OpenDPD: An Open-Source End-to-End Learning & Benchmarking Framework for Wideband Power Amplifier Modeling and Digital Pre-Distortion

Yizhuo Wu, Gagan Deep Singh, Mohammadreza Beikmirza, Leo C. N. de Vreede,
Morteza Alavi and Chang Gao*

Department of Microelectronics, Delft University of Technology, Delft, The Netherlands

Abstract—With the rise in communication capacity, deep neural networks (DNN) for digital pre-distortion (DPD) to correct non-linearity in wideband power amplifiers (PAs) have become prominent. Yet, there is a void in open-source and measurement-setup-independent platforms for fast DPD exploration and objective DPD model comparison. This paper presents an open-source framework, `OpenDPD`, crafted in `PyTorch`, with an associated dataset for PA modeling and DPD learning. We introduce a Dense Gated Recurrent Unit (DGRU)-DPD, trained via a novel end-to-end learning architecture, outperforming previous DPD models on a digital PA (DPA) in the new digital transmitter (DTX) architecture with unconventional transfer characteristics compared to analog PAs. Measurements show our DGRU-DPD achieves an ACPR of -44.69/-44.47 dBc and an EVM of -35.22 dB for 200 MHz OFDM signals. `OpenDPD` code, datasets and documentation are publicly available at **https://github.com/lab-emi/OpenDPD**.

Index Terms—digital pre-distortion, behavioral modeling, deep neural network, power amplifier, digital transmitter

I. INTRODUCTION

In the evolution of 6G and Wi-Fi 7 wireless systems and beyond, the need for enhanced Digital Pre-distortion (**DPD**) is underscored by growing communication capacities and data accuracy demands [1]. DPD mitigates the non-linearity of Power Amplifiers (**PA**) by deriving its inverse function. While traditional methods like the general memory polynomial (**GMP**) [2] are popular in the industry, the increasingly wider signal bandwidths introduce more intricate non-linearity and memory effects in the PA, presenting challenges to conventional linearization techniques.

Deep learning has catalyzed significant advances in DNN-based DPD for wideband PAs. The Time Delay Neural Network (**TDNN**) was the earliest DNN architecture explored to model PA's memory effects [3]–[6]. Gated RNNs, notably LSTM [7] and GRU [8], excel in sequence modeling for state-of-the-art DPDs. Integrating vector decomposition (**VD**) with LSTM led to VDLSTM [9], while Phase-Gated JANET [10] enhanced linearization by combining an amplitude-phase extractor with JANET [11]. The DVR technique with JANET, explored in [12], has further improved linearization for signals up to 200 MHz. Meanwhile, CNNs have been used to analyze PA signals as images to boost linearization [13]. However, the improvements in DPD models entail greater computational demand and model expansion.

In many deep learning fields, standardized datasets direct research. For example, MNIST [14] is fundamental for image classification, while Penn Treebank [15] is pivotal in language modeling. Yet, the DPD field presents inconsistent experimental setups and diverse learning methodologies. This variability complicates objective evaluations of DPD models and emphasizes the need for a benchmark similar to MNIST.

Regarding the discrepancy in DPD learning architectures, traditional methods present notable challenges. The Indirect Learning Architecture (**ILA**) hinges on the assumption of commutability of PAs [16]. Conversely, the Direct Learning Architecture (**DLA**) faces difficulty obtaining the ideal pre-distorted PA input [17]. Although the Iterative Learning Control (**ILC**) addresses the challenge of pinpointing ideal values within DLA, it introduces added complexity through the necessity of multiple measurement iterations [18]. Considering experimental setups, prior studies use varied PAs. Although RF-WebLab [19] is publicly accessible, it focuses on analog PA. The emerging digital transmitter (**DTX**) [20]–[23] architecture remains underexplored despite its unique transfer characteristics [24].

This work presents the main contributions below:

1) We present `OpenDPD`, an open-source, end-to-end (**E2E**) learning framework crafted in `PyTorch` [25], streamlining the prototyping of non-linearity correction methods and establishes a standardized benchmark for DNN-based DPD models.

