TUDelft

Delft University of Technology

TCN-DPD

Parameter-Efficient Temporal Convolutional Networks for Wideband Digital Predistortion

Duan, Huanqiang; Versluis, Manno; Chen, Qinyu; De Vreede, Leo C.N.; Gao, Chang

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

TCN-DPD: Parameter-Efficient Temporal Convolutional Networks for Wideband Digital Predistortion

Huanqiang Duan[#], Manno Versluis[#], Qinyu Chen[$], Leo C. N. de Vreede[#], Chang Gao[#]

[#]Department of Microelectronics, Delft University of Technology, The Netherlands

[$]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, The Netherlands

chang.gao@tudelft.nl

Abstract — **Digital predistortion (DPD) is essential for mitigating nonlinearity in RF power amplifiers, particularly for wideband applications. This paper presents TCN-DPD, a parameter-efficient architecture based on temporal convolutional networks, integrating noncausal dilated convolutions with optimized activation functions. Evaluated on the `OpenDPD` framework with the `DPA_200MHz` dataset, TCN-DPD achieves simulated ACPRs of -51.58/-49.26 dBc (L/R), EVM of -47.52 dB, and NMSE of -44.61 dB with 500 parameters and maintain superior linearization than prior models down to 200 parameters, making it promising for efficient wideband PA linearization.**

Keywords — **temporal convolutional networks, power amplifiers, digital predistortion, behavioral modeling, dilated convolution.**
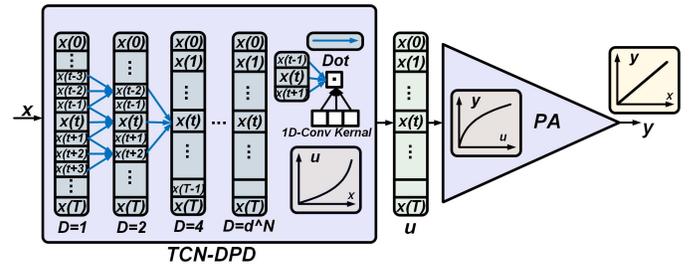
Fig. 1. Basic concept of TCN-DPD with pre-trained PA model for linearization; example with noncausal 1D convolution layers with dilation sizes $D = 1, 2, 4$ up to d^N in the last layer, and kernel size of 3

I. INTRODUCTION

The increasing demand for high-efficiency, wideband communication systems has intensified the need for effective power amplifier (PA) linearization techniques. As modern wireless systems push toward broader bandwidths and higher efficiency requirements, digital predistortion (DPD) has become indispensable for maintaining signal quality while allowing PAs to operate in their efficient nonlinear regions [1].

The Generalized Memory Polynomial (GMP) model has long been the industry standard for PA linearization [2]. However, as communication bandwidths expand and modulation schemes grow more complex, GMP's limitations in modeling sophisticated memory effects have become increasingly apparent [3], [4]. This challenge has driven the exploration of more advanced modeling approaches capable of capturing intricate PA behavior across wider bandwidths.

Deep learning has emerged as a promising solution for next-generation DPD systems. The field evolved from Time Delay Neural Networks (TDNNs) [5] to enhanced variants like Phase-Normalized Real-valued Time Delay (PN-TDNN) [6], followed by various recurrent architectures including LSTM [7], VDLSTM [8], and GRU [9]. Recent innovations incorporate convolutional neural networks (CNNs), such as the RVTDCNN model [10], which simultaneously models spatial and temporal signal characteristics.

Despite these advances, Temporal Convolutional Networks (TCNs) remain largely unexplored in DPD applications, even though they have demonstrated significant advantages in capturing long-range temporal dependencies with reduced computational complexity in time series prediction tasks [11], [12]. Traditional CNN-based approaches for DPD typically require 2D convolutions operating on framed feature maps extracted from in-phase (I) and quadrature (Q) signals, leading to redundant processing of overlapping time segments. In contrast, TCNs employ 1D convolutions directly on the I/Q time series, combined with dilated convolutional layers that efficiently capture long-term memory effects, as illustrated in Fig. 1. This architectural difference enables TCNs to achieve more parameter-efficient wideband DPD while maintaining or improving modeling accuracy. Importantly, the reduction in model parameters directly translates to lower computational complexity, reduced memory requirements, and ultimately decreased power consumption in practical implementations, a critical consideration for battery-operated devices and energy-efficient base stations.

