Deep Neural Network based Stable Digital Predistortion using ELU Activation 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*—In this paper, a deep neural network (DNN) based digital pre-distortion (DPD) technique is proposed. It utilizes the exponential linear unit function instead of the sigmoid one to activate the hidden layer neurons, consequently mitigating the gradient vanishing problem and accelerating the DNN training. For validation, its linearization effects on a SLCG PA that is excited by a 256QAM signal with various modulation bandwidths (swept from 20MHz to 200MHz) are carefully compared with other existing DPD algorithms. This comparison reveals that, when applied under the signal with a modulation bandwidth of 200MHz, the proposed DPD method can further reduce the adjacent channel power ratio and the normalized mean square error of the PA output signal by 4-8.4dB and 1.3-4.3dB, compared to other polynomial/NN-based DPD methods. This demonstrates the superiority of the proposed technique in linearizing the nonlinear SLCG PA under wideband modulated signal excitations.**

***Index Terms*—Digital pre-distortion, deep neural network, wideband modulation, exponential linear unit, switchless class-G power amplifier.**

I. Introduction

Digital pre-distortion (DPD) is a crucial technique for modern wireless communication systems where the ever-increasing signal modulation bandwidth and Peak-to-Average Power Ratio (PAPR) make transmitting signals extremely sensitive to the nonlinear characteristics of the RF power amplifier (PA). Various DPD methods have been proposed to linearize the PA, maintain signal quality, and improve overall system efficiency. These include many polynomial-based techniques, such as Memory-Polynomial (MP) and Dynamic Deviation Reduction (DDR) [1] techniques. Recently, the concept of artificial Neural Networks (NNs) has been introduced into DPD techniques, achieving excellent outcomes [2]-[5]. Typical NN algorithms include real-valued focused time-delay NNs (RVFTDNN) [2][3] and augmented real-valued time-delay NN (ARVTDNN) [4] where the former utilizes the real-valued I/Q and their delays as the input of NN and the latter further introduces the envelope signal into the input vector.

While the above techniques focused on shallow NN, recently, the success of deep NNs (DNNs) in image recognition and classification has made DNN a promising solution for nonlinear PA linearization [8]. In addition, the innovation in the PA back-off efficiency enhancement (BEE) technique also brings new challenges for PA linearization. Switchless Class-G (SLCG) PA is a recently developed BEE technique that adopts two-quadrant modulation (TQM). Unlike other one-quadrant load-modulated amplifiers (i.e. Doherty PA), its linearity and linearizability under wideband modulated signal excitations have not been discussed or exposed.

This paper presents a deep ARVTDNN DPD method and discusses its application to the SLCG PA. The proposed DNN DPD method utilizes the exponential linear unit (ELU) function instead of the sigmoid one to fulfill the activation of the hidden layer neurons, so as to eliminate the gradient vanishing problem and realize stable DNN training and linearization. Through discussing the nonlinear characteristics of the SLCG PA and comparing the linearization results obtained by applying the proposed and other DPD techniques to the SLCG PA, the advantage of the proposed method can be well demonstrated by its relatively better linearization effect that yields improved adjacent channel power ratio (ACPR) of <-48.5dBc and enhanced normalized mean square error (NSME) of <-40dB over wideband modulation bandwidth range (from 20MHz to 200MHz). This corresponds to ACPR and NSME improvement of 4-8.4dB and 1.3-4.3dB, respectively, when compared to those obtained from other polynomial/NN-based DPD methods.

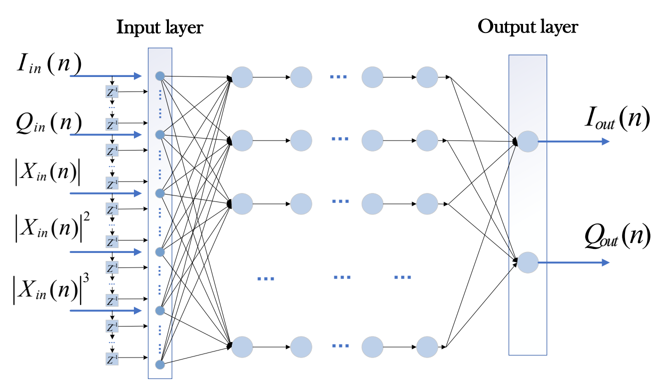

Fig. 1. The proposed augmented real-valued time-delay deep neural network.

II. THE PROPOSED DPD

Fig. 1 illustrates the proposed deep ARVTDNN, where the input signal contains the current and past samples of the Cartesian I/Q components $I_{in}(n)$ and $Q_{in}(n)$ along with the augmented envelope terms $|X_{in}(n)|$:

TABLE I
PERFORMANCE OF DPDs

BW	Model	Num of Neu-rons	Activa-tion	Mean-ACPR (dBc)	NMSE (dB)	EVM (%)
200 MHz	w/o DPD	—	—	-25.8	—	14.4
	ML	—	—	-34.9	-25.9	4.4
	DDR	—	—	-40.4	-35.9	2.0
	Shallow-RVFTDNN	17	Sigmoid	-43.9	-36.6	1.6
	ARVTDNN	17	Sigmoid	-43.9	-38.7	1.9
	Deep-RVFTDNN	122	Sigmoid	-44.8	-38.9	1.8
	PROPOSED	82	ELU	-48.8	-40.2	1.2

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

Where, $X_{in}(n) = I_{in}(n)+j \bullet Q_{in}(n)$ and k is the nonlinear order of $|X_{in}(n)|$, and m is the memory depth of the input vector. The delayed response (see Fig. 1) 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 $I_{in}(n\text{-}1)$ when operating on $I_{in}(n)$.