2) The `OpenDPD` platform offers a public dataset with I/Q modulated signals from a digital PA (**DPA**) in a DTX. Additionally, `OpenDPD` features functions to evaluate simulated (**SIM**)-ACPR and SIM-EVM using a provided pre-trained PA model. This not only simplifies DPD design exploration but also ensures consistent testing conditions, enabling researchers to make objective comparisons across DPD studies.

3) We present the Dense GRU (**DGRU**) for PA modeling and DPD, surpassing the performance of LSTM, GRU, and prior works in DTX non-linearity correction regarding the measured Adjacent Channel Power Ratio (**ACPR**) and Error Vector Magnitude (**EVM**).

4) `OpenDPD` is designed to be extensible, with plans to encompass datasets from diverse PAs, and we cordially invite contributions from various research disciplines.
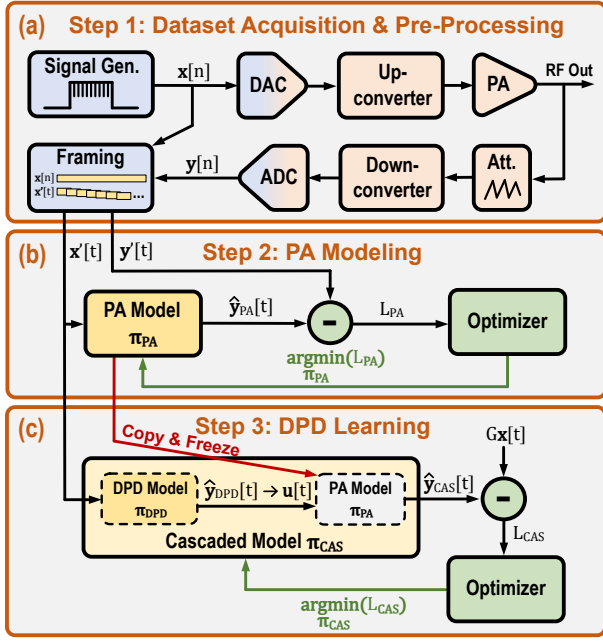
*Corresponding author: Chang Gao (chang.gao@tudelft.nl)

Fig. 1. Generalized E2E learning architecture. With a DPA, an RF-DAC will be used instead of the DAC, Up-converter, and analog PA.

Fig. 2. The DGRU-DPD architecture.

Fig. 3. Setup for dataset acquisition and DPD performance measurement.

II. THE OPENDPD FRAMEWORK

A. E2E Learning

We introduce the E2E learning architecture that uses back-propagation to train a DPD model with a pre-trained PA behavioral model, eliminating the assumption of a commutable PA in ILA and avoiding the manual determination of optimal pre-distorted PA inputs found in DLA. The idea of backprop-agation through a DNN-based PA model is inspired by [26]. However, previous studies did not evaluate this on wideband PA nor report concrete measurements in ACPR and EVM.

As depicted in Fig. 1, the E2E learning architecture consists of three primary steps:

1) **Data Acquisition & Pre-Processing (Fig. 1a):** The baseband I/Q signal $\mathbf{X} = \{\mathbf{x}[n]|\mathbf{x}[n] = I_\mathbf{x}[n] + jQ_\mathbf{x}[n], I_\mathbf{x}[n], Q_\mathbf{x}[n] \in \mathbb{R}, n \in 0,\ldots,N-1\}$ is sent to the PA. The output signals (labels), $\mathbf{Y} = \{\mathbf{y}[n]|\mathbf{y}[n] = I_\mathbf{y}[n] + jQ_\mathbf{y}[n], I_\mathbf{y}[n], Q_\mathbf{y}[n] \in \mathbb{R}, n \in 0,\ldots,N-1\}$, are sourced from the PA under test. To address gradient vanishing issues and enhance training, feature and label sequences are split into shorter frames, $\mathbf{X}'$ and $\mathbf{Y}'$, each of size T, with a stride of S. Consecutive frames overlap by $T-S$ samples.