This paper proposes TCN-DPD, designed by using noncausal dilated depthwise separable convolutional layers and a residual neural network architecture to achieve superior linearization performance using fewer parameters. As illustrated in Fig. 1, the proposed system pairs a TCN-DPD with a pre-trained PA model.

II. TCN-DPD ARCHITECTURE

A. Noncausal Convolution

Traditional TCN implementations rely on causal convolution, where network predictions are based exclusively on past data [11], [12], [13]. This design philosophy, inherited from TDNNs, ensures that outputs at time step t depend solely on data from t and earlier timesteps. Such causal systems require leading zero-padding along the time dimension at each layer of TCN to maintain temporal alignment, as defined by Eq. 1:

$$\text{Causal Padding Size} = K - 1 \tag{1}$$

where K represents the convolutional kernel size. While one-dimensional fully-convolutional networks (FCN) can partially address this challenge by maintaining consistent input-output sequence lengths [14], they still demand significant computational resources. Our approach implements noncausal convolution, which reduces the required padding size according to Eq. 2:

$$\text{Noncausal Padding Size} = \frac{K-1}{2} \quad (2)$$

This noncausal design enables the TCN to capture both past and future temporal dependencies within each I/Q data context window of the convolutional kernel size, thereby improving DPD modeling accuracy.

B. Dilated Convolutions

Sequential tasks requiring extensive historical context often challenge standard convolutional architectures, which struggle to capture long-range temporal dependencies [13]. The TCN-DPD architecture addresses this limitation by incorporating dilated convolutions, building upon established approaches [15], [16]. These dilated convolutions expand the network's receptive field exponentially without requiring additional layers or larger kernels. By progressively increasing dilation sizes across layers, the network efficiently processes extended temporal sequences while maintaining parameter efficiency. Each input within the effective history receives processing from at least one kernel, ensuring comprehensive modeling of long-term dependencies. Fig. 1 demonstrates how this dilated structure systematically expands the receptive field across network layers.

C. Proposed DPD Network

The TCN-DPD architecture, illustrated in Fig. 2, comprises three main components: input and output 1×1 convolution layers sandwiching a core Depthwise Separable Convolution Block, with activation functions (Act. Funcs.) connecting each layer to enable nonlinear modeling.

1) 1×1 Convolution

The network employs 1×1 convolution layers at its input and output boundaries. These layers serve as dimensional adaptation interfaces, efficiently transforming feature representations between the network's internal processing stages.

2) Depthwise Separable Convolution Block

At the architecture's core lies a Depthwise Separable Convolution Block featuring multiple depthwise separable convolution layers [17]. Each layer implements progressively larger dilation sizes, enabling efficient capture of temporal information through an expanding receptive field. The padding size for these dilated layers follows Eq. 3:

$$\text{Dilated Noncausal Padding Size} = \frac{K-1}{2} \times D \quad (3)$$

where D represents the dilation size and K denotes the kernel size. Each layer incorporates an activation function to model complex nonlinear signal patterns.
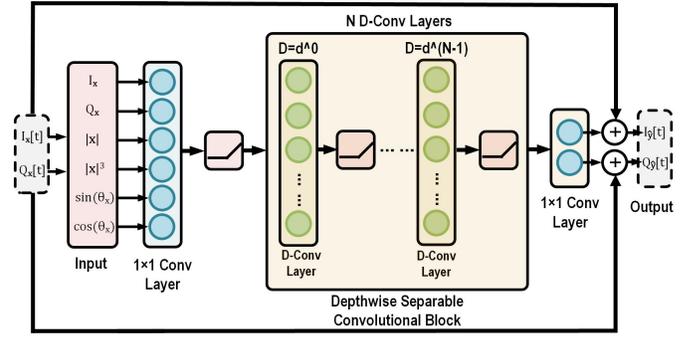

Fig. 2. The residual TCN architecture with depthwise separable convolutional (D-Conv) layers. $D = d^{N-1}$, where D is the dilation size of each D-Conv layer, d is the dilation base, and N is the number of D-Conv layers.

3) Input Features Extracted from I/Q

The network processes both I and Q, along with their derived features, as illustrated in the input layer shown in Fig. 2. This enriched input representation provides comprehensive amplitude and phase information.

4) Residual Connections

The architecture implements residual connections directly linking I and Q inputs to the output, enhancing the model stability in training. These connections preserve critical input information while addressing the vanishing gradient challenges common in deep architectures.

