-44.6 / -45.2	-45.0 / -45.4	-36.9	0.94
	PRO-POSED	82	-50.7 / -49.5	-50.8 / -51.4	-39.17	0.6

Where $I_{in}(n)$ and $Q_{in}(n)$ are the input orthogonal signals, and $|X_{in}(n)|$ is the envelope term, k is nonlinear order of the envelope $|X_{in}(n)|$ introduced at input layer. The delayed response is achieved by using $\sum_1^m z^{-n}$ as the delay operator, where Z^{-1} is the unit delay operator that yields its delayed version $I_{in}(n-1)$ when operating on $I_{in}(n)$. The tapped delay lines store values from the previous time step.

Training process of this network is the same as RVTDFNN, where output of an individual neuron of each layer is given by the following steps:

a) The output of the $p-th$ neuron in the $l-th$ layer (I is the number of neurons in the last layer, and w_{ip}^{l-1} is the synaptic weight connecting the $p-th$ neuron in $l-th$ layer and the $i-th$ neuron in the last layer)

$$O_p^l = \sum_{i=1}^{I} w_{ip}^l \cdot O_i^{l-1} + b_p^l \tag{2}$$

Where O_i^0 is the $i-th$ sample of input layer.

b) The output of activation of the $p-th$ neuron in the $l-th$ layer (f represents ELU activation)

$$net_p^l(\bullet) = net_p^l\left(O_p^l\right) = f^l\left(O_p^l\right) \tag{3}$$

c) Calculate the cost function in the output of network

$$E = \frac{1}{2N}\sum_{n=1}^{N}\left(\left(I_{tar}(n) - I_{out}(n)\right)^2 + \left(Q_{tar}(n) - Q_{out}(n)\right)^2\right) \tag{4}$$

Where I_{tar} and Q_{tar} are the target components, while I_{out} and Q_{out} are the output components of the network.

The backpropagation algorithm used in this paper is the 1D Levenberg-Marquardt algorithm[6], aimed at minimizing the cost function. In the forward pass, the cost function is calculated at each iteration. In the backward pass, the LM algorithm is employed to calculate the latest parameters (synaptic weights and biases) that minimize the cost function as much as possible.

B. Activation

In [3-5][8], the activations of neurons in these NNs are Sigmoid activation (See Fig. 2.), which is mathematically equivalent to hyperbolic tangent given as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

However, the use of the Sigmoid in DNNs may lead to the gradient vanishing problem[7]. In [8], the ReLU activation was employed as a replacement for the Sigmoid to address the mentioned problem and achieved better compensation performance. However, in the reverse modeling process of PAs, for the purpose of achieving time alignment between input and output signals (This can be achieved using cross-correlation techniques[9]) and speeding up the training of NNs, the signals used for modeling (both input and output) are Gain-normalized. During testing, we observed a significant issue known as neuron death in DNNs using ReLU as the activation when these normalized data were used. This problem arises because there are a number of negative values in the normalized samples, which causes the ReLU to shut down, leading to neuron death and ultimately causing the network training to fail[10]. In this paper, we introduce the Exponential Linear Unit (ELU)[11] to replace the Sigmoid. ELU overcomes the gradient vanishing problem and avoids the neuron death associated with the use of ReLU as an activation.

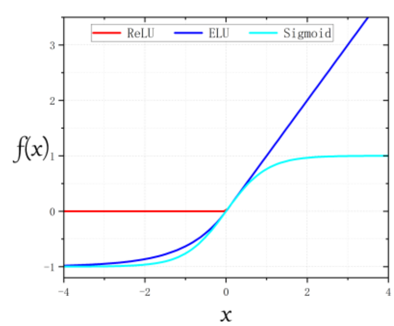

Fig. 2. Activations of Sigmoid, ReLU and ELU

The formula for the ELU activation (See Fig. 2.) is as follows, where α is the hyperparameter whose value is 1 in this paper.

$$f(x) = \begin{cases} x & if \quad x > 0 \\ \alpha\left(e^x - 1\right) & if \quad x < 0 \end{cases} \tag{6}$$

III. Experimental Results & Discussion

For measurement and experimental validation purposes, the proposed ARVTD-DNN-DPD is implemented on the MATLAB platform. The four mentioned DPD methods were compared for linearization performance on a highly nonlinear SLCG PA with memory effects driven by a wideband signal whose saturated output power is 37 dBm and the drain efficiency is 35% [12]. Above four DPD methods adopt an indirect learning structure. Fig. 3. illustrates the measurement set-up and the predistortion implementation process.

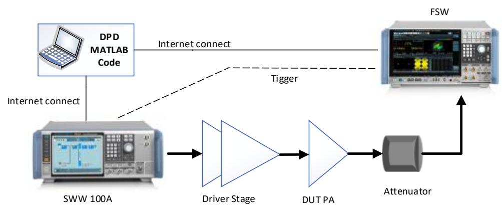

Fig. 3. The measurement setup and the predistortion implementation process.

The RVTDFNN-DPD employs a single hidden layer with 17 neurons, and the number of delay taps is set to 4, according to [10]. The proposed DPD has 4 delay taps, 8 hidden layers (L=8), each with 10 neurons, and uses ELU as the activation. The number of samples for DPD is 9000 in each case. The output layer has two neurons with a linear activation.