2) **PA Modeling (Fig. 1b):** Utilizing framed inputs $\mathbf{x}'[t]$ and framed target outputs $\mathbf{y}'[t]$, a behavioral PA model, π_{PA}, is trained in a sequence-to-sequence learning way via BackPropagation Through Time (**BPTT**). The objective is to minimize the Mean Squared Error (**MSE**) loss function given by $L_{PA} = \frac{1}{B \times T} \sum_{b=1}^{B} \sum_{t=0}^{T-1} (\hat{\mathbf{y}}_{PA}[t] - \mathbf{y}'[t])^2$, where B is the batch size in the mini-batched stochastic gradient descent and $\hat{\mathbf{y}}_{PA}[t]$ is the prediction output of the PA model.
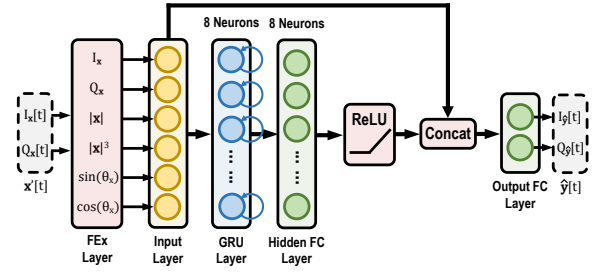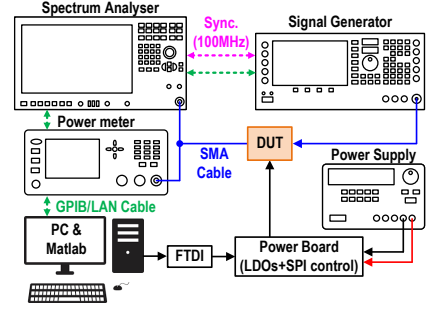
3) **DPD Learning (Fig. 1c):** Before the pre-trained PA behavioral model π_{PA}, a DPD model, π_{DPD}, is instantiated and its output $\hat{\mathbf{y}}_{DPD}[t]$ is fed to the input of π_{PA} to form a cascaded model π_{CAS}. In this configuration, the parameters of π_{PA} remain unaltered. The learning process involves feeding the framed inputs $\mathbf{x}'[t]$ through π_{CAS} and executing BPTT across both π_{PA} and π_{DPD} to minimize $L_{CAS} = \frac{1}{B \times T} \sum_{b=1}^{B} \sum_{n=0}^{T-1} (\hat{\mathbf{y}}_{CAS}[t] - G\mathbf{x}[t])^2$, where G is the target gain of the cascaded DPD-PA system.

After the learning process, $\hat{\mathbf{y}}_{DPD}$ converges to the ideal PA inputs $\mathbf{U} = \{\mathbf{u}[n]|\mathbf{u}[n] = I_\mathbf{u}[n] + jQ_\mathbf{u}[n], I_\mathbf{u}[n], Q_\mathbf{u}[n] \in \mathbb{R}, n \in 0,\ldots,N-1\}$.

B. Benchmarking

OpenDPD has a standardized dataset and benchmarks, including SIMulated (**SIM**)-ACPR and SIM-EVM, for fair comparison of DPD models across different machine learning approaches, enhancing quick DPD prototyping.

1) Dataset: The DPA_200MHz dataset uses 200 MHz 10-channel $\times$ 20 MHz I/Q modulated Orthogonal Frequency Division Multiplexing (**OFDM**) signals, with each channel using 64-QAM and 64 subcarriers. It is partitioned into training (60%), validation (20%), and test (20%) sets, containing the DPA's input-output signals, as captured in Fig. 3.