III. EXPERIMENTAL RESULTS

A. Experimental Setup and Benchmarking

All experiments used the `DPA_200MHz` dataset in `OpenDPD` [18], containing 200 MHz bandwidth signals (10 channels × 20 MHz) with 64-QAM OFDM modulation from a 40 nm digital transmitter [19]. For consistent evaluation, we employed a fixed pre-trained DGRU PA model achieving a SIM-NMSE of -31.84 dB, with approximately 500 real-valued parameters tested across 5 random seeds. While simulation and measurement results may differ, previous research [18] showed consistent relative performance rankings between the two approaches. Hardware validation remains essential for final confirmation and is ongoing.

1) Benchmark of Activation Functions

After establishing the core TCN architecture, we systematically investigated activation functions to optimize network performance. Traditional TCN implementations often employ ReLU or its variants [11], [13]. However, these functions present two significant limitations for DPD applications: their inability to process negative values effectively and their non-smooth characteristics, which can introduce spectral leakage. These limitations typically require additional network parameters to model the target function accurately.

To identify the most effective activation function, we evaluated 22 different options using a baseline TCN configuration with 4 depthwise convolution layers, kernel size $K = 5$, and dilation base $d = 2$. Each activation function

Table 1. DPD Performance Comparison of 22 Act. Funcs. Based on a Fixed DGRU PA Pre-trained Model on `DPA_200MHz` Validation Set, Averaged Over 5 Random Seeds

ID	1	2	3	4	5	**6**	7	8	9	10	11
Act. Func.	CELU	ELU	GELU	Hardshrink	Hardtanh	**Hardswish**	LeakyReLU	LogSigmoid	Mish	ReLU	ReLU6
SIM-NMSE (dB)	-43.62	-43.62	-44.79	-11.10	-42.00	**-44.89**	-38.20	-35.62	-44.33	-36.74	-36.77
SIM-ACLR (dBc)	-50.21	-50.20	-51.52	-32.92	-48.00	**-51.54**	-44.16	-40.88	-51.17	-42.53	-42.53
ID	12	13	14	15	16	17	18	19	20	21	22
Act. Func.	RReLU	SELU	SiLU	Softplus	Softshrink	Softsign	Tanh	Tanhshrink	Hardsigmoid	Sigmoid	PReLU
SIM-NMSE (dB)	-37.25	-41.08	-44.70	-38.59	-11.10	-43.65	-44.58	-11.10	-38.05	-35.62	-41.70
SIM-ACLR (dBc)	-43.06	-47.53	-51.54	-44.02	-32.92	-50.14	-51.37	-32.92	-44.88	-40.88	-47.98

[a] TCN architecture set up with 4 depthwise convolution layers, kernel size $K = 5$, and a predefined dilation base $d = 2$.
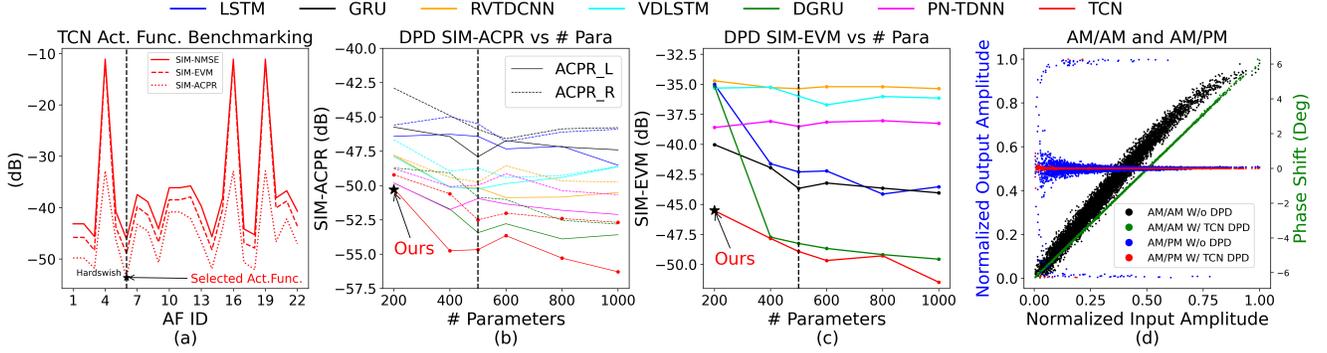