The DPD results for a 100MHz signal applied to the SLCG PA are presented in Table I and Fig. 4.

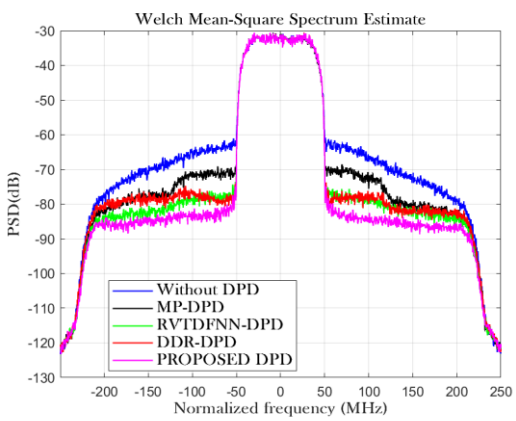

Fig. 4 Power spectral density in each case.

To validate its robustness, the system performance was experimentally measured under a single-carrier 256QAM input signal with a bandwidth of 80MHz to 200MHz (See Fig. 5.). It can be observed that the proposed model, compared to traditional DDR and single-hidden-layer RVTDFNN, exhibits better adaptability to the linearization requirements under large bandwidth signal conditions.

IV. Conclusion

This paper proposes the ARVTD-DNN-DPD model, employing the ELU activation to replace the ReLU activation in DNN to avoid the neuron death. Experimental verification using an SLCG PA demonstrates that the proposed DPD maintains excellent linearization performance and enhanced robustness under a wide bandwidth. Compared to traditional polynomial DDR and single-hidden-layer RVTDFNN, it shows a 4-5dB and 5-6dB improvement in ACLR and ACPR respectively, and about a 3dB improvement in NMSE. When driven by a 140MHz wide bandwidth signal, DDR-DPD loses its linearization capability, while the proposed DPD continues to perform effective linearization.

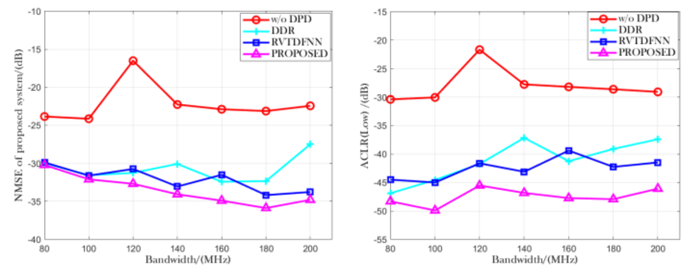

Fig. 5. Performance in NMSE(left) and ACLR(right) v.s. BandWidth.

References

[1] Ghannouchi, Fadhel M. Behavioral Modelling and Predistortion of Wideband Wireless Transmitters. Hoboken: John Wiley & Sons Inc., 2015. Print.

[2] T. Liu, S. Boumaiza, and F. Ghannouchi, “Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks,” IEEE Trans. Microw. Theory Tech., vol. 52, no. 3, pp.1025–1033, Mar. 2004.

[3] Rawat, M., K. Rawat, and F.M. Ghannouchi. “Adaptive Digital Predistortion of Wireless Power Amplifiers/Transmitters Using Dynamic Real-Valued Focused Time-Delay Line Neural Networks.” IEEE transactions on microwave theory and techniques 58.1 (2010): 95–104. Web.

[4] Wang, Dongming et al. “Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters.” IEEE transaction on neural networks and learning systems 30.1 (2019): 242–254. Web.

[5] B. E. Watkins et al., “Neural network based adaptive predistortion for the linearization of nonlinear RF amplifiers,” Military Commun. Conf. 1995, pp. 145-149, San Diego, USA, November 1995.

[6] Hagan, M.T., and M.B. Menhaj. “Training Feedforward Networks with the Marquardt Algorithm.” IEEE transactions on neural networks 5.6 (1994): 989–993. Web.

[7] S. Haykin, Neural Networks: A Comprehensive Foundation. Upper-Saddle River, NJ: Prentice-Hall, 1999.

[8] Hongyo, Reina, Yoshimasa Egashira, and Keiichi Yamaguchi. “Deep Neural Network Based Predistorter with ReLU Activation for Doherty Power Amplifiers.” 2018 Asia-Pacific Microwave Conference (APMC). IEICE, 2018. 959–961. Web.

[9] M. S. O. Alink, E. A. M. Klumperink, M. C. M. Soer, A. B. J. Kokkeler, G. J. M. Smit, and B. Nauta, “A CMOS-compatible spectrum analyzer for cognitive radio exploiting crosscorrelation to improve linearity and noise performance,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 59, no. 3, pp. 479–492, Mar. 2012.

[10] Lu Lu,Yeonjong Shin,Yanhui Su George E. Karniadakis.(2019).Dying ReLU and Initialization: Theory and Numerical Examples..CoRR

[11] Djork-Arné Clevert,Thomas Unterthiner Sepp Hochreiter.(2015).Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)..CoRR

[12] X. Fang, R. Chen and J. Shi, "Switchless Class-G Power Amplifiers: Generic Theory and Design Methodology Using Packaged Transistors," in in IEEE Transactions on Microwave Theory and Techniques, accepted, doi: 10.1109/TMTT.2024.3351852.