2) Comparison between DPD Models: OpenDPD users should train the PA behavioral and DPD models using the training set, optimizing hyperparameters like learning rate and batch size with the validation set, keeping test data unused.
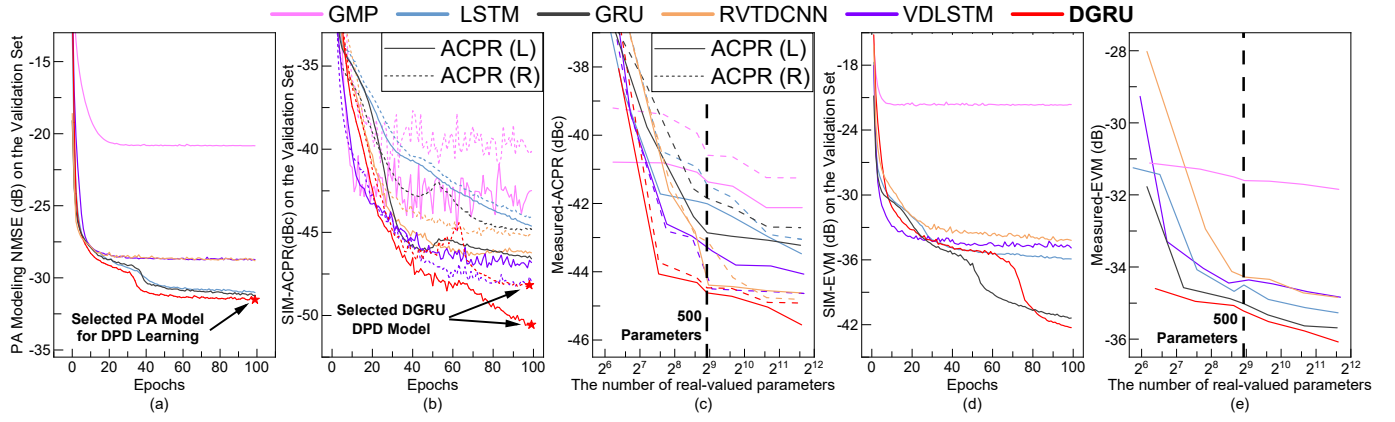
Fig. 4. Training and measurement results on 200 MHz 10-channel×20 MHz OFDM signals from the DPA_200MHz validation set. Each curve represents the best performance of each algorithm over 5 random seeds. (a) The 500-parameter PA modeling NMSE over training epochs. The gold PA model of each algorithm is saved at the lowest NMSE; (b) The 500-parameter DPD learning SIM-ACPR over training epochs. The gold DPD model of each algorithm is saved at the lowest averaged SIM-ACPR; (c) Measured ACPR vs. real-valued model parameters; (d) The 500-parameter DPD learning SIM-EVM over training epochs; (e) Measured EVM vs. real-valued model parameters.

The platform calculates NMSE, SIM-ACPR, and SIM-EVM as follows:

$$\text{NMSE} = \frac{\sum_{n=0}^{N-1}\left((I_{\hat{y}}[n] - I_y[n])^2 + (Q_{\hat{y}}[n] - Q_y[n])^2\right)}{\sum_{n=0}^{N-1}(I_y[n]^2 + Q_y[n]^2)}, \tag{1}$$

$$\text{ACPR and SIM-ACPR} = \frac{P_{\text{adj}}}{P_{\text{main}}}, \tag{2}$$

where P_{main} and P_{adj} are the main and adjacent channel powers, respectively. ACPR is measured while SIM-ACPR is software-derived. For EVM:

$$\text{EVM} = \frac{\sum_{k=0}^{K-1}\left((I_{y,\text{mea}}[k] - I_x[k])^2 + (Q_{y,\text{mea}}[k] - Q_x[k])^2\right)}{\sum_{k=0}^{K-1}(I_x[k]^2 + Q_x[k]^2)}, \tag{3}$$

with $I[k]$ and $Q[k]$ being the FFT-derived in-phase and quadrature signal components and "mea" denoting measured PA outputs. SIM-EVM is analogously computed in software.

C. Dense GRU-based DPD

We introduce a novel DGRU-DPD architecture shown in Fig. 2. This architecture has a dense skip path with the conventional GRU-RNN structure to concatenate features to the input of the output fully connected (**FC**) layer. The motivation is to mitigate the gradient vanishing issue and reuse input features to augment the linearization efficacy of the model [27]. DGRU equips an online Feature Extractor (FEx) calculating $|x[t]|$, $|x[t]|^3$, $\sin(\theta_{\mathbf{x}}[t]) = Q_{\mathbf{x}}[t]/|x[t]|$, and $\cos(\theta_{\mathbf{x}}[t]) = I_{\mathbf{x}}[t]/|x[t]|$ from $I_{\mathbf{x}}[t]$, $Q_{\mathbf{x}}[t]$ and sending them to downstream layers. The OpenDPD platform offers a pre-trained DGRU model, facilitating ease of reproduction, deployment, and fine-tuning by subsequent research endeavors.