Fig. 3. Training simulation results on 200 MHz 10-channel × 20 MHz OFDM signals from the `DPA_200MHz` validation set. Each curve represents the best performance of each architecture over 5 random seeds with a fixed DGRU PA model. (a) DPD performance comparison of 22 Act.Funcs. Based on a fixed DGRU PA model. (b) The 500-parameter DPD learning SIM-ACPR_L/R over training epochs. (c) SIM EVM vs. real-valued model parameters. (d) The 500-parameter DPD modeling AM/AM and AM/PM plot.
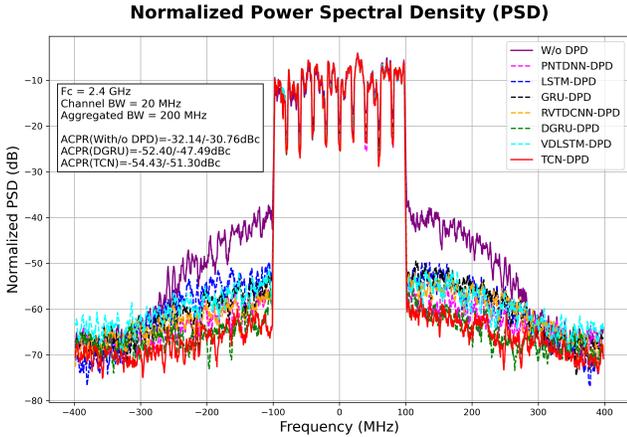

Fig. 4. PSD comparison of the output signal with different 500 real-valued parameter DNN-DPD models on a 200 MHz signal from the `DPA_200MHz` test set. Each curve represents the best DPD performance of each model over 5 random seeds based on a fixed DGRU PA model.
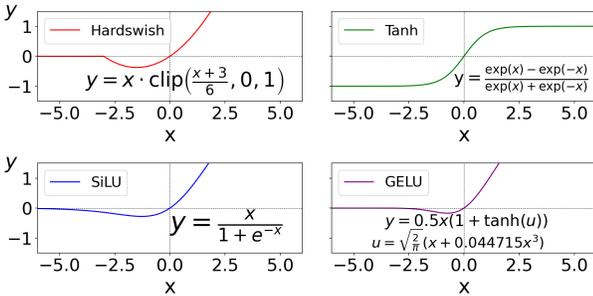

Fig. 5. The Top-4 performing activation functions in Table 1.

Table 2. Performance Comparison of DPD Models Based on a Fixed DGRU PA Model with Approximately 500 Real-Valued Parameters on `DPA_200MHz` Test Set, Averaged Over 5 Random Seeds ± Standard Deviations

Classes	DPD Models	SIM-NMSE (dB)	SIM-ACPR (dBc, L/R)	SIM-EVM (dB)
W/o DPD[a]	-	-	-31.90±0.24 / -30.45±0.31	-34.02±1.42
Base RNN	LSTM	-35.22±3.86	-43.60±1.14 / -42.68±0.36	-37.52±5.11
	GRU	-40.01±1.68	-44.95±0.65 / -43.76±1.60	-42.70±2.23
Prior DPD	RVTDCNN [10]	-32.03±1.19	-48.04±0.71 / -46.26±1.27	-34.61±1.81
	VDLSTM [8]	-32.50±0.71	-47.04±1.44 / -45.85±1.32	-34.94±1.54
	PN-TDNN [6]	-35.49±0.47	-49.25±0.64 / -48.43±1.11	-37.70±0.89
	DGRU [18]	-41.82±2.87	-50.57±1.82 / -49.16±1.67	-44.04±2.38
This Work	**TCN-500**	**-44.61±1.37**	**-51.58±2.84 / -49.26±2.04**	**-47.52±1.49**
	TCN-200	-41.27±1.55	-45.83±2.39 / -46.76±1.23	-43.81±1.67
	TCN-1000	-46.37±1.13	-52.58±2.43 / -50.84±1.44	-49.40±1.90

[a] Based on a fixed DGRU pre-trained PA model over 5 random seeds, the average SIM-NMSE is -31.84 dB.

was tested across 5 random seeds, with results presented in Fig. 3(a). To ensure reliability, we computed averaged performance metrics across all seeds, as detailed in Table 1.

2) DPD Benchmarking

Following activation function optimization, we conducted comprehensive performance benchmarking of the TCN-DPD architecture. We evaluated configurations ranging from 200 to 1000 parameters against other state-of-the-art DNN-DPD models. Each configuration underwent testing across 5 random seeds to ensure statistical significance. The best-performing models for each parameter count are presented in Fig. 3(b) and (c). In contrast, Table 2 provides detailed performance metrics for the 500-parameter configuration, averaged across all random seeds to demonstrate robustness.