While the proposed deep ARVTDNN DPD adopts the same forward propagation algorithm as [2]-[5] and uses the Levenberg-Marquart (LM) [6] algorithm for back-propagation, it utilizes different hidden layer activation to enhance the training stability. Conventionally, the sigmoid function (i.e. hyperbolic tangent function) is adopted for the activation in the NN training. However, this leads to the gradient vanishing problem in the DNN scenario [7]. Authors in [8] utilize the Rectified Linear Unit (ReLU) function, i.e. the half-wave rectifying function with the form of $f(x)=max(x,0)$, to replace sigmoid one and address the above problem. However, large amounts of negative valued I/Q signals exist in modern wireless communication systems. These negative-valued I/Q samples fall into the rectification range of the ReLU function and make ReLU inactive, which ultimately causes the failure of the network training [9]. To mitigate this issue, this work introduces the ELU function, which takes the form of (2), to replace the sigmoid one:

$$f(x) = \begin{cases} x & if \ \ x > 0 \\ \alpha\left(e^{x}-1\right) & if \ \ x < 0 \end{cases} \quad (2)$$

Where α is the hyper-parameter whose value is set to 1 in this study. Note that the above equation alters the negative input part of the ReLU function to an exponential function, therefore, the ELU function can remain active for negative input. Meanwhile, it is also capable of producing a sufficiently high gradient under an extremely large positive input, and consequently, it can avoid the gradient vanishing problem. This enables successful and stable linearization.

III. EXPERIMENTAL RESULTS & DISCUSSION

For experimental validation, the proposed DPD technique is applied to linearize a SLCG PA using the indirect learning architecture. The adopted SLCG PA is capable of operating over a frequency band of 1-3GHz and delivering saturation output power of 36.6-38.7 dBm with 6dB back-off and saturation drain efficiency (DE) of 45.4-51.2% and 47.8-57.7%, respectively [10]. Fig. 2. illustrates the results of applying a 7-order memoryless DPD to the SLCG PA under a 64QAM signal with a PAPR of 7.5dB and a modulation bandwidth of 10MHz. Based on Fig.2, before the ML-DPD, the PA can deliver an EVM of 2.64-7.08% and an ACPR of -28.9 to -35.6dBc at an average output power (P_{avg}) of around 30dBm with an average drain efficiency (DE_{avg}) of 33.7-48.6% over 1-3GHz. After applying MLDPD, EVM, and ACPR can be improved to levels of <3% and <-46dBc, respectively. This indicates good linearizability of the adopted SLCG PA in the case of narrowband modulated signal excitations.

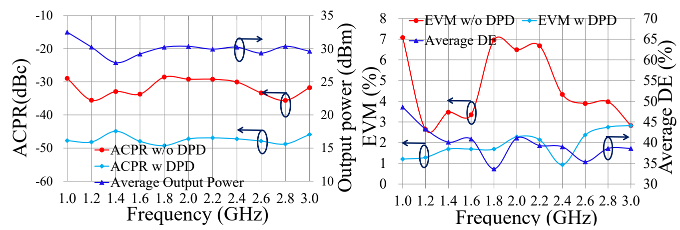

Fig. 2. The modulated signal measurement results of the SLCG PA under 10MHz 64QAM signals.

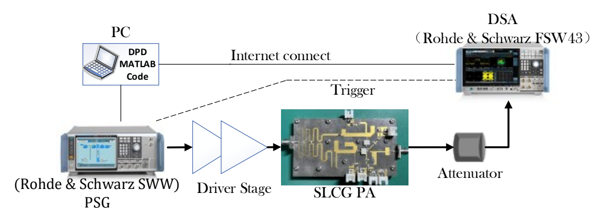

Fig. 3. The DPD measurement setup.

In this measurement, the above SLCG PA was excited with 256QAM signals at the center frequency of 2.4GHz with various modulation bandwidths varying from 20MHz to 200MHz, and further linearized using various DPD methods at a P_{avg} of 30dBm with a DE_{avg} of 38%. Fig. 3 illustrates the diagram of the DPD measurement testbench. The employed DPD methods include the traditional polynomial-based ML and DDR DPD, shallow and deep RVFTDNN DPD, ARVTDNN-DPD, and the proposed DPD techniques. Table I summarizes the configurations and outcomes of the adopted DPD techniques. The shallow RVFTDNN and ARVTNN DPD both employ 17 neurons, where the latter is an extended version of the former with two envelope terms augmented. Moreover, the deep RVFTDNN method adopts 8 hidden layers where each layer consists of 15 neurons and is activated with the sigmoid function, and the proposed DPD is configured to exhibit total neurons of 82 with 8 hidden layers where each contains 10 neurons and the ELU function is used for activation. For all the NN-based DPD, the output layer is set as 2 neurons with linear activations, and *m* (delay tap) and *k* (nonlinear order) are set to 2 and 4, respectively.