III. EXPERIMENTAL RESULTS

A. Experimental Setup, Dataset and Training

The experimental setup is shown in Fig. 3. The I/Q data was processed using a 40 nm CMOS-based DTX [28] and

then digitized by an R&S-FSW8 analyzer. The DPA's average power is 13.42 dBm, and the PAPR of the signal without DPD is about 9.6 dB. With DPD, the PAPR of the signal is about 11 dB. OpenDPD has the flexibility to set the target gain G to arbitrary values. In this work, we particularly used:

$$G = \frac{1}{N}\sum_{n=0}^{N-1}\frac{\sqrt{I_y[n]^2 + Q_y[n]^2}}{\sqrt{I_x[n]^2 + Q_x[n]^2}} \tag{4}$$

The collected dataset named DPA_200MHz comprises 38,400 samples of 200 MHz OFDM signals sampled at an 800 MHz rate. The training is done using the ADAMW [29], [30] optimizer for 100 epochs with the ReduceLROnPlateau scheduler, starting at 1×10^{-3} with a batch size of 64, a frame length of 50 and stride of 1.

The performance of our proposed DGRU model in PA modeling and DPD learning was benchmarked against baseline models GMP, GRU, and LSTM, complemented by recent works VDLSTM [9] and RVTDCNN [13], chosen for their architectures that entail relatively less domain-specific pre-sumptions, making them apt for E2E learning. Except for the DPD parameter scan experiments, all PA and DPD models have around 500 parameters, and only the DPD model is needed during deployment. The input features of baseline GRU and LSTM models are $I_{\mathbf{x}}$, $Q_{\mathbf{x}}$, while VDLSTM and RVTDCNN used their best configuration reported in [9] and [13], where their model sizes were respectively adjusted by increasing the hidden LSTM layer and the hidden FC layer. We trained and measured all models five times with distinct random seeds to account for training stochasticity, reporting mean metrics and corresponding standard deviation values.

B. Results and Discussion

As illustrated in Fig. 4(a), the NMSE reduces with training iterations during the PA modeling process. DGRU demon-strates superior accuracy among the various algorithms by yielding the minimum error on the validation set data after

TABLE I
PERFORMANCE COMPARISON OF DPD MODELS WITH APPROXIMATELY 500 REAL-VALUED PARAMETERS ON 200 MHz 10-CHANNEL×20 MHz
OFDM SIGNALS FROM DPA_200MHz TEST SET AVERAGED ON 5 RANDOM SEEDS ± STANDARD DEVIATIONS.

Classes	DPD Models[a]	DPD-NMSE[b] (dB)	SIM-ACPR (dBc, L/R)	Measured-ACPR (dBc, L/R)	SIM-EVM (dB)	Measured-EVM (dB)
Without DPD	-	-	-32.19±0.67/-30.71±0.75[c]	-32.74/-31.71	-33.06±0.73[c]	-27.36
Baseline Models	GMP [2]	-21.07±0.06	-42.60±1.40/-39.72±1.28	-41.37±0.67/-40.59±0.24	-21.50±0.27	-31.60±0.30
	LSTM	-36.27±2.73	-44.39±0.78/-42.80±0.80	-42.01±0.58/-41.47±0.86	-35.98±2.79	-34.50±0.81
	GRU	-39.63±0.51	-44.84±1.03/-43.79±0.52	-42.86±0.53/-41.85±1.01	-42.66±1.10	-35.02±0.62
Previous Works	RVTDCNN [13]	-31.63±0.19	-44.51±0.97/-44.39±0.29	-44.39±0.47/-43.14±0.47	-33.80±0.28	-33.85±0.21
	VDLSTM [9]	-32.15±0.15	-45.42±0.79/-45.92±0.32	-43.39±0.40/-44.50±0.71	-34.27±0.12	-34.42±0.16
This Work	Features[d]+GRU	-39.84±2.23	-47.22±1.79/-46.44±0.23	-44.72±0.82/-43.38±0.63	-42.80±1.85	-35.02±0.49
	DGRU	**-41.85±0.51**	**-49.20±0.64/-46.44±0.67**	**-44.69±0.57/-44.47±0.27**	**-44.20±0.56**	**-35.22±0.38**