B. Results and Discussion

Our analysis reveals insights regarding activation function selection and model performance. The evaluation in Fig. 3(a) and Table 1 demonstrates that Hardswish consistently achieves superior performance across all metrics. Three other functions, GELU, SiLU, and Tanh, also perform strongly. As shown in Fig. 5, these functions share key characteristics: differentiability across their domains and smooth gradient transitions near $x = 0$, making them well-suited for DPD applications.

The comprehensive performance evaluation of the TCN-DPD architecture, illustrated in Fig. 3(b) and (c), demonstrates exceptional modeling capabilities across different parameter configurations. Using SIM-ACPR and SIM-EVM as primary metrics, our TCN-based model consistently outperforms existing architectures across parameter ranges from 200 to 1000. Most notably, with just 200 parameters, the TCN achieves remarkable performance with average ACPRs (L/R) of -45.83 and -46.76 dBc and an exceptional average EVM of -43.81 dB. This achievement is particularly significant as competing architectures struggle to maintain comparable EVM performance at such low parameter counts, highlighting the TCN's potential as a highly efficient DPD solution.

The performance of TCN-DPD extends across all parameter configurations, especially with advantages at lower parameter counts. This consistent performance underscores the TCN architecture's fundamental efficiency in capturing PA nonlinearities while maintaining parameter efficiency. Table 2 presents a detailed performance comparison at the 500-parameter configuration point, where our architecture demonstrates clear advantages over existing approaches. The AM/AM and AM/PM characteristics plotted in Fig. 3(d), derived from the best-performing model over 5 random seeds, further validate the TCN's modeling accuracy. The power spectral density (PSD) comparison in Fig. 4 provides clear visual evidence of the TCN's superior linearization capabilities.

IV. Conclusion

This paper introduces TCN-DPD, a novel parameter-efficient architecture for wideband power amplifier linearization. By integrating noncausal dilated convolutions with optimized activation functions, our approach achieves superior simulated ACPR, EVM, and NMSE metrics while using significantly fewer parameters than prior solutions. Our ongoing work focuses on hardware implementation to verify real-world performance, with results to be reported in a future publication.

Acknowledgment

References

[1] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband rf power amplifiers," *IEEE Microwave Magazine*, vol. 15, no. 7, pp. 84–99, 2014.

[2] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[3] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for doherty power amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 146–148, 2019.

[4] X. Lu, Q. Zhou, L. Zhu, Z. Wei, Y. Wu, Z. Liu, and Z. Chen, "A low-computational-complexity digital predistortion model for wideband power amplifier," *Sensors (Basel, Switzerland)*, vol. 24, no. 21, p. 6941, 2024.

[5] 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 Transactions on Microwave Theory and Techniques*, vol. 58, no. 1, pp. 95–104, 2010.

[6] A. Fischer-Bühner, L. Anttila, M. D. Gomony, and M. Valkama, "Phase-normalized neural network for linearization of rf power amplifiers," *IEEE Microwave and Wireless Technology Letters*, 2023.

[7] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[8] H. Li, Y. Zhang, G. Li, and F. Liu, "Vector decomposed long short-term memory model for behavioral modeling and digital predistortion for wideband rf power amplifiers," *IEEE Access*, vol. 8, pp. 63 780–63 789, 2020.

[9] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[10] X. Hu, Z. Liu, X. Yu, Y. Zhao, W. Chen, B. Hu, X. Du, X. Li, M. Helaoui, W. Wang *et al.*, "Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3923–3937, 2021.

[11] A. Pandey and D. Wang, "Tcnn: Temporal convolutional neural network for real-time speech enhancement in the time domain," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6875–6879.

[12] Y. Chen, Y. Kang, Y. Chen, and Z. Wang, "Probabilistic forecasting with temporal convolutional neural network," *Neurocomputing*, vol. 399, pp. 491–501, 2020.

[13] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[15] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu *et al.*, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, vol. 12, 2016.

[16] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations*, 2016.

[17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[18] Y. Wu, G. D. Singh, M. Beikmirza, L. C. De Vreede, M. Alavi, and C. Gao, "Opendpd: An open-source end-to-end learning & benchmarking framework for wideband power amplifier modeling and digital pre-distortion," in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2024, pp. 1–5.

[19] M. Beikmirza, Y. Shen, L. C. N. de Vreede, and M. S. Alavi, "A wideband energy-efficient multi-mode cmos digital transmitter," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 3, pp. 677–690, 2023.