Fig. 4 shows the output spectrum of the SLCG PA driven by a 200MHz 256QAM signal with and without DPD. The associated linearization results are also compared in Table I. According to Fig.4 and Table I, compared to existing DPD methods, the proposed one offers a 4-8.4dB improvement in ACPR as well as about 1.3-4.3dB improvement in NMSE. In addition, while deep NN training using the sigmoid function requires a large amount of training time due to the gradient vanishing problem, the ELU function applied in the proposed method can greatly accelerate the training process. Thereby, a stable and reliable linearization effect can be obtained with reasonable training time resources.

Fig. 5 illustrates the DPD results for the signal modulation bandwidth swept from 20MHz to 200MHz. Due to the long training time required by the deep RVFTDNN DPD with the sigmoid function, it was not evaluated in this experiment. Based on Fig. 5, one can find that the effects of all DPD methods degrade with the rise of signal bandwidth. This is mainly due to the enlargement of the memory effect of the SLCG PA under wideband modulated signal, and partly due to the increase of noise floor in wideband sampling. Nevertheless, over the entire range of modulation bandwidth, the proposed method is capable of providing the best linearization that maintains ACPR levels of <-48.5dBc. On the contrary, the ACPR of other methods starts to degrade to levels of >-46dBc for modulation bandwidth of >140MHz. These highlight the advantages of the proposed deep ARVTDNN DPD in linearizing the nonlinear PA excited with wideband modulated signals.

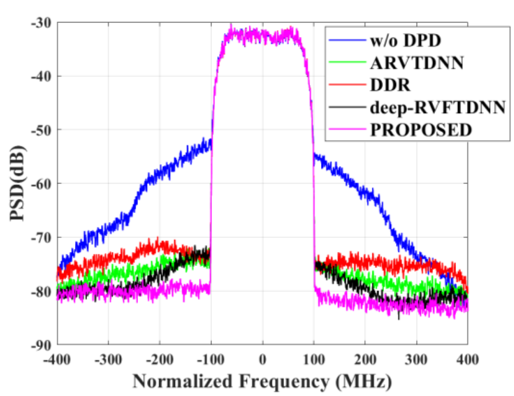

Fig. 4. The PA output power spectral density with and without DPD.

IV. Conclusion

This paper proposes a deep ARVTDNN DPD technique, which employs ARVTD-based input vector and adopts ELU function to replace the sigmoid one in the neuron activation, and consequently avoids the gradient vanishing problem and accelerates the DNN training. Experimental verification conducted on an SLCG PA demonstrates that the proposed DPD maintains excellent linearization performance and enhanced stability in the case of large signal modulation bandwidth. For instance, when exciting the SLCG PA using a 256QAM signal with a modulation bandwidth of 200MHz, compared to other existing DPD, the proposed technique can provide ACPR and NSME improvement of 4-8.4dB and 1.3-4.3dB, respectively.

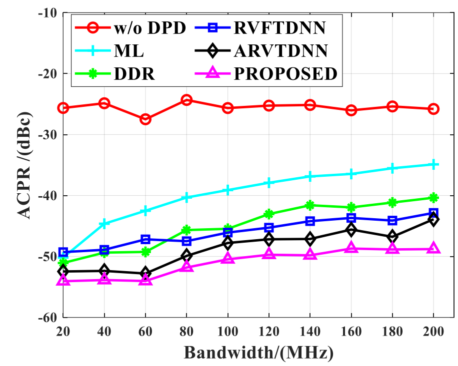

Fig. 5. The DPD measurement results under various modulation bandwidths.

References

[1] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *in IEEE Microwave Magazine*, vol. 10, no. 7, pp. 52-64, Dec. 2009.

[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] M. Rawat, 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 Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 95–104, Jan. 2010.

[4] D. Wang, M. Aziz, M. Helaoui and F. M. Ghannouchi, "Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters," *in IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 242-254, Jan. 2019.

[5] P. Jaraut, M. Helaoui, W. Chen, M. Rawat, N. Boulejfen and F. M. Ghannouchi, "Review of the Neural Network based Digital Predistortion Linearization of Multi-Band/MIMO Transmitters," *2021 IEEE MTT-S International Wireless Symposium (IWS)*, Nanjing, China, 2021, pp. 1-3.

[6] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *in IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993, Nov. 1994.

[7] Haykin, Simon, and N. Network. "A comprehensive foundation." *Neural networks* 2.2004 (2004): 41.

[8] R. Hongyo, Y. Egashira and K. Yamaguchi, "Deep Neural Network Based Predistorter with ReLU Activation for Doherty Power Amplifiers," *2018 Asia-Pacific Microwave Conference (APMC)*, Kyoto, Japan, 2018, pp. 959-961.

[9] Lu, Lu, et al. "Dying relu and initialization: Theory and numerical examples." *arXiv preprint arXiv:1903.06733* (2019).

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