[a] The numbers of parameters are 495 (GMP), 488 (GRU), 488 (LSTM), 538 (VDLSTM), 500 (RVTDCNN), and 486 (DGRU).
[b] The best PA modeling NMSE values over 100 epochs on the test set are -20.85 dB (GMP), -30.85 dB (GRU), -31.02 dB (LSTM), -28.79 dB (VDLSTM), -28.76 dB (RVTDCNN), and -31.61 dB (DGRU). DPD models in this table were all trained in pairs with the DGRU PA model to ensure the same initial condition in DPD learning.
[c] Calculated from the DGRU PA models with 5 different random seeds.
[d] The extracted features include I_x, Q_x, $|\mathbf{x}|$, $|\mathbf{x}|^3$, $\sin\theta_x$, $\cos\theta_x$.
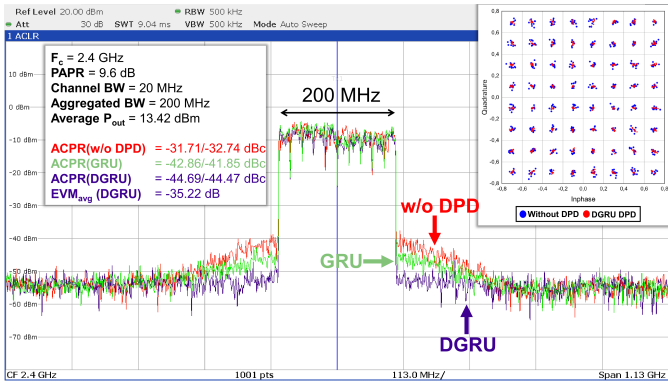

Fig. 5. Measured spectrum and constellation map on the 200 MHz Signal.

100 epochs. To standardize the initial conditions for DPD learning across all algorithms, the gold DGRU PA model with the lowest NMSE over 100 epochs was adopted.

Figs. 4(b) and (d) delineate the SIM-ACPR and SIM-EVM on the validation set across 100 epochs, respectively. The selection of the gold DPD model per algorithm was based on the lowest mean SIM-ACPR throughout the epochs. The neural network-based models notably outperformed the GMP in linearizing the wideband DPA. Furthermore, the DGRU attained the best SIM-ACPR and SIM-EVM after training.

Figs. 4(c) and (e) further reveal the measured ACPR and EVM across a parameter range spanning from 100 to 3200 measured on the gold DPD model of each algorithm. The DGRU models show better performance with a mere excess of 128 parameters. As the parameter count escalates, the DGRU-DPD consistently outperforms other algorithms.

Moreover, the comparison between the SIM-ACPR, SIM-EVM curves, and their measured counterparts with around 500 parameters (vertical dashed lines in Figs. 4(c) and (e)) exhibits a reliable congruence regarding Measured-ACPR and Measured-EVM performance rankings. For instance, simulated results accurately reflect the measurement-derived left (L) and right (R) ACPR for each DPD model. They also capture the performance hierarchy among different algorithms, indicating that the simulated metrics could be a trustworthy proxy for evaluating DPD performance in real-world applications.

Table I summarizes the performance of DPD models with around 500 parameters on the DPA_200MHz test set. The findings underscore the effectiveness of the E2E learning architecture in training DNN-based DPD models such as VDLSTM, RVTDCNN, and even baseline GRU and LSTM models to accomplish commendable linearization with ACPRs surpassing -41 dBc and EVMs around -34 dB. The DGRU-DPD is the top performer, attaining an ACPR of -44.69/-44.47 dBc, and an EVM of -35.22 dB.

Table I also delineates the incremental performance gains achieved by DGRU by integrating an FEx layer and implementing a dense skip path. By replacing the basic $I_\mathbf{x}$ and $Q_\mathbf{x}$ inputs with their feature-extracted counterparts while retaining the GRU backbone, an improvement in ACPR, from -42.86/-41.85 dBc to -44.72/-43.38 dBc, was observed. The transition to the DGRU backbone leads to the best values in both simulated and measured ACPR and EVM.

Finally, Fig. 5 shows the measured spectrum and the constellation diagrams for PA outputs, comparing scenarios with and without DPD, underscoring the capability of OpenDPD in training DNN-based DPD models to correct non-linearity in a wideband DPA effectively.

IV. CONCLUSION

This paper introduces the OpenDPD framework, aiming to streamline DPD exploration and standardize datasets for fair comparisons. Results show our E2E learning architecture can effectively train various DNN-based models for DPD purposes. Our proposed DGRU achieved superior results in the benchmark, exceeding both baselines and prior models. We plan to regularly update OpenDPD with new analog and digital PA datasets and pre-trained models with the latest DNN backbones and welcome interdisciplinary contributions.

REFERENCES

[1] S. Wesemann, J. Du, and H. Viswanathan, "Energy efficient extreme mimo: Design goals and directions," *IEEE Communications Magazine*, vol. 61, no. 10, pp. 132–138, 2023.

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

[3] P. Jaraut, M. Rawat, and F. M. Ghannouchi, "Composite neural network digital predistortion model for joint mitigation of crosstalk, i/q imbalance, nonlinearity in mimo transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5011–5020, Nov 2018.

[4] 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.

[5] Y. Zhang, Y. Li, F. Liu, and A. Zhu, "Vector decomposition based time-delay neural network behavioral model for digital predistortion of rf power amplifiers," *IEEE Access*, vol. 7, pp. 91 559–91 568, 2019.

[6] 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," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 242–254, 2019.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[8] K. Cho, B. van Merriënboer, Ç. Gülçehre, 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)*, Oct. 2014, pp. 1724–1734. [Online]. Available: http://www.aclweb.org/anthology/D14-1179

[9] 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.

[10] T. Kobal, Y. Li, X. Wang, and A. Zhu, "Digital predistortion of rf power amplifiers with phase-gated recurrent neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 6, pp. 3291–3299, 2022.

[11] J. Westhuizen and J. Lasenby, "The unreasonable effectiveness of the forget gate," *CoRR*, vol. abs/1804.04849, 2018. [Online]. Available: http://arxiv.org/abs/1804.04849

[12] T. Kobal and A. Zhu, "Digital predistortion of rf power amplifiers with decomposed vector rotation-based recurrent neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4900–4909, 2022.

[13] X. Hu, Z. Liu, X. Yu, Y. Zhao, W. Chen, B. Hu, X. Du, X. Li, M. Helaoui, W. Wang, and F. M. Ghannouchi, "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, 2022.

[14] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[15] M. Marcus, G. Kim, M. A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger, "The Penn Treebank: Annotating predicate argument structure," in *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*, 1994. [Online]. Available: https://aclanthology.org/H94-1020

[16] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 159–165, 2004.

[17] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, pp. 1417–1418(1), November 2001. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/el_20010940

[18] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative learning control for rf power amplifier linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 9, pp. 2778–2789, 2016.

[19] P. N. Landin, S. Gustafsson, C. Fager, and T. Eriksson, "Weblab: A web-based setup for pa digital predistortion and characterization [application notes]," *IEEE Microwave Magazine*, vol. 16, no. 1, pp. 138–140, 2015.

[20] M. S. Alavi, R. B. Staszewski, L. C. N. de Vreede, and J. R. Long, "A Wideband 2× 13-bit All-Digital I/Q RF-DAC," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 732–752, 2014.

[21] D. Jung, S. Li, J.-S. Park, T.-Y. Huang, H. Zhao, and H. Wang, "A cmos 1.2-v hybrid current- and voltage-mode three-way digital doherty pa with built-in phase nonlinearity compensation," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 525–535, 2020.

[22] M. Beikmirza, Y. Shen, L. C. N. de Vreede, and M. S. Alavi, "A wideband four-way doherty bits-in rf-out cmos transmitter," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 12, pp. 3768–3783, 2021.

[23] D. P. N. Mul, R. J. Bootsman, M. Beikmirza, M. S. Alavi, and L. C. N. de Vreede, "The efficiency and power utilization of current-scaling digital transmitters," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–17, 2023.

[24] M. Beikmirza, L. C. de Vreede, and M. S. Alavi, "A low-complexity digital predistortion technique for digital i/q transmitters," in *2023 IEEE/MTT-S International Microwave Symposium - IMS 2023*, 2023, pp. 787–790.

[25] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[26] C. Tarver, L. Jiang, A. Sefidi, and J. R. Cavallaro, "Neural network dpd via backpropagation through a neural network model of the pa," in *Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 358–362.

[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[28] 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.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[30] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7